# A Visual Search Inspired Computational Model for Ship Detection in Optical Satellite Images

Fukun Bi, Bocheng Zhu, Lining Gao, and Mingming Bian

***Abstract*—In this letter, we propose a novel computational model for automatic ship detection in optical satellite images. The model first selects salient candidate regions across entire detection scene by using a bottom-up visual attention mechanism. Then, two complementary types of top-down cues are employed to discriminate the selected ship candidates. Specifically, in addition to the detailed appearance analysis of candidates, a neighborhood similarity-based method is further exploited to characterize their local context interactions. Furthermore, the framework of our model is designed in a multiscale and hierarchical manner which provides a plausible approximation to a visual search process and reasonably distributes the computational resources. Experiments over panchromatic SPOT5 data prove the effectiveness and computational efficiency of the proposed model.**

***Index Terms*—Appearance analysis, context, remote sensing image, ship detection, visual search.**

## I. Introduction

AUTOMATIC ship detection in satellite images is very important for remote sensing technology. Its typical applications include maritime management, illegal fishing surveillance, and smuggling activities monitoring [1]–[5]. Ship detection in synthetic aperture radar (SAR) images has been intensively studied due to SAR's ability to work in all-weather conditions and all-time situations [3]. However, few works have been done on ship detection in optical satellite images. Compared with SAR images, optical satellite images can provide more detailed and easily interpreted characteristics of ships that can be further used in classification applications. Moreover, the signatures of nonmetallic ships in optical images are more visible than those in SAR images. Therefore, as a significant and complementary information source, it is necessary to exploit an automatic technique for ship detection in optical satellite images.

In a bottom-up manner, a number of studies focused on separating ship signatures from their surroundings according to the difference of intensity contrast or statistical distribution (e.g., [1] and [3]). These bottom-up approaches include the popular constant false alarm rate-based techniques [3], [4], and the recently addressed Bayesian decision-based approach (BD) [1]. However, in real optical satellite scenes, many geospatial objects (e.g., islets, coastlines, strong waves, and cloudlets) are also conspicuous against their surroundings. Although some existing methods use shoreline data to perform detection in sea area [1], [2], it still suffers from the interference of coastal areas due to low precision of the shoreline data. Recently, in order to reduce false alarms, some top-down cues have been introduced to characterize ship candidates' appearance, such as shape, texture, and intensity-distribution features [2], [5]. Nevertheless, these object-centered approaches are usually ambiguous when the image's resolution cannot provide sufficient details about the appearance of ships and the influences of various ship wakes. As another essential source of top-down cues, contextual information can further improve the detection performance and has been successfully applied to some object detection tasks [6]–[8]. However, available contextual interactions of automatic ship detection have seldom been investigated. This is because the computational modeling of contextual interactions poses great challenges in real optical satellite scenes.

On the other hand, it is of great significance to rapidly detect the ships in large-scene remote sensing images by using limited computational resources. The conventional bottom-up based methods are generally performed over the entire scene and cost equal computational time to analyze each region [1], [3]. Recently, a hierarchical detection method has been proposed to reduce the computational cost [2], [5]. This method is divided into several function phases which are all performed in a same spatial scale. However, it is believed that different image cues can be well represented in their suitable spatial scales, and the computational resources can be further reasonably distributed in a multiscale analysis manner.

It is well known that human visual system can find task-related objects effectively, even in a highly cluttered scene [7]–[10]. Many studies have attempted to reveal the visual search process and interpret its crucial mechanisms [8], [9]. It has been shown that, in addition to the object's appearance information, the contextual interaction also plays an essential role in a visual search process [6]–[8]. Furthermore, in order to effectively utilize neural resources, different phases of a visual search process usually operate under various spatial scales [8].

In this letter, inspired by these studies of human visual search, we propose a computational model of automatic ship detection in optical satellite images. In our model, both bottom-up attention mechanism and multiple top-down analysis are integrated into a unified framework. The main contributions of the letter can be summarized as follows. First, the proposed model is a front-end and operational detection model that can provide a plausible approximation to a visual search process. Second, to effectively discriminate the ship candidates with ambiguous appearance, a neighborhood similarity-based technique is proposed to characterize the local context interactions.

Manuscript received September 9, 2011; revised December 4, 2011; accepted December 10, 2011 Date of publication February 3, 2012; date of current version May 7, 2012.

F. Bi and B. Zhu are with the School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: bifukun@163.com; zhubc@pku.edu.cn).

L. Gao and M. Bian are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: gaolining@mail.tsinghua.edu.cn; mingmb@bit.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2011.2180695

1545-598X/$31.00 © 2012 IEEE

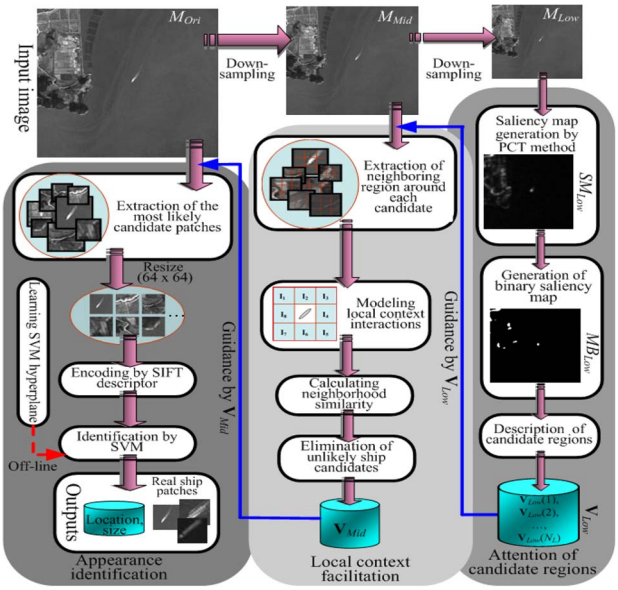


Fig. 1. General framework of the proposed model for ship detection.

Third, we design a multiscale and hierarchical framework for the detection model. Particularly, in such a hierarchical and gradually concentrative manner, the computational resources can be reasonably distributed as that in human visual search. For the available panchromatic SPOT5 satellite images with 5-m spatial resolution, we focus on detecting the ships that are longer than 50 m. Thus, the appearance features of these ships can be discriminated and described. Moreover, the ships in harbor are not considered due to our practical applications. The comparative experiments validate the effectiveness and computational efficiency of the proposed model.

## II. Detection Model

The computational framework of our detection model is shown in Fig. 1. Before performing the detection, an original scene $M_{Ori}$ is downsampled to be a middle-resolution scene $M_{Mid}$ and a low-resolution scene $M_{Low}$, respectively. The sampling rates are both set to 2 in terms of our detection task and the image's resolution. The main framework can be divided into three serial stages that mimic the visual search process. Specifically, $M_{Low}$ is first employed to select salient candidate regions by using a visual attention algorithm. Note that our model at this stage only highlights the locations of candidates and provides their approximate ranges without performing detailed analysis. Hence, $M_{Low}$ can be sampled by a mean filter. Then, in $M_{Mid}$, contextual interactions around the selected candidates are formulated, so that many negative candidates can be rejected quickly. $M_{Mid}$ is sampled by subsampling to retain representative information about intensity and texture of the candidate's neighborhood. Finally, from $M_{Ori}$, detailed appearance information of the remaining "ship-like" candidates is encoded with robust local descriptors, and then true ships can be ultimately identified by a classification technique. Such a hierarchical and multiscale analysis strategy can fully utilize different image cues at their corresponding detection stages. In addition, it also can distribute computational resources reasonably. The proposed detection model will be further described in the following subsections.

### A. Attention of Candidate Regions

The aim of attention of candidate region (ACR) stage is to rapidly select ship candidates (patches) across the entire scene with missing alarms as low as possible, and the detailed discrimination of selected candidates will be analyzed at subsequent stages. It has been noted that bottom-up-based visual attention is an effective mechanism that can quickly detect multitypes of salient regions in a clustered scene [9]. Recently, many effective approaches for saliency detection have been proposed. However, at ACR stage, we employed Yu's pulsed cosine transform (PCT)-based attention method because of its good performance in saliency detection and fast speed in computation [10]. According to this PCT method, the saliency map $SM_{Low}$ of a given low-resolution scene $M_{Low}$ can be calculated as

$$P = \text{sign}\left(C(M_{Low})\right) \tag{1}$$

$$F = \text{abs}\left(C^{-1}(P)\right) \tag{2}$$

$$SM_{Low} = G * F^2 \tag{3}$$

where $C(\cdot)$ and $C^{-1}(\cdot)$ are the 2-D discrete cosine transform and its inverse transform, respectively. $\text{sign}(\cdot)$ is the signum function, $\text{abs}(\cdot)$ is the absolute value function, and $G$ is a two-dimensional low-pass filter. We notice that in addition to the intensity-contrast saliency, PCT can also highlight multitypes of salient ship candidates and can heavily suppress the homogeneous and broad cover of cloud. Next, to describe the selected candidates, a binary saliency map $MB_{Low}$ is generated from $SM_{Low}$ by using an adaptive threshold $T_B = u - \sigma$, where $u$ and $\sigma$ denote the mean and standard deviation of $SM_{Low}$, respectively. After that, according to $MB_{Low}$, the basic information of candidate patches can be represented as a recording set $\mathbf{V}_{Low} = [\mathbf{V}_{Low}(1), \mathbf{V}_{Low}(2), \cdots, \mathbf{V}_{Low}(N_L)]$, where $N_L$ is the number of candidate patches. In particular, for the $k$th vector $\mathbf{V}_{Low}(k) = [x_k, y_k, L_k]$, the location $(x_k, y_k)$ is the centroid coordinate of the $k$th binary candidate region, and the patch size is $L_k \times L_k$ pixels. $L_k$ is the long side of smallest rectangle that can enclose the $k$th binary candidate region.

However, in real optical satellite images, many factors can cause false alarms. As a bottom-up-based method, PCT cannot discriminate the ships from other selected salient objects. Therefore, inspired by the characteristics of human visual search [6]–[8], two complementary types of top-down cues including context and appearance analysis are introduced in our detection model for further discrimination.

### B. Local Context Facilitation

Many authors in cognitive neuroscience have shown that context plays an important role in human visual search. It can effectively facilitate object search and recognition [6]–[8]. Many computational models of contextual interactions have been successfully employed in target detention tasks [7]. However, these computational models are largely task dependent, so that they cannot be directly used in our ship detection task. Compared with global contextual interactions, it is much easier to calculate and obtain local contextual interactions in optical satellite scenes. Therefore, we attempt to exploit a neighborhood similarity-based method to characterize local context interactions. It is based on a simple and reliable rationale that most of the neighboring region (i.e., local sea area) around a ship tends to have similar characteristics in both intensity

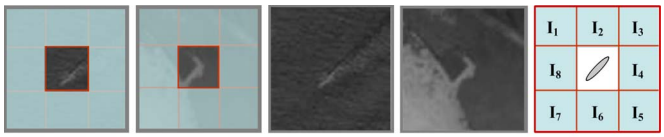


Fig. 2. Example of local context facilitation. (a) A real ship in isolation. (b) A false detection (coastline) in isolation. (c) The completely local region of (a). (d) The completely local region of (b). (e) A schematic sketch of the proposed local context facilitation.

and texture, whereas other selected geospatial objects cannot always meet this prerequisite. This method depends on the reasonable assumptions that a small local region in a wide sea scene is generally homogenous, and the ships outside the harbor always keep enough safe distance from each other. Before a detailed analysis of appearance characteristics, some false candidates are expected to be quickly eliminated by using this specified context facilitation. Fig. 2 gives an example to illustrate the local context facilitation (LCF). As can be seen from Fig. 2(a) and (b), only using the appearance features of ship candidates, we can hardly discriminate the real ships. However, when adding the local context interactions as in Fig. 2(c) and (d), the ship candidates can be disambiguated and easily identified. The proposed neighborhood similarity-based context facilitation is formulated as follows:

Under a guidance of the recording set $\mathbf{V}_{Low}$, the detection system will focus on those candidate patches in $M_{Mid}$. Then, the neighboring region around each candidate patch is evenly divided into blocks $I_1, I_2, \cdots, I_N$. Each block has the same size as the candidate patch. In this letter, we set $N = 8$. A schematic example is shown in Fig. 2(e). After that, typical intensity and texture features are extracted from each predefined neighboring block. Let $F_1(I_k), F_2(I_k), \cdots, F_M(I_k)$ denote a set of features that are extracted from a given neighboring block $I_k$, where $M$ is the features' total number. Specifically, $F_1(I_k)$ and $F_2(I_k)$ are the mean and the standard deviation of block $I_k$, respectively. In order to characterize the texture features of $I_k$, we calculate its gray-level co-occurrence matrix (GLCM), which is regarded as an effective approach to describe the texture information. $F_3(I_k)$, $F_4(I_k)$, $F_5(I_k)$, and $F_6(I_k)$ denote the GLCM's contrast, correlation, energy, and homogeneity, respectively. Thus, after the ith feature $F_i$ is obtained, we can determine whether $I_k$ is a dissimilar block against other neighboring blocks according to a rule as

$$\overline{R}_k(i) = \frac{1}{N-1} \sum_{j \neq k} F_i(I_j) \tag{4}$$

$$\begin{cases} DifFlag(i,k) = 0, & \text{if } \left|F_i(I_k) - \overline{R}_k(i)\right| / \overline{R}_k(i) \\ & \geq T_{Flag} \\ DifFlag(i,k) = 1, & \text{if } \left|F_i(I_k) - \overline{R}_k(i)\right| / \overline{R}_k(i) \\ & < T_{Flag} \end{cases} \tag{5}$$

where $\overline{R}_k(i)$ is the mean of $F_i$ that are averaged over all neighboring blocks except $I_k$. $DifFlag(i,k)$ is a flag that labels whether $I_k$ is a dissimilar block in terms of $F_i$. $T_{Flag}$ denotes the flag threshold. Next, given a candidate patch, by considering each feature type and each around block, its neighborhood similarity is defined as

$$Similarity = \frac{\sum_i \sum_j DifFlag(i,j)}{MN}. \tag{6}$$

If $Similarity \geqslant T_{accept}$, it implies that the neighboring region of a candidate patch tends to have similar characteristics in both intensity and texture, and a ship is expected to appear in such local context. Here, $T_{accept}$ is an accepted threshold. Finally, basic information of the remaining "ship-like" candidate patches are recorded using a new set $\mathbf{V}_{Mid}$, which is recorded as the same manner as $\mathbf{V}_{Low}$. Essentially, the discrimination capability of LCF is mainly influenced by the parameters $T_{Flag}$ and $T_{accept}$. On the other hand, they also make LCF tolerate several outliers in practical applications, such as some ships with long wakes, or with other geospatial objects nearby. In our detection model, two serial and complementary mechanisms are employed to eliminate false alarms. Hence, the main task of LCF is designed to remove those obvious false candidates according to their context interactions, and retain the uncertain ship candidates for the following appearance analysis. According to this principle, a grid-search-based approach is designed in our experiments (see Section IV-A) in order to discuss and test such a dual-parameter setting problem.

### C. Appearance Identification (AI)

After performing LCF stage, there still exist some subtle false alarms (e.g., islets and cloudlets) that may have similar surroundings with real ships. Therefore, we need to further analyze their appearance features that can provide complementary information for final ship identification. To obtain more detailed appearance information, the detection model extracts high-resolution candidate patches from the original scene $M_{Ori}$ under a guidance of the recording set $\mathbf{V}_{Mid}$.

An important problem in appearance analysis is that of how to extract robust characteristics of a candidate's appearance. The shape and gray-level features have been widely used in relative studies [2], [5]. However, since the ships in optical satellite images may have low intensity contrast against their surroundings, these segmentation-based features often cannot provide accurate information about the appearance of ship candidates. Considering candidate patches are small local regions in a detection scene, local descriptor technique is a potentially optimal choice due to its robust capability of local region representation. In this stage, we employ the classic scale invariant feature transform (SIFT) descriptor from [11] to encode the retained candidate patches. Considering the experimental image's resolution and our detection task, the size of concerned ship patch is typically in a range from $10 \times 10$ to $80 \times 80$. In order to encode these candidate patches effectively, we resize each patch to $64 \times 64$ pixels before extracting its appearance descriptor. Then, for each resized patch, a 128-D feature vector is calculated by the SIFT descriptor.

After that, we can ultimately identify the true ships by using the support vector machine (SVM) technique with a radial basis function kernel [12]. Moreover, we note that strong contrast cases and low contrast cases have distinctive appearance characteristics. This can largely affect the classification result. Therefore, at the SVM's training and classifying phases, we consider three typical categories including strong contrast ship, low contrast ship, and false alarm. Finally, the ship identification result is combined with both ship subcategories.

## III. Visual Search Plausibility

Human visual system can effectively identify the task-related objects from massive highly cluttered visual inputs [6], [9]. Although the exact neural process of visual search is unclear,

a number of studies in cognitive neuroscience and computer vision have shown that visual attention, contextual interaction, and appearance analysis are three crucial mechanisms in a visual search process [6]–[11]. Recently, Bar proposed a biologically plausible and simplified model that represents these crucial mechanisms in different brain areas and reveals how human brain integrates these individual elements [8]. This neuronal model emphasized that different spatial frequencies convey different information for the corresponding phases in a visual search process. Specifically, low spatial frequency representation of visual inputs is first projected into prefrontal cortex (PFC) and parahippocampal cortex (PHC). It has been shown that the PFC is more sensitive to the possible candidates, which are selected by a bottom-up attention mechanism in the primary visual cortex. In parallel, task-related contextual interactions are encoded by PHC. Then, the inferior temporal cortex (ITC) is activated by a neuronal equivalent operation "AND" of PFC and PHC. After that, the activated ITC can provide a unique identification of the task-related object by using detailed appearance information from high spatial frequency representation. Inspired by this visual search plausibility, our proposed detection model can provide a reasonable approximation to a visual search process in the human brain. As described in Section II, its computational framework is performed in a multiscale and hierarchical analysis manner as that in human visual search, and it can be divided into three serial stages including ACR, LCF, and AI, which can simulate corresponding three neuronal functions of PFC, PHC, and ITC, respectively. Moreover, their organization order and structure are similar to the Bar's neuronal model.

## IV. Experimental Results and Discussions

In this section, two comparative experiments are designed to evaluate the performance of the proposed model in terms of the detection effectiveness and the computational cost. Moreover, the parameter setting problem of LCF stage is also discussed. All experiments are performed over panchromatic SPOT5 satellite images with 5-m spatial resolution. Note that most previous works only focus on detecting simple scenes that contain relatively less interference factors [1], [2], [5]. In our experiments, we specially collected three extended training and testing datasets: the patch training data set (PTD), the easy testing data set (ETD), and the difficult testing data set (DTD). Specifically, PTD is used to train the SVM that will be used in the AI stage, and the neighboring regions around these patches are also extracted for setting parameters in the LCF stage. It consists of three kinds of training patches including 282 strong contrast ships, 246 low contrast ships, and 330 false alarms. ETD consists of 68 scenes in size of 512 × 512 pixels, which mainly contain simple cases, such as strong contrast ships, calm sea conditions, and homogeneous cloud covers. DTD consists of 62 scenes in size of 512 × 512 pixels, which mainly involve difficult cases including low contrast ships, many "ship-like" geospatial objects (e.g., small island, cloudlet), coastal areas, and heterogeneous cloud covers.

### A. Tests on Parameter Settings of LCF Stage

Based on the available data, the grid-search approach is usually used to test multiparameter setting problems (e.g., [13]). In our dual-parameter setting tests of LCF, given various pairs of $(T_{Flag}, T_{accept})$ values, the corresponding metric values can be calculated to form a topographic grid map. In this letter, two types of metrics are employed to quantify the tests: the percentage of missing true ships out of all training ships is called a missing true ships rate (MTSR), and the percentage of identified false alarms out of all training false alarms is called an identified false alarms rate (IFAR). Thus, two grid maps of MTSR and IFAR can be, respectively, generated by varying the parameter values $(T_{Flag}, T_{accept})$ over the PTD data set. According to the analysis of LCF's task in detection scheme as described in Section II-B, the principle of setting suitable parameters $(T_{Flag}, T_{accept})$ is to preferentially ensure low MTSR, and then achieve IFAR as high as possible. Hence, we first find the low value regions in MTSR grid map (less than 2%). Then, under the limitation of these regions, the suitable parameters ($T_{Flag} = 2, T_{accept} = 0.8$) can be obtained by searching the maximum in IFAR grid map.

TABLE I
Comparative Results of Ship Detection Effectiveness

| Detection strategy | *Recall* (%) | | 1-*Precision* (%) | |
|---|---|---|---|---|
| | ETD | DTD | ETD | DTD |
| **(1) Effectiveness of main stages in the proposed model** | | | | |
| ACR (case 1) | 98.89 | 96.84 | 43.22 | 62.74 |
| ACR+AI (case 2) | 98.27 | 95.97 | 14.01 | 19.19 |
| ACR+LCF (case 3) | 97.83 | 95.73 | 17.44 | 26.35 |
| **Our model (case 4)** | **97.20** | **95.04** | **7.37** | **10.59** |
| **(2) Comparison with other detection methods** | | | | |
| BD (method 1) | 94.88 | 90.50 | 39.17 | 54.90 |
| ACA (method 2) | 93.45 | 88.95 | 13.11 | 18.63 |

### B. Effectiveness of the Proposed Model

In this subsection, two types of experiments are employed to validate the effectiveness of our model. Particularly, we pay attention to test the effectiveness of modeling the local contextual interaction and integrating it with appearance analysis for false-alarm elimination. In the experiments, we employ $Recall$ and 1-$Precision$ to measure the detection performance. They are defined as

$$Recall = N_{\mathrm{DT}}/N_{\mathrm{TS}} \tag{7}$$

$$1 - Precision = N_{\mathrm{DF}}/(N_{\mathrm{DT}} + N_{\mathrm{DF}}) \tag{8}$$

where $N_{\mathrm{DT}}$ is the number of detected real ships, $N_{\mathrm{DF}}$ is the number of detected false alarms, and $N_{\mathrm{TS}}$ is the manually labeled ship amounts of a certain testing data set.

On one hand, to assess the capabilities of main stages in our model, we perform each stage alone as well as their combination. As can be seen from Table I, all detection cases can obtain high values of $Recall$, even performed over DTD data set. The values of 1-$Precision$ are relatively high when only using ACR stage (case 1). This means that pure bottom-up ACR stage generates many false alarms, although it can achieve high values of $Recall$. Note that 1-$Precision$ decreases obviously when AI or LCF stage is integrated into the detection framework (case 2, 3). The best result is achieved when both LCF and AI are used (case 4). The results indicate that LCF and AI can provide complementary top-down information for discrimination of ship candidates. To summarize, in the proposed detection model, high $Recall$ is provided by ACR stage while low 1-$Precision$ is controlled by the integration of LCF and AI stages.

On the other hand, we compare our model with two typical methods. As described in Section I, the methods of ship detection in optical satellite images can be summarized into two categories. One is primarily driven in a bottom-up manner, and the other is improved by adding top-down cues of appearance analysis. Therefore, we compare our model with the BD from [1] and the recently proposed appearance-characteristic analysis-based method (ACA) from [2]. Since the bottom-up-based method BD has a similar task and image's characteristics with our work, we perform it under the default parameters [1]. Note that, the main aim of comparing ACA with the proposed model is to validate the potential improvements of introducing contextual cues in our multiple top-down analysis. For an operational and fair comparison, the appearance feature extraction of ACA is conducted as the same manner as our model. Moreover, the parameter settings of ACA are tuned to generate as good results as it can.

As can be seen from Table I, our model can obtain higher $Recall$ than other two methods, particularly in DTD data set. This is because ACR stage of our model can detect multisaliency of candidate regions by using a visual attention mechanism, but the statistical-distribution-based BD and the segmentation-based ACA methods cannot. The analysis of 1-$Precision$ is divided into the following two aspects. First, both the proposed model and ACA achieve better performance than BD in all testing data sets obviously. This is because BD, which is performed in a pure bottom-up manner, can hardly remove those false alarms that have similar characteristics with real ships, even in simple scenes. Second, although ACA has introduced some top-down cues of appearance analysis, its effectiveness of discriminating ship candidates is still limited. The results show that, in addition to the appearance features, contextual interactions can provide complementary and important top-down information for the discrimination of ship candidates. Such superiority is more obvious in complex scenes, in which the ship candidates are usually ambiguous.

### C. Computational Efficiency of the Proposed Model

In order to validate the computational efficiency of the proposed model, we have compared the computational time of our model with that of BD and ACA. The experiments are accomplished in Matlab 7.0 with a hardware environment of Intel 2.4 GHz CPU. We record the CPU time for each testing image, and the final computational time is averaged over all results of the testing data sets. Note that, in order to conduct in-depth comparisons, we divide the total computational time into two parts: extraction of ship candidates (*step* 1) and discrimination of ship candidates (*step* 2). In our model, ACR belongs to *step* 1, while both LCF and AI belong to *step* 2.

Table II gives these comparative results. It can be seen that BD, which is performed over the entire scene and equally analyze each region, is much slower than other two methods. By using a hierarchical strategy, ACA costs less computational time than BD. This is because such a hierarchical strategy can rapidly discard the nontarget information and focus on the relevant parts over a scene. Note that our model, which is performed in a multiscale and hierarchical manner, is faster than both BD and ACA. Although our model needs more time in *step* 2 (containing an additional contextual analysis) than ACA, the total time cost of our model is still lower than that of ACA. It is because our model performs in a multiscale manner. Specifically, in *step* 1, our model only needs to highlight the locations of ship candidates and provide approximate ranges. Thus, *step* 1 can be performed over an entire scene with low resolution. In *step* 2, only the selected candidate patches need to be analyzed and discriminated at higher resolution.

TABLE II
COMPARATIVE RESULTS OF COMPUTATIONAL EFFICIENCY

| Method | Analysis strategy | Step 1 (s) | Step 2 (s) | Total Time(s) |
|---|---|---|---|---|
| BD | Over entire scene | / | / | 12.37 |
| ACA | Hierarchical | 2.76 | 2.02 | 4.78 |
| **Our model** | **Multi-scale and hierarchical** | **0.72** | **2.57** | **3.29** |

## V. CONCLUSION

This letter proposed a front-end and operational model for automatic ship detection in optical satellite images. In our model, inspired by the visual search process, a multiscale and hierarchical framework is designed. In particular, a neighborhood similarity-based method was exploited to characterize the local context interactions. Thus, two complementary types of top-down cues (i.e., contextual and appearance information) were introduced to discriminate ship candidates. Experimental results over SPOT5 data validated the detection effectiveness and computational efficiency of the proposed model. Hence, our model is suitable for further real-time applications. Particularly, in addition to the appearance analysis, the capability of eliminating false alarms is significantly enhanced by using the proposed context facilitation.

## REFERENCES

[1] N. Proia and V. Pagé, "Characterization of a bayesian ship detection method in optical satellite images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 226–230, Apr. 2010.
[2] C. Zhu, H. Zhou, R. Wang, and J. Guo, "A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3446–3456, Sep. 2010.
[3] D. Crisp, "The state-of-the-art in ship detection in synthetic aperture radar imagery," Defence Sci. Technol. Org., Melbourne, Australia, 2004.
[4] C. C. Wackerman, K. S. Friedman, W. G. Pichel, P. Clemente-Colon, and X. Li, "Automatic detection of ships in RADARSAT-1 SAR imagery," *Can. J. Remote Sens.*, vol. 27, no. 5, pp. 568–577, 2001.
[5] C. Corbane, L. Najman, E. Pecoul, L. Demagistri, and M. Petit, "A complete processing chain for ship detection using optical satellite imagery," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5837–5854, Jul. 2010.
[6] A. Torralba, A. Oliva, M. S. Castelhano, and J. M. Henderson, "Contextual guidance of eye movements and attention in real-world scenes: The role of global features on object search," *Psychol. Rev.*, vol. 113, no. 4, pp. 766–786, Oct. 2006.
[7] C. Galleguillos and S. Belongie, "Context based object categorization: A critical survey," *Comput. Vis. Image Understand.*, vol. 114, no. 6, pp. 712–722, Jun. 2010.
[8] M. Bar, "Visual objects in context," *Nat. Rev. Neurosci.*, vol. 5, no. 8, pp. 617–629, Aug. 2004.
[9] L. Itti and C. Koch, "Computational modeling of visual attention," *Nat. Rev. Neurosci.*, vol. 2, no. 3, pp. 194–203, Mar. 2001.
[10] Y. Yu, B. Wang, and L. Zhang, "Bottom-up attention: Pulsed PCA transform and pulsed cosine transform," *Cogn. Neurodyn.*, vol. 5, no. 4, pp. 321–332, Nov. 2011.
[11] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
[12] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
[13] C. Hsu, C. Chang, and C. Lin, A Practical Guide to Support Vector Classification, Jul. 2004. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf