A Spectral Attentional Mechanism Tuned to Object Configurations

Philippe Burlina, Rama Chellappa, Fellow, IEEE, and C. L. Lin

Abstract—This paper describes an attentional mechanism based on the interpretation of spectral signatures for detecting regular object configurations in areas of an image delineated using context information. The proposed global operator relies on the spectral analysis of edge structure and exploits spatial as well as frequency-domain constraints derived from known geometrical models of monitored objects. A decision theoretic method for learning acceptance detection regions is presented. Applications of this attentional mechanism are demonstrated for several aerial image interpretation tasks for attentional as well as recognition purposes. Specific examples are described for detecting vehicle formations (such as convoys), qualifying the geometry of detected formations, or monitoring the occupancy of regions of interest (such as parking areas, roads, or open areas). Experiments and sensitivity analysis results are reported.

I. INTRODUCTION

T HIS PAPER describes an attentional mechanism based on the characterization and analysis of spectral signatures. This attentional mechanism is based on a global operator detecting regular and periodic edge configurations in areas of the image selected using contextual information. Attentional mechanisms are important components of human and machine vision systems [7], [13]. Examples of tasks that could profit from the inclusion of such mechanisms include computationally intensive tasks such as visual guidance, recognition, or the detection of activities in aerial image analysis. Attentional mechanisms are essential in that they limit and focus processing to specific regions of interest (ROI's) or indicate the occurrence of an activity calling for further analysis.

The presence of local regularities or specific organization is one indication of human or natural activities. Such organization can take many forms and can be detected by various means: the presence of specific geometric relationships between image features such as line symmetry, orthogonality, parallelism, etc.; statistical redundancy as measured by information theoretic measures such as entropy, mutual information [11] or minimum description length techniques [14]; or geometric redundancy in the form of periodicity, possibly detected by spectral analysis techniques. Attentional mechanisms based on spectral responses are consistent with some neurophysiological evidence pointing to processes equivalent to Fourier operations in the human visual system [17].

Frequency-domain techniques have also been employed for various image analysis tasks such as texture analysis [3], object recognition, and the design of spatial invariants [1], or motion flow estimation [9]. The broader problem of detecting specific geometrical feature configurations has also been addressed in works such as [4], [12], and [21]. In [4], object grouping into formations is achieved using a knowledge based association scheme. In [21], line configurations are recognized from their Hough transform signatures. In [12], an optimal algorithm for recognizing equally spaced collinear subsets is reported.

Our method relies on the spectral analysis of regular edge structure and exploits frequency domain constraints derived from known geometrical models of monitored objects. The use of such global operators and spectral methods in image analysis has often been limited by their lack of robustness and high sensitivity to segmentation errors. In this paper, these drawbacks are alleviated by using site model information. A site model is used which contains a geometrical description of the area under scrutiny and of relevant site features (ROI's, buildings, roads, etc.). Along with collateral and auxiliary information [2], it also includes imaging and photometric parameters associated with available images. When a site model has been constructed from previously acquired imagery, a typical model supported exploitation cycle is as follows: The image is automatically registered to the existing site, ROI's are delineated, and specific image analysis algorithms are applied according to the monitoring task (change detection, site model refinement, or verification). Here, site model information enables the robust use of spectral analysis techniques by allowing for spatial domain as well as frequency domain search localization: Searches are conducted along initial ROI's derived from site models. Object and configuration models are then used to restrict frequency domain searches to compliance regions. These ideas are clarified in Section II. We describe decision theoretic strategies for designing and learning the decision rule from a set of training images. Applications of the attentional mechanism described here are demonstrated for specific aerial image interpretation tasks such as change detection or object recognition [6], [15]. Several tasks considered here include detecting vehicle formations such as convoys, qualifying the geometry of detected formations, or monitoring the occupancy of regions of interest such as parking areas, roads, or open areas.
Experimental results on real images are reported for each of these tasks. The sensitivity to misspecification of the model parameters is examined. The method is formalized in Section II, it is specialized to aerial image understanding tasks in Section III. Finally, experiments and sensitivity analysis are described in Section IV.

II. SPECTRAL CHARACTERIZATION OF REGULAR EDGE STRUCTURES

A. Overview

The attentional mechanism described here is triggered by the presence of regular object configurations. Partial or complete prior information in the form of site and object models is used to infer information about the object geometry, the object groupings’ configuration geometry, or the hypothesized locations or ROI’s including such configurations. Generic ROI’s are either one-dimensional (1-D), and as such modeled as ribbons (e.g., road, railroad, landing strip, river, etc.), or two-dimensional (2-D), and modeled as regions (e.g. parking lot, open area, etc.). The detection of regular and periodic structure is carried out either by spectral analysis of the brightness intensity directional derivative in the 1-D case, or analysis of the gradient magnitude in the 2-D case. In the case of ribbons, the directional derivative is computed. Evidence of a specific periodic structure is found by searching for a peak within a frequency window referred to as compliance window derived from known object model constraints. The 2-D case is similar but yields compliance regions, which entail the use of various search strategies depending on the type of prior information available about the configuration geometry and object geometry. The observation space for the detection rule includes the peak absolute spectrum magnitude and its value relative to its median computed within a specified frequency window. Acceptance regions are derived by using a decision theoretic approach by exploiting a set of training images. These ideas are expanded and formalized in this section.

B. Characterization

Consider first the 1-D case where ROI’s are represented by ribbons in the image. Let \( \mathbf{r} \) denote the image position of a point. Assume that a ribbon’s skeleton describes a smooth curve \( \mathcal{C} \), given by \( r_{\mathcal{C}}(s) \), the image point position as a function of the curvilinear abscissa \( s \). Let \( \mathbf{N}(s) \) and \( \mathbf{T}(s) \) be the normal and tangent vectors to \( \mathcal{C} \) at \( r_{\mathcal{C}}(s) \). The ribbon \( R \) characterized by width \( W(s) \) at \( r_{\mathcal{C}}(s) \) is defined by the set of points satisfying

\[
R = \{ \mathbf{r} : \mathbf{r} \in L(s) \} \quad \text{with} \quad L(s) = \{ \mathbf{r} : \mathbf{r} = \lambda \mathbf{N}(s) + r_{\mathcal{C}}(s), \lambda \in \left(-\frac{W(s)}{2}, \frac{W(s)}{2}\right) \}.
\]

Construct \( k(s) \), the directional derivative of the brightness intensity \( I(\mathbf{r}) \) averaged along the width of the ribbon

\[
k(s) = \frac{1}{\mathcal{R}} \int_{\mathcal{R}} \frac{d\mathbf{r} D T I(\mathbf{r})}{W(s)}.
\]

Periodically organized objects are detected from spectral analysis of directional edges \( k(s) \). Let \( F(f) = F[k(f) \mathbf{f}] \) denote the Fourier transform (FT) of \( k(s) \). Consider the occurrence of periodically situated objects, such as a road convoy, vehicles parked in a parking lot, or railroad cars on tracks. These are represented by a multiply replicated function \( c(s) \)—derived from the directional derivative profile of the object—within a window of length \( p \), as

\[
k(s) = \text{rect} \left( \frac{s}{p} \right) \left( c(s) * \sum_{n=-\infty}^{\infty} \delta(n/p^*) \right) (s)
\]

Periodic edge structure is characterized by detecting peaks and their corresponding harmonics from the amplitude spectrum \( |K(f)| \). The strategy used to reliably measure this occurrence when an additive clutter noise component is present, i.e. for \( k'(s) = k(s) + n(s) \), is described in Section II-D. The 2-D case allows for many possible object configurations and scenarios and is relevant for analyzing situations involving regular object configurations in open areas. Let \( k(\mathbf{r}) = \|\nabla I(\mathbf{r})\| \), the gradient magnitude at \( \mathbf{r} \), and consider the 2-D FT

\[
K(f) = F[k(f)] = \int_{-\infty}^{\infty} \mathbf{d} \mathbf{r} k(\mathbf{r}) \exp(-j2\pi\mathbf{r}^T \mathbf{f})
\]

with \( \mathbf{f} = [f_0, f_1]^T \). Consider a grouping of objects with arbitrary replications and orientation. Their gradient magnitude is described by

\[
k(\mathbf{r}) = \sum_{\mathbf{n}} c(\mathbf{R}_{\mathbf{n}} \mathbf{r} - \mathbf{v}_n)
\]

with \( \mathbf{v}_n \) and \( \mathbf{R}_{\mathbf{n}} \) respectively denoting an arbitrary 2-D translation vector and rotation matrix. Since

\[
F[\mathbf{A}^{T} \mathbf{f}] = \det(\mathbf{A})^{-1} F[\mathbf{A} \mathbf{f}]
\]
for \( \det(A) \neq 0 \), then with \( k(r) \) as in (2) we have
\[
K(f) = \sum_{n} \exp(-j2\pi n^T R_{n} f) P(r)[R_{n} f].
\]
If the objects satisfy periodic configurations, i.e., the directions and intervals of replication are given by the vectors \( v_0 \) and \( v_1 \), with
\[
v_i = n_0 + j v_1
\]
and the objects have identical orientations, i.e., \( R_{n_0} = R_{n_1} \), within a support area \( A \), then \( k(r) \) becomes
\[
k(r) = I_A(r) \langle e^{j(D_0 \cdot n_1)} \rangle \delta(r + V n_1), \quad n = [i, j]^T.
\]
The resulting spectral representation [8] is
\[
K(f) = \left| I_A(f) \right| \langle C(J_{10} \cdot n_{0}) \rangle \delta(f - V^{-1} n_1),
\]
where \( I_A(f) = F[I_A(r) \delta(r + V n_1)] \). Its 2-D periodicity matrix \( U \) equals
\[
N = V^{-1} n_1.
\]
The situation describes, by simple selection of the shape of the support region \( A \), a wide array of configurations of special interest such as circular, semicircular, wedge, pyramid, block, or linear formations. As in the 1-D case, a periodic configuration can be detected by searching for its corresponding 2-D base and harmonic peaks whose positions are given by the frequency domain periodicity matrix \( U = V^{-1} T \). The rule for reliably detecting such peaks in the presence of noise is described in Section II-D.

**C. Spectral Searches on Compliance Windows**

Constraints on three-dimensional (3-D) geometrical object and configuration models yield spectral compliance windows where dominant components characteristic of certain types of regular configurations are searched for. Let \( f^* \) denote the base component arising as a result of a regular object formation on a ribbon. Knowledge of the object geometry and context enables us to derive bounds on the positions of the peaks: Knowledge of the average image object length \( l \) taken on the image region covered by the ribbon is derived from the known object model and the camera orientation parameters. Context also allows for knowledge of the minimum \((1 + \epsilon)l\) as well as maximum \((1 + \mu)l\) replication distance of this object. \( f^* \) must lie within a spectral compliance window \( W_{C_r} \), defined as
\[
W_{C_r} = [1/(1 + \mu) l, 1/(1 + \epsilon) l].
\]
A regular structure is hypothesized if a peak is found at frequency \( f^* \), also referred to here as the base or dominant spectral component
\[
f^* = \arg \max_{f} |K(f)|, \quad f \in W_{C_r}.
\]
When the dominant component is found in the spectral compliance window, corroborating evidence of harmonic components at \( nf^*, n = 2, 3 \cdots \) is verified. The decision rule is built on a 2-D observation space composed of the logarithm of the spectrum magnitude and the logarithm of the spectrum magnitude normalized by the median (see Section II-D). Similar compliance windows can be derived in the 2-D case. Consider the periodicity matrix decomposition
\[
V = \hat{V} \text{diag}(\tau_0, \tau_1), \quad \hat{V} = \hat{U_0} \hat{U_1}, \quad |\hat{U_0}| = |\hat{U_1}| = 1.
\]
The spectral periodicity matrix \( U \) equals \( U = V^{-1} T \), so that \( U = V^{-1} \text{diag}(\tau_0, \tau_1) \). Note also that \( u_i \cdot v_j = \delta_{i,j} \), where \( i, j = 0, 1 \) since \( U^{-1} V = (V^{-1} T)^T \). The spectral periodicity vectors \( u_i \) are orthogonal to their alternate spatial counterparts, since \( u_i \cdot v_j = 0 \) if \( i \neq j \). Using \( u_i \cdot v_i = 1 \), and \( v_i = \tau_i u_i \), we get
\[
|\hat{U}_i| = |\cos(u_i, v_i)|^{-1} \tau_i^{-1} = |\sin(u_i, v_i)|^{-1} \tau_i^{-1}, \quad i = 0, 1
\]
which was assumed in the previous section that \( \det(u_i, v_i) \neq 0 \). The case where \( \det(u_i, v_i) = 0 \) consists of linear configurations in a 2-D domain, with a unique direction of replication, and one may write the gradient magnitude map as
\[
k(r) = \sum_{n} \delta(r - n\tau_1)
\]
for which
\[
K(f) = C(\tau_1) \sum_{i} \delta(f - V^{-1} R_{n_1} n_1)
\]
which gives rise to impulsive lines orthogonal to the direction of \( v_1 \). In sum, the spectral signature offers clues to the spatial geometry (orientation and dimension) of the detected configuration, and can be used to determine the spatial orientations of the object configurations. Alternatively, if site model information is available, compliance windows can be derived to guide the detection process. Using geometrical considerations similar to the ones presented in the 1-D case, the following compliance (longitudinal and lateral) constraints may be inferred:
\[
\frac{1}{(1 + \mu_0) l_0} < f^* R_{n_1} \sin(u_0, v_1) < \frac{1}{(1 + \epsilon_0) l_0},
\]
\[
\frac{1}{(1 + \mu_1) l_1} < f^* R_{n_1} \sin(u_1, v_1) < \frac{1}{(1 + \epsilon_1) l_1}
\]
where \( l_0, l_1, \mu_0, \mu_1, \epsilon_0, \epsilon_1 \) are derived from the task at hand (examples are given in Section III). How much is known about \( V \) depends on the context. The search strategy for detecting a compliant impulse component makes use of various schemes depending on the available model information.

- If only the orientation direction of the object configuration is available, then the constraints in (4) translate into radial search spaces on a half-plane along directions \( u_0 \) and \( u_1 \), or alternatively into a search sector if an uncertainty on these directions is specified.
- If only the object dimensions are known, the search is conducted along annular rings, generated by a full \( 2\pi \) revolution of the sets in (4).
Finally, if both dimensions and directions can be inferred from the site model or from the context, the search space reduces to annular sectors.

D. Detection

Various factors can impair the detection process: the clutter noise component present across the spectral compliance window, poor imaging and illumination conditions that lead to weak gradient magnitude, and inaccuracies in site and region boundary models. The factors interfering with the existence of a uniform critical region across experimental conditions are variations in viewing conditions and illumination.

We design a detection rule that tests the dominant component at the base and corresponding harmonic frequencies. This rule takes into account the spectrum magnitude $|K(f^*)|$ associated with the maximum peak at $f^*$ within a compliance window. To keep critical/detection regions constant across viewing directions, the spectral estimates need to be normalized by the aspect ratio. The resulting normalized spectrum magnitude associated with the peak is denoted by $K_\alpha$. The second element in our decision rule is the ratio of the spectrum magnitude over the median of this magnitude computed over a specified frequency window $K_{\text{med}}(f^*) = |K(f)|/K_{\text{med}}$ with $K_{\text{med}}$ denoting the “median” magnitude. The detection acceptance and rejection regions are designed over the 2-D observation vector $Y$ whose components are the logarithms of these two measures, i.e., $Y = \left( \ln(K_\alpha), \ln(K_{\text{med}}) \right) = (L_\alpha, L_{\text{med}})$. The harmonics $n f^*$ or $n^2 f^*$ are also checked for the presence of an impulsive component. In many situations, it is difficult to test more than the first harmonic component(s).

We turn to the design of the detection rule. Let $H_0$ and $H_1$ correspond to the two hypotheses, with $H_0$ the hypothesis that no peak is present, as follows:

$$H_i: P(Y|\theta) = P_\theta(Y).$$

The decision rule is $d(L_\alpha, L_{\text{med}}) = I_R(L_\alpha, L_{\text{med}})$, with $I_R$ the indicator function on the acceptance region $\mathcal{R}$. The acceptance region could be manually tuned using heuristic rules, an example of which would be that peaks with lower spectral magnitude should be tolerated for larger magnitude ratios. Manual tuning, however, is often not a desirable nor practical option due to the large amount and diversity of imagery typically available for exploitation. We use a decision theoretic approach with a Bayesian strategy for deriving the acceptance region from a training set of images. This strategy consists in minimizing the expected risk [18] function chosen here with the following costs:

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

where $P_0(\mathcal{R})$ and $P_1(\mathcal{R}^c)$ are the false alarm and nondetection probability and $\pi_0$ the priors. The cost factors $C_{\text{fa}}$ and $C_{\text{td}}$ balance the costs associated with a false alarm and a nondetection. $C_{\text{fa}}$ and $C_{\text{td}}$ are designed so as to reflect the priorities or the final function of the attentional mechanism (preprocessor, or stand-alone). If this mechanism is used exclusively as an attentional preprocessor to local detection schemes, then higher false alarm rates can be tolerated. The acceptance region boundary is parameterized by vector $\mathcal{V}$ and chosen as

$$\mathcal{R} = \mathcal{R}_V = \{ (L_\alpha, L_{\text{med}}), \text{such that } b(L_\alpha, L_{\text{med}}; \mathcal{V}) \geq 0 \}.$$

Assume that the joint conditional probability distribution on $Y = (L_\alpha, L_{\text{med}})$ is Gaussian, i.e.,

$$H_i: P(Y|\theta) \sim N(m_i; \Sigma_i), \quad i = 0, 1$$

then the log-likelihood ratio function is a quadratic function in $Y$ [10], i.e., $(Y - m_0)^t \Sigma_0^{-1}(Y - m_0) - (Y - m_1)^t \Sigma_1^{-1}(Y - m_1)$. If the covariance matrices are equal, the boundary region is a line. If they are dissimilar, this boundary is either elliptic, parabolic, or hyperbolic, depending on the covariance matrices. We assume dissimilar covariances for which the boundary equation $b(\cdot, \cdot) = 0$ is a conic section. The acceptance region is determined by finding $\mathcal{V}^*$, which minimizes the empirical expected value of the conditional risk computed over the training set, i.e.,

$$\mathcal{V}^* = \arg\min \{ E\{R(d)\} \}. \quad (5)$$
III. APPLICATION TO AERIAL IMAGE UNDERSTANDING TASKS

The above spectral-based method is applied for attentional as well as recognition purposes in the context of aerial image understanding. Several specializations of the proposed attentional mechanism are used for i) detecting vehicle formations such as convoys; ii) qualifying the geometry of detected formations; and iii) monitoring the occupancy of regions of interest such as parking areas, roads, and open areas.

We first discuss the procedural structure of a context-based exploitation cycle [5]. Most of the following operations are carried out on a platform allowing for the definition and manipulation of CAD-like objects (site models) and including the basic photogrammetric tools (resection, etc.) [19]. These operations assume a site model has already been constructed. A site model contains a geometrical description of the site under scrutiny and of relevant site features (areas, buildings, roads, rivers, railroad tracks, landing strips, highways, etc.); it contains description and models of specific objects (vehicles, etc.), and it includes imaging and photometric parameters associated with available images along with collateral and auxiliary information (time, weather conditions, etc.) [20]. When a new image is acquired, it is automatically registered to the site. This is accomplished using a multiresolution image-to-image registration method [22]. After registration, site model information is used in the newly acquired image: ROI’s are delineated and image object dimensions are derived. Fig. 1(a) shows an initial image, and Fig. 1(b) shows the new image precisely registered to the site model. From the camera model, 3-D models of ROI’s can be transformed and registered to the image plane: roads and parking and staging areas are delineated, and the attentional mechanism and local object detection schemes are applied to the monitored areas highlighted in Fig. 2(a).

As previously explained, compliance window parameters are derived from context information. For example, if the task is convoy detection on roads, the average image vehicle length $l$ taken on the image region covered by the road is first derived from the known vehicle model and the camera orientation parameters. For safety reasons, vehicles follow each other at distances greater than $\epsilon l$, for some $\epsilon > 0$; an upper bound for $f^*$ is then given by $(1+\epsilon)^{-1}$. Similarly, vehicles following each other at too great distances are not considered to belong to convoys. The maximum distance between vehicles can then be expressed in terms of vehicle lengths $l$ as $(1+\mu)l$.
Fig. 3. Monitoring of parking lot occupancy and road convoys.

with $\mu > c$. This yields a lower bound $(1 + \mu)l_0^{-1}$ for the dominant spectral component and for the minimum length. Similarly, when dealing with monitoring 2-D regions such as parking lots, partial or complete information can be derived from the site model about the longitudinal and lateral bounds $l_0, l_1, \mu_0, \mu_1, c_0, c_1$ as well as the orientation yielding the periodicity matrix. These can correspond to vehicle dimensions as well as spacing between vehicles for parking lots. The various degrees to which this information is known yields the specific search strategies described in the previous section.

In the case of train detection, these constraints are even more precise, and can be readily derived from the type of monitored trains (freight, passenger, etc.).

The above detection scheme can be applied directly to convoy detection or the detection of full parking areas. Alternatively, the above mechanism can serve as a preprocessing attentional mechanism prior to the application of more computationally intensive detection and counting algorithms. We use it also to cue a local vehicle detector. To this end, we infer the locations of edges giving rise to the periodic structure determined through frequency-domain analysis, referred to here as periodic loci, and thereafter apply local object detection schemes to corroborate the presence and refine the count of objects at these positions. Let $\mathcal{L}$ denote the set of periodic loci. A local vehicle detector can be applied to the reduced spatial window centered at these locations $\mathcal{L}$. As we are primarily concerned with high-altitude imagery, in our implementation vehicles are modeled as 3-D cuboids with specified width, length, and height specifications. The vehicle detector consists of an extractor and a verifier and relies on template matching using information derived from the geometric models. Vehicles’ locations are hypothesized at the periodic loci. The extractor then performs template matching at the hypothesized vehicles locations. The verifier then checks for shadows to guarantee the correctness of the results. More details on the vehicle detection scheme may be found in [15].

IV. EXPERIMENTAL RESULTS

Our experimental results were obtained using a set of 40 images from the RADIUS-MB2 series. While still a statistically modest test sample, these images are interesting since they offer an array of varying viewing and illumination conditions as well as activity scenarios. Using the techniques for building
the site models reported in [19], a site model is constructed from multiple resected images. Fig. 1(a) shows an original image (M17). In Fig. 1(b), the site model is superimposed on the image. After registration of M17 [5], a selected number of delineated regions of interest (roads, parking areas, and staging areas) are derived from the camera external orientation parameters and the site model. Fig. 2(a) shows the parking and road areas monitored in our experiment. The needed apparent image dimensions of monitored objects, as well as road and parking area boundaries and directions, are derived from the known site model and resected camera parameters for the image under analysis.

In the 1-D detection case, directional derivatives are calculated on ROI's made up of ribbons. For roads, this direction is the direction of the tangent to the road; for unidimensional parking areas, this direction is the dominant orientation of the parking area. Ad hoc tuned as well as learned critical regions were used for these experiments. Sample detection results are shown in the following figures: Fig. 2(a) shows an newly acquired image, and our method is applied to the operator-selected areas (shown in yellow). Fig. 2(b) and (c) shows detected active parking areas. The resulting detected active roads are highlighted in Fig. 2(d) and Fig. 3(a). Fig. 3(b)–(d) shows the results when the attentional mechanism is applied to all the monitored areas shown in Fig. 2(a). Active areas are shown in red, and inactive roads are shown in green. Inactive parking lots are not highlighted. When the attentional mechanism is used as a preprocessor to vehicle detection and counting, the resulting set of periodic loci is shown in Fig. 4(a). After application of the vehicle detector, the resulting detected vehicles are shown in Fig. 4(b).

Detection and false alarm rates resulting from running the global activity detector on all or a subset of the 40 model board MB2 images are reported here for hand-tuned rules and learned decision rules. The detection of convoys is tested on some of these images using a tuned detection rule (see Table II), yielding a compound nondetection probability of 0.107143 and a false alarm probability of 0.161290. The detection of active (full) parking areas on all the MB2 images yields a compound nondetection probability of 0.071429 and a false alarm probability of 0.173077 for hand-tuned critical region boundaries. Ten of the 40 images were chosen as a training set. As discussed previously, among all possible conic sections, \( \nu(\cdot, \cdot) \) is assumed to be an elliptic boundary. The parameters of this elliptic boundary are optimized on the set of control images. The expected value \( E[R(d)] \), computed over the training set, is a noisy function of \( V \); this in part is due to the modest size of the training set. \( V^* \) is determined by using Nelder–Mead Simplex algorithm [16]. This function is nonconvex, and therefore the simplex algorithm is not guaranteed convergence. Furthermore, the minimum is not unique. The resulting boundary for \( C_{rd} = 0.55 \) and \( C_{fa} = 0.45 \) is shown in Fig. 5. In this case, the compound detection performance yields a significantly
TABLE III
DETECTION OF ACTIVE PARKING AREAS: HAND-TURNED RULE

<table>
<thead>
<tr>
<th>Image No.</th>
<th>Total Visible Areas</th>
<th>Active Areas</th>
<th>Missed Areas</th>
<th>False Positive Areas</th>
<th>Image No.</th>
<th>Total Visible Areas</th>
<th>Active Areas</th>
<th>Missed Areas</th>
<th>False Positive Areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>29</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>31</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>34</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>35</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>36</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>37</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>38</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>39</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig. 5. Decision region obtained from training images. Active parking lots are indicated in “•,” inactive “+” in the $(L_x, L_y)$ plane.

better false alarm probability, 0.115385 and a comparable nondetection probability, 0.085714. The breakdown of the nondetection and false alarms is provided for the MB2 images in Tables III and IV.

The previous result shows that the detection rule is not too sensitive to the varying illumination and viewing conditions present in the data set. We wish to better characterize the sensitivity of the detection performance to misspecification of the object model. To this end, the 3-D dimensions (width and lengths) of the vehicle were varied, and the compound detection and false alarms were computed on the set of test images. The resulting probabilities are displayed as a function of these dimensions in Fig. 6 and the resulting expected risk in Fig. 7 for detection of active parking areas. In Fig. 6, the upper surface represents the probability of detection as a function of the 3-D width $W$ and length $L$. The lower surface in the same figure represents the false alarm probability. Situations where width is greater than 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 a reasonable variation in size. (The empirical risk function for this case is shown in Fig. 7.) For convoy detection, as seen from simulation results in Fig. 8, as the values deviate from their optimal specifications (the middle of the grid), performance degrades. As $W$ and $L$ increase, the false alarm probability decreases along with the detection probability. This highlights the importance of context in this particular application.
TABLE IV
DETECTION OF ACTIVE PARKING AREAS: LEARNED RULE

<table>
<thead>
<tr>
<th>Image No.</th>
<th>Total Visible</th>
<th>Active Areas</th>
<th>Missed Areas</th>
<th>False Positive</th>
<th>Image No.</th>
<th>Total Visible</th>
<th>Active Areas</th>
<th>Missed Areas</th>
<th>False Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>28</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>31</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>34</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>35</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>36</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>37</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>38</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>39</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig. 7. Risk function computed on all images for active parking areas detection.

For specific applications such as those described here, this method compares favorably with the direct application of symbolic recognition schemes. By using a global operator we can readily identify specific ROI’s, while the complexity of this method is minimal compared to symbolic recognition. Delineation from known site models, computation of the gradient, and computation of the spectrum are of modest computational complexity.

The same process is carried out in 2-D image regions. The global analysis is applied to the RADIUS Fort Hood image data set, one image of which is shown in Fig. 9. An enlarged area is shown in Fig. 10 for which the gradient magnitude is computed. We show the corresponding gradient magnitude in Fig. 11. This image exhibits two types of periodicities; one corresponds to the row periodicity, the second to the car periodicity. Fig. 13 is a contour plot of the magnitude spectrum. The “principal” direction is parallel to the spatial row periodicity direction. The detected vehicles from the local detector are shown in Fig. 12. A contour plot of the spectrum with additional level sets is displayed in Fig. 14. The base
frequency corresponding to the car periodicity is searched for along a direction orthogonal to the principal direction, since the image is a nadir view. This leads to a directional tolerance area whose boundaries are indicated in Fig. 14. On the other hand, geometric constraints yield a compliance window in the form of an annular ring, as explained in Section II. The base frequency corresponding to the car periodicity lies at the maximum of the spectrum in a compliance window equal to the intersection of the annular ring and the angular sector, as shown in Fig. 14. For this experiment $e = 1.2$ and $\mu = 5$ were chosen.

Inferring the orientations is interesting both for registration and site model construction purposes: If these orientations are included in the site model, this information allows the refinement of camera exterior orientation parameters. If this information is not already contained in the site model, it is relevant since it may be used as a site model construction tool. Parking lots are labeled $P_1$ through $P_{23}$ from the bottom...
left to the top right of image 2 in Fig. 9. The principal and secondary directions, inferred from searching for the maximum and second largest peaks in an annular search area (itself determined from vehicle dimension constraints), lead to spatial orientations that are correct to within 4° when compared to the orientations measured from the actual images. These are 84° and 162° for P_6, P_7 and P_8 in image 2, 142° and 42° for P_1 through P_3 in image 3, and 30° and 114° for image 4 and parking lots P_6, P_7 and P_8.

V. CONCLUSION

We have presented an attentional mechanism based on the global detection of organized object configurations. This method uses site and object model information and relies on the application of global detectors and frequency-domain analysis to directional edge information computed on ROI’s. The proposed approach is illustrated for monitoring movable objects in aerial image exploitation in order to i) detect vehicle formations such as convoys; (b) qualify the geometry of detected formations; and (c) monitor the occupancy of regions of interest such as parking areas, roads, or open areas.

REFERENCES


Philippe Burlina received the B.S. degree in 1988 from Université de Technologie de Compiègne, France, and the Ph.D. degree in 1994 from the Electrical Engineering Department, University of Maryland, College Park.

He is an Assistant Research Scientist with the Center for Automation Research, University of Maryland. His current research interests include image/video processing, model-based image analysis, and image/video communications.
Rama Chellappa (M’81–SM’83–F’92) is a Professor of electrical engineering and an Affiliate Professor of computer science at the University of Maryland, College Park. He is also affiliated with the Center for Automation Research (Associate Director) and the Institute for Advanced Computer Studies. Prior to joining the University of Maryland, he was an Associate Professor and Director of the Signal and Image Processing Institute, University of Southern California (USC), Los Angeles. Over the last 15 years he has published 20 book chapters and 60 peer-reviewed journal papers. Several of his journal papers have been reproduced in collected works published by IEEE Press, IEEE Computer Society Press, and MIT Press. He is editor of Papers on Digital Image Processing (IEEE Computer Society Press), co-authored a research monograph, “Artificial neural networks for computer vision,” with Y. T. Zhou (Springer-Verlag), and co-edited a book on Markov random fields with A. K. Jain (Academic Press). His current research interests are image compression, automatic target recognition from stationary and moving platforms, surveillance and monitoring, automatic design of vision algorithms, synthetic aperture radar image understanding, and commercial applications of image processing and understanding.

Dr. Chellappa has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, and IEEE TRANSACTIONS ON NEURAL NETWORKS, and currently is a Co-Editor-in-Chief of Graphical Models and Image Processing (Academic Press). He chaired the Image and Multidimensional Signal Processing Technical Committee of the Signal Processing Society during 1994–1995. He was the General Chair of both the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, and the 1989 IEEE Computer Society Workshop on Artificial Intelligence for Computer Vision. He served as the Technical Program Chair for the 1989 NSF-sponsored workshop on Markov random fields and their applications, the 1993 Signal Processing Society Conference on Neural Networks for Signal Processing, and the 1995 International Conference on Image Processing. He has received several awards, including the 1985 NSF Presidential Young Investigator Award, a 1985 IBM Faculty Development Award, the 1991 Excellence in Teaching Award from the School of Engineering at USC, and the 1992 Best Industry Related Paper Award from the International Association of Pattern Recognition (with Q. Zheng).

C. L. Lin received the B.S. degree from National Taiwan University, Taipei, Taiwan, R.O.C., in 1988 and the M.S. and Ph.D. degrees from the University of Maryland, College Park, in 1994 and 1996, respectively, all in computer science.

He was a Research Assistant at the Center for Automation Research, University of Maryland, from 1992 to 1995. Currently, he is with Microsoft, Redmond, WA, working on video compression. His research interests also include image/video processing, object recognition and computer graphics.