PERFORMANCE ANALYSIS AND LEARNING APPROACHES FOR VEHICLE DETECTION AND COUNTING IN AERIAL IMAGES

V. Parameswaran *P. Burlina* *R. Chellappa*

Computer Vision Laboratory
Dept. of Electrical Engineering
University of Maryland, College Park, MD 20742

ABSTRACT

Robustness as well as the ability to work in an unsupervised mode are two desirable features of algorithms employed on large image databases. This paper describes parameter optimization strategies for such algorithms and motivates these strategies by focussing on aerial image exploitation and studying certain specific aerial image understanding algorithms, namely local vehicle detection and global vehicle configuration detection. The paper first gives a brief introduction to the problem in the context of aerial imagery. Next, a high level description of the algorithms and parameters that need to be optimized is given. Strategies for parameter optimization are illustrated using examples. Finally a discussion on the applicability and scope for improvement of the strategies is given.

1. INTRODUCTION

Context based aerial image understanding (AIU) has been studied quite extensively recently, the approach often referred to as 'site model based image exploitation' *e.g.*, [1], [3]. Such an approach is very well suited for routine AIU work such as detection and counting of vehicles and global vehicle configurations, because it enables discrimination between irrelevant changes (*e.g.* illumination change, seasonal variations etc.) and actual changes. The approach consists of maintaining an abstraction of a site, referred to as the *site-model*. The site-model consists of a coordinate system and various *features* which are models of the objects that the site consists of (*e.g.* parking lots, roads, buildings etc.). A typical AIU task would be to detect changes in a newly acquired image of the site. Note that the new image could have been taken with a different illumination level or in a different season from the earlier images. Prior to doing any processing on the image, it is necessary to *register* [9] the image with the site, which means calculating the image's transformation with respect to the site origin. In other words, registration provides a mapping from features in the site, which are image independent, to the image in question. This enables running the detection algorithms on selected portions of the image, for example, parking lots or certain areas of interest. The detection algorithms, however, depend upon several parameters as input. The optimal choice of these parameters can be very image dependent and clearly, for reasonably accurate results one cannot use the same parameters for every image. Optimal choice of these parameters appears to need extensive input from the image analyst on a per-image basis. This presents an obvious problem as regards the application of these algorithms in batch (i.e. unsupervised) modes over large image databases. In this paper, we address various issues involved in solving the above problem. We consider limiting or completely eliminating, the number of tuning parameters using the following strategy: preliminary sensitivity analysis is used to identify those parameters which the results of the algorithm are most sensitive to. A compound measure of sensitivity is chosen as the expected risk function, computed empirically over a set of training images. The parameters which the algorithm is the least sensitive to, are "frozen" to their best values (off-line parameter optimization). We provide 'on-line' training tools for the remaining parameters which the performance depends the most on. For such on-line training tools, we introduce the concept of *control patches* that are fixed portions of a given site and are used for automatic parameter tuning. The concept is explained in more detail in Section 4. The approaches used for parameter training and optimization are set in a classic hypothesis testing framework using Bayesian and Neyman-Pearson strategies. A related issue is the assessment of the sensitivity of the algorithms to non-image dependent parameters, for example, to any model or template parameters. This problem is important for site-model based exploitation and some initial results are reported here.

The support of the Advanced Research Projects Agency (ARPA order No. 8979), the U. S. Army Topographic Engineering Center under contract DACA 76-92-C-0024 is acknowledged.

2. OUTLINE OF ALGORITHMS

Vehicle Detection and Counting

The aim of this module [4] is to reliably detect and count vehicles in aerial images. The vehicle detection process consists of two stages, namely, an edge detection stage and a testing stage. The edge detection stage uses the Canny edge operator [2]. This stage involves specification of two thresholds, referred to as the high and low thresholds, and a mask size for convolving the image with the Canny operator. The testing phase operates on the edge data generated from the edge detection phase. A generalized Hough transform of the image is calculated using the known shape and size (user specified) of the sample vehicle provided, and a vote for possible centers of vehicles is thus calculated for each pixel in the image. A hypothesis is made for a vehicle at each such candidate center, and the edge map is exam-

ined. The quality of the match between the edge map and the candidate vehicle outline is judged using a threshold referred to as the overlap threshold. If the degree of overlap exceeds the threshold, a vehicle is declared. Thus, the overall vehicle detection process involves the specification of four parameters - the *Canny mask size, low threshold, high threshold, overlap threshold*, and the model vehicle.

Configuration Detection

The purpose of this module [4] is to detect vehicle formations in images. The configuration detection process uses spectral analysis, wherein spectral compliance windows are inferred from model information to search for impulsive components representing periodic object configurations such as convoys on roads or vehicles in parking lots. We have designed a detection rule on the observation space $\mathcal{O}$ which consists of the absolute spectrum magnitude associated with the impulsive component and its value relative to the median spectrum magnitude. The rule tests the dominant spectral component within a compliance window at the base and corresponding harmonic frequencies and takes into account the normalized spectrum magnitude K_a associated with the maximum peak at f^* within a compliance window and the ratio of the spectrum magnitude to the median K_{med} of this magnitude computed over the compliance window, denoted by K_r. Thus, the parameters to be optimized are K_a, and K_r.

3. DETECTION AND TRAINING

On-line parameter training and off-line parameter optimization are set in a hypothesis testing framework. Let H_0 and H_1 correspond to the two hypotheses (absent/present),

$$H_i : P(\mathbf{Y}|i) = P_i(\mathbf{Y})$$

The acceptance and rejection regions are designed over some observation space $\mathcal{O}$. The observation vector $\mathbf{Y}$ in the case of the vehicle detector is simply the overlap value. In the case of the configuration detector, it is composed of the two spectral measures described in Section 2.

A decision rule d is simply given by $d(\mathbf{Y}) = \mathcal{I}_{\mathcal{R}}(\mathbf{Y})$, with $\mathcal{I}_{\mathcal{R}}$ the indicator function on the acceptance region $\mathcal{R}$. The acceptance region admits a parametric form

$$\mathcal{R} = \mathcal{R}_{\mathcal{V}} = \{\mathbf{Y} \text{ such that } b(\mathbf{Y}; \mathcal{V}) \geq 0\}$$

where the form of the critical/acceptance regions can be inferred from the distributions satisfied by the observation vector under either hypothesis. The design of the detection rule is set up according to the following two strategies:

- (A) Bayesian strategy: Find $\mathcal{V}^* = \text{argmin}(E\{R(d)\})$. This strategy consists of minimizing the expected risk [8] where the cost factors C_{fa} and C_{nd} are chosen to balance the cost associated with a false alarm and a non-detection:

  $$E\{R(d)\} = C_{nd} P_0(\mathcal{R}_{\mathcal{V}})\pi_0 + C_{fa} P_1(\mathcal{R}_{\mathcal{V}}{}^c)\pi_1$$

  where $P_0(R)$ and $P_1(\mathcal{R}_{\mathcal{V}}{}^c)$ are the false alarm and non-detection probabilities, and π_i are the priors.
- (B) Neyman-Pearson: Find $\mathcal{V}^* = \text{argmax}(P_1(R))$ subject to $P_0(R) < \alpha$.

4. VEHICLE DETECTOR OPTIMIZATION

The relative impact of the parameters involved was studied by carrying out experiments on images with known ground truths. Specifically, the empirical expected risk $E\{R(d)\}$ including both false alarm and non-detection rates were computed for the Canny mask size, the Canny thresholds and the overlap thresholds. The results for the Canny mask size and the overlap thresholds are shown in figures 1 and 2 respectively. From this empirical risk it was inferred that the variation in detection performance was smaller for the Canny parameters when they varied within their operational limits, while the performance varied significantly with the overlap threshold. Thus, for the former parameters it would be sufficient to derive optimal estimates (off-line optimization), while for the latter, we need to consider a training tool (on-line optimization).

For off-line Canny parameter optimization we use the strategy (A) with the expected risk computed over the training set. We choose $C_{nd} = 0.5$ and $C_{fa} = 0.5$ and we assume equal priors, which is equivalent to a minimum probability error rule. The empirical value of $E\{R(d)\}$ is computed over the range of Canny parameters as a function of the Canny threshold and mask sizes. The Canny parameters are set off-line to the values minimizing the expected risk. These values were found to be 5 for the mask size and $(200, 400)$ for the minimum and maximum thresholds. For the on-line overlap threshold training, we use strategy (B). Empirical detection and false alarm probability are derived from predetermined control patches. For each newly acquired image, the overlap threshold is automatically computed from the empirical probabilities derived from the control patches. Denote by $P_1(\mathcal{R}_{\mathcal{V}}) = r(P_1(\mathcal{R}_{\mathcal{V}}{}^c))$ the empirical ROC curve derived by varying the overlap threshold $\mathcal{V}$; then we want to find $\mathcal{V}^*$ satisfying $\mathcal{V}^* = \text{argmax}(P_1(\mathcal{R}_{\mathcal{V}}))$ subject to

- $P_0(\mathcal{R}_{\mathcal{V}}) < \alpha$
- $\frac{d(r(p))}{dp}\,|_{p=P_1(\mathcal{R}_{\mathcal{V}}{}^c)} > \beta$

The second condition ensures that the slope of the ROC curve for the given value of $\mathcal{V}$ is not too small, i.e. that an increase in false alarm probability can be traded for a significant increase in detection probability. One issue remains, as to the selection of control patches.

When a new image is acquired, the overlap threshold is automatically "calibrated" using control patches as explained above. In site model based image exploitation of aerial imagery, control patches with their associated ground truth can be specified once by representing them as specific features of the site (in terms of position, orientation etc.). This way, whenever a new aerial image of the site needs to be subjected to any of the above algorithms, the control patches will be mapped to portions of image, as part of the registration process (where the image is registered with the site model). The above is possible when we can identify control patches in the parking lot that are known to be always empty (such as passageways, parking lot exits, etc.), or always full (areas located near building entrances). If such patches cannot be reliably identified, an interactive method could be used, where the image analyst inspects and validates the control patches over a subset of the images to be processed before initiating a batch procedure.

5. CONVOY DETECTOR OPTIMIZATION

In the case of the convoy detector, the components of $\mathbf{Y}$, the 2D observation vector, are the logarithms of the parameters K_a and K_r that were described in section 2, i.e. $\mathbf{Y} = (ln(K_a), ln(K_r)) = (L_a, L_r)$. Let H_0 and H_1 correspond to the two hypotheses, with H_0 the hypothesis that no peak is present; the decision rule is simply $d(L_a, L_r) = \mathcal{I}_{R_\mathcal{V}}(L_a, L_r)$. We use a Bayesian strategy (A) for deriving the acceptance region from a training set of images. The acceptance region boundary is parameterized by vector $\mathcal{V}$ and chosen as

$$R = R_\mathcal{V} = \{(L_a, L_r), \text{such that } b(L_a, L_r; \mathcal{V}) \geq 0\}$$

Assuming that the joint conditional probability distributions on $\mathbf{Y} = (L_a, L_r)$ are Gaussian, i.e. $H_i : P(\mathbf{Y}|i) \sim N(\mathbf{m}_i, \mathbf{\Sigma}_i), i = 0, 1$; then the log-likelihood ratio function is a quadratic function in $\mathbf{Y}$[5], i.e. $(\mathbf{Y} - \mathbf{m}_1)^t \mathbf{\Sigma}_1^{-1} (\mathbf{Y} - \mathbf{m}_1) - (\mathbf{Y} - \mathbf{m}_0)^t \mathbf{\Sigma}_0^{-1} (\mathbf{Y} - \mathbf{m}_0)$. We assume dissimilar covariances for which the boundary equation $b(.,.;) = 0$ is a conic section. The acceptance region is determined by finding $\mathcal{V}^*$, which minimizes the expected value of the conditional risk computed over the training set, i.e. $\mathcal{V}^* = \text{argmin}(E\{R_\mathcal{V}(d)\})$. As an example, ten images from a particular site were chosen as a training set, and $b(.,.;)$ was assumed to be an elliptic boundary. The parameters of this elliptic boundary were optimized on the set of control images. The expected value $E\{R(d)\}$, computed over the training set, is a noisy function of V, in part due to the modest size of the training set. V* is determined by using the Nelder-Mead Simplex algorithm [7]. This function is non-convex, and therefore the simplex algorithm is not guaranteed to converge. Furthermore, the minimum is not unique. The resulting boundary for $C_{nd} = 0.55$ and $C_{fa} = 0.45$ is shown in Fig. 3. In this example, the compound detection performance yields a false alarm probability equal to 0.115385 with a non-detection probability equal to 0.085714.

6. MODEL PARAMETER MISSPECIFICATION

We also characterized the sensitivity of parking lot occupancy detection to misspecification of the model parameters. The 3D dimensions of the vehicle were varied, and the detection and false alarm probabilities were computed on the set of test images. The resulting probabilities are displayed as functions of these dimensions in Figure 4 for detection of active parking lots. In Figure 4, the upper surface represents the probability of detection as a function of the 3D width W and length L. The lower surface represents the false alarm probability. Situations where the width is greater than the length constitute a misspecification by $\pi/2$ of the actual vehicle orientation. In this figure we see that the resulting performance is not too sensitive to reasonable variations in size.

7. CONCLUSION

We addressed the problem of automatic tuning of parameters for image understanding algorithms with specific reference to local vehicle detection and global vehicle configuration detection algorithms in AIU tasks. We proposed two different approaches for optimization, depending upon the relative impact of the parameters on the performance of the detection algorithms, the relative impact being inferred from runs on a set of test images. Naturally, it should be ensured that the test images are large in number so as to be statistically comprehensive. This would further justify freezing parameters of lesser significance to their off-line optimal values. Also, in the present schemes, the sensitivity to tuning and model parameters is evaluated empirically. This study can be complemented by a more in-depth analytical study of these algorithms' sensitivity, as is done in [6] where each step of the algorithm can be described analytically and a first order sensitivity analysis can be carried out. In the present study, empirical distributions are used. Instead, the moments can be estimated and simple hypothesis testing techniques can be used to verify the consistency of the observed data with the assumed distribution and estimated moments. Also, on line parameter optimization techniques have been introduced by use of the concept of 'control patches' which is very useful and practicable in the context of site-model based image exploitation and merits further investigation for parallels in other applications.

8. REFERENCES

[1] P. Burlina, R. Chellappa, C. Lin, and X. Zhang, "Context-Based Exploitation of Aerial Imagery," in *Proc. Workshop on Model-Based Vision*, (Boston, MA), June 1995.

[2] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 8, pp. 679–698, November 1986.

[3] R. Chellappa et al, "Site-Model Based Monitoring of Aerial Images," in *Proc. ARPA Image Understanding Workshop*, (Monterey, CA), pp. 295–318, November 1994.

[4] R. Chellappa et al, "An Integrated System for Site Model Supported Change Detection," in *Proc. ARPA Image Understanding Workshop*, (Palm Springs, CA), March 1996.

[5] K. Fukunaga, *Statistical Pattern Recognition*, Academic Press, 1990.

[6] X. Liu, T. Kanungo, and R. Haralick, "Statistical Validation of Computer Vision Software," in *Proc. ARPA Image Understanding Workshop*, 1996.

[7] J. Nelder and R. A. Mead, "A Simplex Method for Function Minimization," *The Computer Journal*, Vol. 7, 1965.

[8] V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, 1988.

[9] X. Zhang, P. Burlina, Q. Zheng, and R. Chellappa, "Automatic Image to Site Model Registration," in *Proc. IEEE Conference on Acoustics, Speech and Signal Processing*, (Atlanta, GA), March 1996.

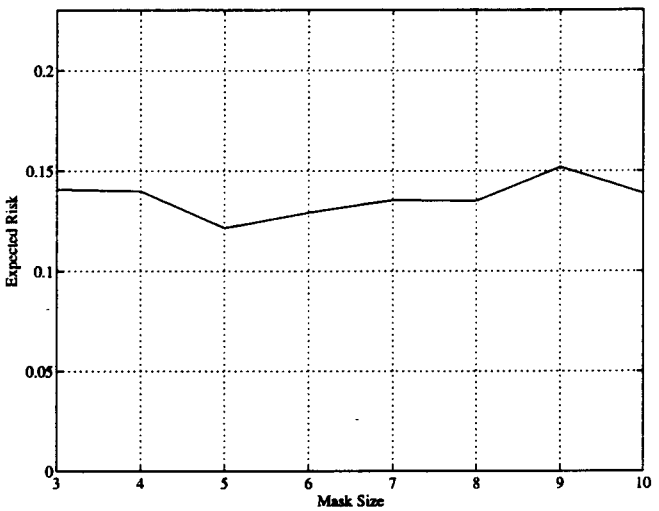

Figure 1: Vehicle detection: expected risk as a function of Canny mask size.

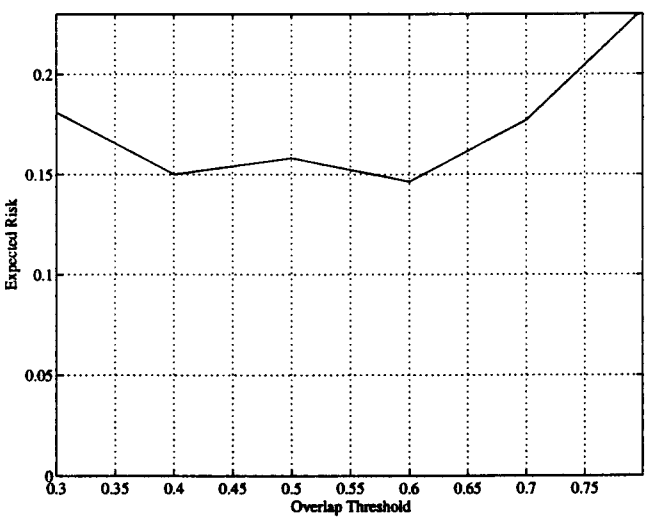

Figure 2: Vehicle detection: Expected risk as a function of overlap threshold.

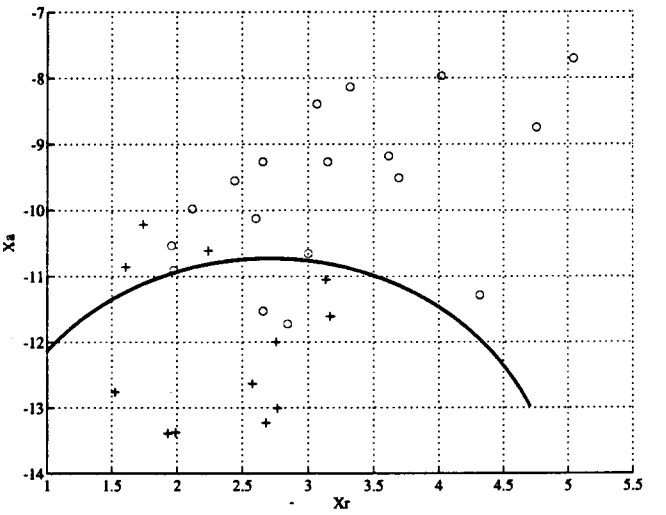

Figure 3: Decision region obtained from training images. Active parking lots are indicated by 'o's and inactive by '+'s in the (L_r, L_a) plane.

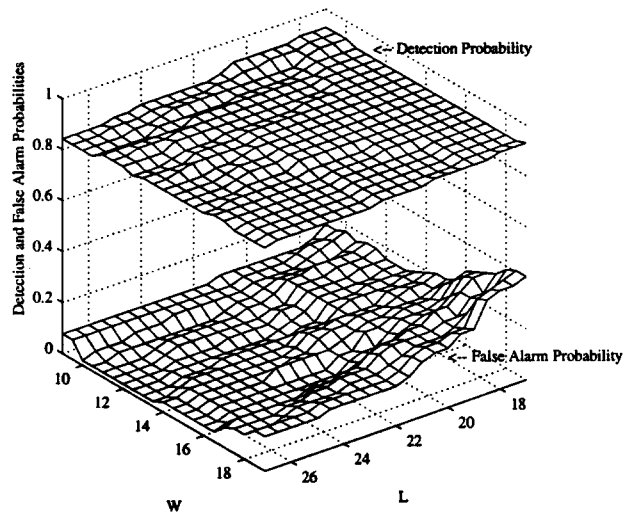

Figure 4: Sensitivity of the detection of active parking lots to misspecification of vehicle dimensions.