Higher Order Statistical Learning for Vehicle Detection in Images

A.N. Rajagopalan   Philippe Burlina   Rama Chellappa*
Center for Automation Research
University of Maryland
College Park, MD - 20742, USA.

abstract
Abstract

The paper describes a scheme for detecting vehicles in images. The proposed method approximately models the unknown distribution of the images of vehicles by learning higher order statistics (HOS) information of the 'vehicle class' from sample images. Given a test image, statistical information about the background is learnt 'on the fly'. An HOS-based decision measure then classifies test patterns as vehicles or otherwise. When tested on real images of aerial views of vehicular activity, the method gives good results even on complicated scenes. It does not require any a priori information about the site. However, it is amenable to augmentation with contextual information. The method can serve as an important step towards building an automated roadway monitoring system.

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 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]. Although the work described in this paper deals with static images, it can be extended to video data.

The vehicle detection scheme presented in this paper uses higher order statistics (HOS) of the images of vehicles to get a better approximation to their unknown distribution. Training data samples of vehicles are first clustered and the statistical parameters corresponding to each cluster are estimated. Clustering is based on an HOS-based decision measure which is obtained by deriving a series expansion for the multivariate probability density function in terms of the Gaussian function and the Hermite polynomial. Given a test image, the background information is learnt 'on the fly'. Detection is then performed by searching the test image for patches of vehicles at all points in the image and across different scales. A vector of difference measurements of the test pattern with respect to each of the clusters is computed using the HOS-based closeness measure. A simple thresholding scheme then determines whether the test pattern belongs to the class of vehicles or not. The HOS-based measure has good discriminating capability and the results are quite encouraging. No pre-processing operations are carried out on the vehicle patterns.

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

2. HOS-Based Decision Measure

In this section, we derive a series expansion for a multivariate probability density function (p.d.f) in terms of the Gaussian function and the Hermite polynomial. An HOS-based decision measure is then derived from this expansion.

Let the random vector $\underline{X} = [X_1 \ X_2 \ \ldots X_N]^T$ and $\underline{X} \sim N(\underline{0}, I)$. If $\underline{t} = [t_1 \ t_2 \ \ldots t_N]^T$, then the moment generating function of $\underline{X}$ is given by $\Phi(\underline{t}) = E\left[\exp\left(\underline{t}^T \underline{X}\right)\right]$. Since these random variables are statistically independent, $\Phi(\underline{t}) = \exp\left(\frac{1}{2}\underline{t}^T\underline{t}\right)$. Therefore, $E\left[\exp\left(\underline{t}^T\underline{X} - \frac{1}{2}\underline{t}^T\underline{t}\right)\right] = 1$. Replacing $\underline{t}$ by $\underline{t}+\underline{s}$, we get

$$E\left[\exp\left(\underline{t}^T\underline{X} - \frac{1}{2}\underline{t}^T\underline{t}\right)\exp\left(\underline{s}^T\underline{X} - \frac{1}{2}\underline{s}^T\underline{s}\right)\right] = \exp\left(\underline{t}^T\underline{s}\right).$$

Expanding this equation in Taylor series, we obtain

$$E\left[\sum_{m,n=0}^{\infty}(\underline{t}^{\otimes n})^T\frac{\overline{H}_n(\underline{X})}{n!}\frac{\overline{H}_m^T(\underline{X})}{m!}(\underline{s}^{\otimes m})\right] =$$
$$\sum_{m,n=0}^{\infty}\frac{1}{n!}(\underline{t}^{\otimes n})^T I_{q_n q_m}(\underline{s}^{\otimes m}),$$

where $\otimes$ represents the tensor product and $\underline{t}^{\otimes n} = \underbrace{\underline{t} \otimes \underline{t} \otimes \ldots \otimes \underline{t}}_{n \text{ times}}$. The matrix $I_{q_n q_m} = O_{q_n q_m}$ for $n \neq m$ and $I_{q_n q_m} = I_{q_n}$ for $n = m$, where $O_{q_n q_m}$ is the zero matrix with q_n rows and q_m columns while I_{q_n} is the identity matrix of dimension q_n. The vector $\overline{H}_n(\underline{x})$ is given by

$$\overline{H}_n(\underline{x}) = \left[(\underline{D}_t^{\otimes n})\exp\left(\underline{t}^T\underline{x} - \frac{1}{2}\underline{t}^T\underline{t}\right)\right]_{\underline{t}=0}$$

and $\underline{D}_t = \left[\frac{\partial}{\partial t_1} \ \frac{\partial}{\partial t_2} \ldots \frac{\partial}{\partial t_N}\right]^T$. The dimensions of $\overline{H}_n(\underline{x})$ and $\overline{H}_m(\underline{x})$ are given by q_n and q_m, respectively. By equating the coefficients of $\underline{t}$ and $\underline{s}$ on both sides, we obtain the important orthogonality relation

$$E\left[\underline{H}_n(\underline{X})\underline{H}_m^T(\underline{X})\right] = I_{p_n p_m} \ . \tag{1}$$

Note that the expectation is w.r.t $N(\underline{0}, I)$. In (1), $\underline{H}_n(\underline{x})$ 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 dimensions of the vectors $\underline{H}_n(\underline{x})$ and $\underline{H}_m(\underline{x})$ are given by p_n and p_m, respectively. The term $H_{k_i}(x_i)$ is the Hermite polynomial of order k_i and is defined as $H_{k_i}(x_i) = \left[\frac{\partial^{k_i}}{\partial t_i^{k_i}}\exp\left(t_i x_i - \frac{1}{2}t_i^2\right)\right]_{t_i=0}$. Similarly, one can derive

an orthogonality relation in terms of $N(\underline{\mu}, R)$ as

$$E\left[\underline{H}_n\left(R^{-\frac{1}{2}}(\underline{X} - \underline{\mu})\right) \ \underline{H}_m^T\left(R^{-\frac{1}{2}}(\underline{X} - \underline{\mu})\right)\right] = I_{p_n p_m} \ .$$

Let $\underline{Y} = R^{-\frac{1}{2}}(\underline{X} - \underline{\mu})$ and $\underline{y} = R^{-\frac{1}{2}}(\underline{x} - \underline{\mu})$. If $\underline{X}$ has mean $\underline{\mu}$ and covariance R, then using the above orthogonality relation, the multivariate probability density function $f(\underline{x})$ can be written as

$$f(\underline{x}) = N(\underline{\mu}, R)\left(1 + \sum_{n=3}^{\infty} E\left[\underline{H}_n^T(\underline{Y})\right]\underline{H}_n(\underline{y})\right) \ . \tag{2}$$

We define the HOS-based closeness measure as $-\log f(\underline{x})$. It is interesting to note here that when $f(\underline{x})$ is Gaussian, the HOS-based decision measure neatly reduces to the normalized Mahalanobis distance.

3. Vehicle Detection

In this section, we propose an HOS-based vehicle detection scheme that finds vehicles by searching an image for square patches of different views of the vehicle at all points of the image and across different scales.

We use a statistical distribution-based model for detection. It is to be expected that the joint density function for the class of vehicles is unlikely to be well-modeled by a strict/simple Gaussian fit $N(\underline{\mu}, R)$. Since only a finite number of moments would be computable in practice, the unknown p.d.f. is approximated up to its mth order joint moment using (2) as

$$f(\underline{x}) = N(\underline{\mu}, R)\left(1 + \sum_{n=3}^{m} E\left[\underline{H}_n^T(\underline{Y})\right]\underline{H}_n(\underline{y})\right) \ , \tag{3}$$

where $\underline{Y} = R^{-\frac{1}{2}}(\underline{X} - \underline{\mu})$ and $\underline{y} = R^{-\frac{1}{2}}(\underline{x} - \underline{\mu})$. Thus, higher order statistics are used to get a better approximation to $f(\underline{x})$. The corresponding HOS-based finite order decision measure is given by

$$-\log N(\underline{\mu}, R)\left(1 + \sum_{n=3}^{m} E\left[\underline{H}_n^T(\underline{Y}))\right]\underline{H}_n(\underline{y}))\right) \ . \tag{4}$$

3.1. Clustering using HOS

We model the distribution of vehicles by fitting the data samples of vehicles with multi-dimensional clusters. The idea of using multi-dimensional clusters to model the p.d.f may be traced back to the works in [8, 9]. Traditional k-means clustering algorithms based on the Euclidean or the Mahalanobis distances [10, 11] work satisfactorily under Gaussian assumptions. However, if the actual distribution of the data is non-Gaussian, then traditional k-means may fail to yield

satisfactory results. Hence, we propose a clustering algorithm that uses higher order (> 2) statistics for improved clustering. The closeness measure that we use for clustering is given by (4).

The Clustering Algorithm

1. Obtain k initial pattern centers from the image database of vehicles. Divide the data set into k clusters by assigning each data sample to the nearest pattern center in Euclidean space.

2. Initialize the joint moments (second and onwards up to order m) of all k clusters.

3. Recompute pattern centers to be the centroids of the current data partitions.

4. Using the current set of k pattern centers and their higher order moments, recompute data partitions by re-assigning each data sample to the nearest cluster using the HOS-based decision measure defined in (4). If the data partitions remain unchanged or if the maximum number of inner-loop (i.e steps 3 and 4) iterations have been exceeded, proceed to step 5. Otherwise, return to step 3.

5. Re-compute the moments (second and onwards up to order m) of all k clusters from their respective data partitions.

6. Using the current set of k pattern centers and their cluster moments, recompute data partitions by re-assigning each data sample to the nearest cluster using the HOS-based measure. If the data partitions remain unchanged or if the maximum number of outer loop (i.e steps 3 to 6) iterations have been exceeded, proceed to step 7. Otherwise, return to step 3.

7. Return the current set of k pattern centers and their joint moments (up to order m), for each cluster.

It was found, after some experimentation, that six clusters were adequate for our purpose.

3.2. Classification

We describe two methods for classifying vehicles. Both the methods use the HOS-based decision measure for classification. However, the first method does not use the background information in the test image while the second method incorporates dynamic background learning while testing. The two methods are compared to demonstrate the advantages of learning the background scene. As would be shown, the second method yields better overall detection rate while simultaneously reducing the number of false matches.

Method 1: (No Background Learning)
In this method, given a test image, vehicle detection is performed by searching the image for square patches of vehicles at all points in the image and across different scales. A vector of difference measurements of the test pattern with respect to each of the vehicle clusters is computed using the HOS-based closeness measure. The minimum difference value is then determined and simple thresholding is used to decide whether the test pattern belongs to the class of vehicles or not.

Method 2: (Dynamic Background Learning)
The background information can be used to improve the performance of the vehicle detector algorithm. This can be done by first learning the background dynamically as follows. Initially, the test image is scanned at its highest resolution for square patches that are not vehicles. As non-vehicles usually far outnumber the vehicles in a given test image, we use a loose threshold to classify non-vehicles based on the already available statistical knowledge of the vehicles. Since the background usually constitutes a major portion of the test image, one can obtain a sufficient number of samples that are not vehicles. The non-vehicle patterns are next distributed into six clusters using the HOS-based closeness measure and the statistical parameters corresponding to each of the six clusters are estimated.

Finally, the test image is again searched for square patches of vehicles but now at all points in the image and across different scales. A vector of difference measurements of the test pattern is computed with respect to each cluster (six corresponding to vehicles and six corresponding to the background) using the HOS-based closeness measure. If the minimum difference value corresponds to that of a vehicle cluster and is less than a specified threshold, the test pattern is declared as a vehicle, else not. The above modified condition helps to reduce the false patterns considerably. Knowledge about the background allows us to relax the threshold which in turn leads to an improvement in the vehicle detection rate while simultaneously keeping down the number of false matches. It must be noted that the above condition could lead to a few misses when some of the vehicles look quite like the background.

4. Experimental Results

We present results on the performance of the proposed HOS-based vehicle detection systems with and without background learning. The training set consisted of about 500 grey-scale patterns of vehicles (cars here), each of dimension 16×16 pixels only. The methods were then tested on real images that were aerial views of vehicular activity on roadways captured with a stationary camera. The training set was distinct from the test set. As a compromise between accuracy of representation and computational complexity, we choose

$m = 3$ (in (3) and (4)) for our experiments.

Figures 1 and 2 show the output results corresponding to these methods for some test images. Multiple boxes represent detection at different scales. 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. Although the figures are self-explanatory, for the purpose of comparison, a quantitative breakdown of the performance of the methods is also tabulated. Considering the complexity of the problem which lies in detecting very small objects against a cluttered background, we note that both the methods are able to detect vehicles reasonably well in all the images. Even those with non-frontal views are detected. However, there are some misses; these include vehicles that were smaller than the chosen window size and hence could not be reliably detected. For the calculations in Table 1, vehicles that were much smaller than the training image size set were not taken into consideration. We note that the average detection rate for

System	Vehicles	Detected	False Alarms
Method 1	90	56	40
Method 2	90	66	13

Table 1. Performance comparison.

Method 1 is about 62%. However, the method suffers from quite a number of false alarms. This is due to the fact that it completely ignores information about the background scene. On the other hand, the scheme described in Method 2 which incorporates dynamic background learning clearly outperforms Method 1. It has a higher average detection rate (about 73%) and only one-third of the false alarms as compared to Method 1. Clearly, dynamic background learning enhances performance significantly.

5. Conclusions

We have described a scheme for vehicle detection against a cluttered background. The proposed method (Method 2 in the paper) uses higher order statistics of data samples of vehicles to get a better approximation to the distribution of the image patterns of vehicles. The background is learnt dynamically while testing. An HOS-based decision measure is used for classification. The scheme gives good results, even on fairly complicated scenes. The method can also detect non-frontal views. It may be possible to improve the performance further by using contextual information.

We are currently working on extending this scheme to video. The vehicle detection rate is expected to improve significantly due to the motion information that is available in video data.

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), pp. 249-253, Feb. 1999.

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

[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), pp. 3-24, Nov. 1998.

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

[8] 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.

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

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

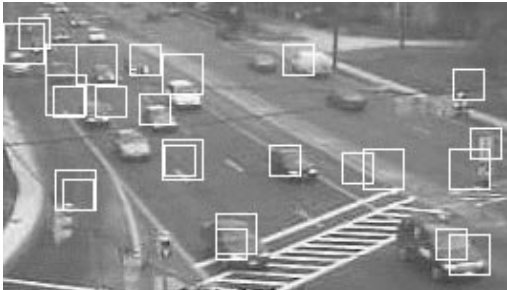[11] R. O Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons Inc., 1973.
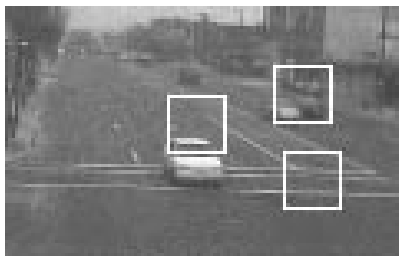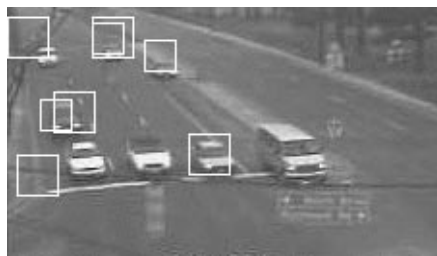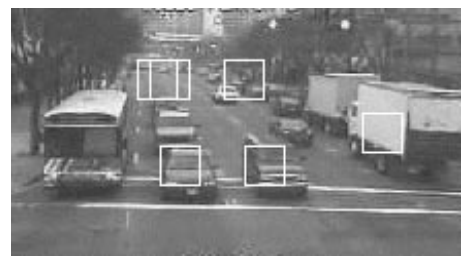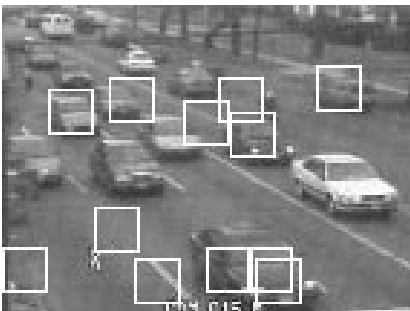
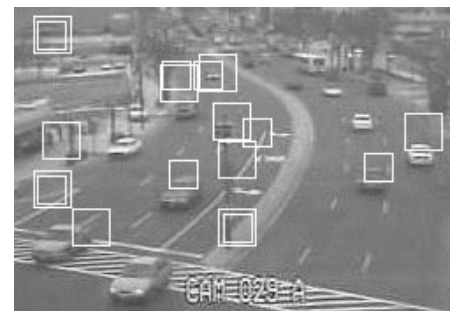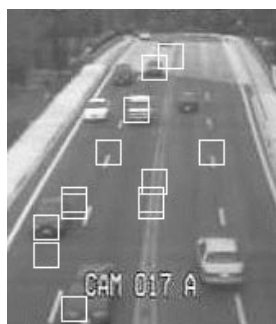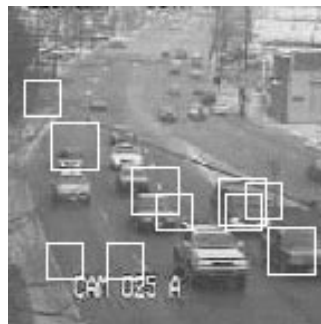A: 1/1/1   B: 5/8/1   C: 3/5/1

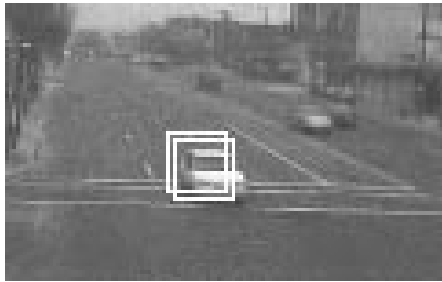D: 11/13/5   E: 8/15/5   F: 4/9/5

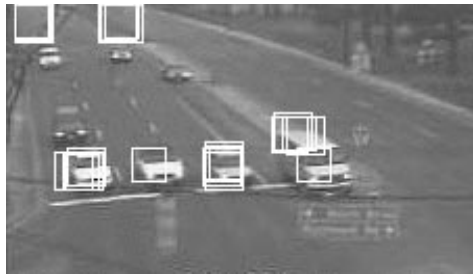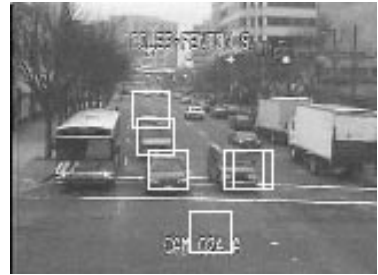G: 6/10/3   H: 4/4/5   I: 5/8/6

J: 4/8/5   K: 5/9/3

Figure 1. Results of the HOS-based vehicle detection scheme without any background learning. Multiple boxes represent detection at different scales. Below each image, three values are given: they correspond to the number of vehicles detected, the number of vehicles present in the image, and the number of false alarms, respectively. Vehicles that are much smaller than the training image are ignored in the calculations.
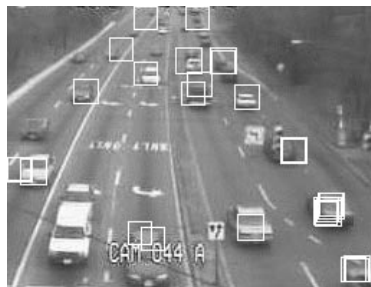
A: 1/1/0   B: 6/8/1   C: 3/5/1
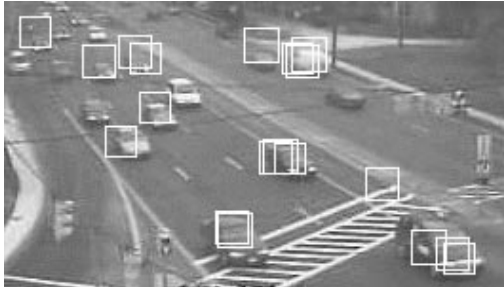
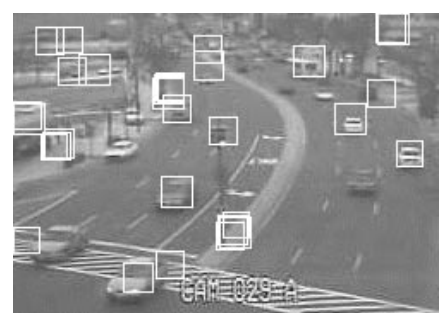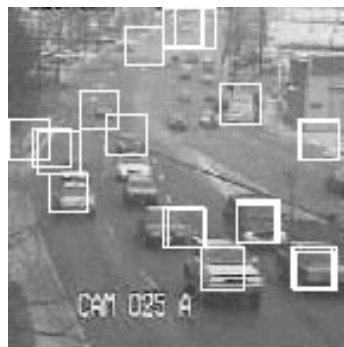D: 10/13/1   E: 12/15/0   F: 5/9/2

G: 8/10/0   H: 3/4/2   I: 6/8/5

J: 5/8/0   K: 7/9/1

Figure 2. Results of the HOS-based vehicle detection scheme with dynamic background learning. Multiple boxes represent detection at different scales. Below each image, three values are given: they correspond to the number of vehicles detected, the number of vehicles present in the image, and the number of false matches, respectively. Vehicles that are much smaller than the training image are ignored in the calculations.