Robust Moving Ship Detection Using Context-based Motion Analysis and Occlusion Handling

Xinfeng Bao\textsuperscript{a}, Svitlana Zinger\textsuperscript{a}, Rob Wijnhoven\textsuperscript{b} and Peter H. N. de With\textsuperscript{a}

\textsuperscript{a} SPS-VCA, EE Faculty, Eindhoven Univ. Technol., 5600 MB Eindhoven, the Netherlands
\textsuperscript{b}ViNotion B.V., Horsten 1, 5600 CH Eindhoven, the Netherlands

ABSTRACT

This paper proposes an original moving ship detection approach in video surveillance systems, especially concentrating on occlusion problems among ships and vegetation using context information. Firstly, an over-segmentation is performed to divide and classify by SVM (Support Vector Machine) segments into water or non-water, while exploiting the context that ships move only in water. We assume that the ship motion to be characterized by motion saliency and consistency, such that each ship distinguish itself. Therefore, based on the water context model, non-water segments are merged into regions with motion similarity. Then, moving ships are detected by measuring the motion saliency of those regions. Experiments on real-life surveillance videos prove the accuracy and robustness of the proposed approach. We especially pay attention to testing in the cases of severe occlusions between ships and between ship and vegetation. The proposed algorithm outperforms, in terms of precision and recall, our earlier work and a proposal using SVM-based ship detection.

Keywords: Ship detection, port surveillance, video analysis, context, motion saliency, motion similarity

1. INTRODUCTION

In port areas, condensed waterway traffic of various ships easily creates hazardous situations, which requires reliable traffic management and security control. Large vessels in harbors are typically detected by a radar system. Despite the radar accuracy and efficiency, the major problem of widely used radar systems is their high miss detection rates of small or non-metal ships. Video-based surveillance is newly introduced to port monitoring as a complementary system to radar with the aim to improve the ship detection rate.

Automatic ship detection is a crucial task in port surveillance and is explored in recent research. There are two main strategies that are broadly elaborated in the literature: background registration and appearance modeling. In recent work \cite{1, 2}, research aims at first detecting the horizon line in the frame and separate the ships from the modeled sky or water using image registration. Socek et al. \cite{3} propose a more general approach where they cluster color features through segmentation and feed them to a background registration. Using the modeled background, they detect the foreground. Arshad \textit{et al.} \cite{4} use edge information instead of color and use them in morphological operations based on background modeling. However, all previously mentioned algorithms employ background registration and are unable to handle illumination changes and scintillation in water regions. For the appearance modeling approaches, Sullivan \textit{et al.} \cite{5} build appearance templates by training a set of filters for common ship classes. However, their algorithm tends to miss a target if the appearances of the ship differ from the pre-trained templates. Instead of modeling appearances of the whole ship, Wijnhoven \textit{et al.} \cite{6} construct a cabin detector utilizing local features of representative and characteristic parts of ships. Unfortunately, the simplified and generalized descriptors rely on local features which make ships not distinctive from other textured objects along the coast/canal with similar features. Moreover, the algorithm cannot detect ships without cabins. Although local properties are fundamental in object detection, appearance modeling relying on local features suffers from large limitations in describing the complex and various appearances of ships. Driven by the success of context-based object detection in recent research \cite{7}, our previous work \cite{8} explores a novel framework for detecting ships by utilizing the contextual information and motion of ships. The algorithm works robustly for various types of ships relying on spatial and temporal \textit{co-occurrence consistency} between ships and their surroundings. However, the approach cannot make a distinction between two ships or between a ship and vegetation if one occludes the other.
In this paper, we further exploit the temporal properties of our context-based ship detection by fusing a motion similarity analysis with motion saliency. Firstly, a context model is established based on the results of graph-based segmentation and region-level water detection, assuming ships only travel inside water regions. The modeled context is then used to analyze motion similarity among non-water segments employing SRM (Statistical Region Merging) [9]. Adjacent non-water segments with statistically similar motion are merged into regions. Each of those regions possibly contains a ship. Our assumption is that each ship has a distinguishing region-level motion pattern. Finally, moving ships are detected by performing a motion saliency check for potential ship regions. The main advantage of our context-based motion analysis is that it can detect each individual ship even when there are significant occlusions among ships and vegetation, which commonly occurs in surveillance. It is evident that we want to separate a ship although being visible in an occluded situation, like occlusion with surrounding trees, cranes, or another ship. It will be shown that this separation is possible because we consider the motion speed and direction at a detailed level of segments and jointly moving globally in the same direction with approximately the same speed.

This paper is organized as follows. In Section 2, we present the object-centric context model with analysis of motion similarity and saliency. Section 3 provides the experimental results for real-life port surveillance videos with comparisons. Finally, Section 4 discusses conclusions.

2. CONTEXT-BASED MODELING AND MOTION ANALYSIS

The flowchart of the moving ship detection involves a sequence of processing steps, as depicted in Figure 1. The contextual information is extracted as the basis for the motion similarity and saliency analysis within regions, which will be described in this section.

![Flowchart of the moving ship detection using context and motion analysis.](image)

2.1 Object-centric Context Model

Our dataset contains surveillance videos with typical port scenery. This implies that some spatial relationships and co-occurrences can be expected among ships and their surroundings. Therefore, we focus on building an object-centric context model that provides prior knowledge for the motion analysis. Considering that water is dominant in our scenarios, we perform a water region detection [10] to obtain the contextual information. Let us now shortly describe our water detection approach. A video frame is over-segmented using a graph-based algorithm [11] while preserving the balance between fine and coarse segmentation. We expect the resulting segments to contain meaningful information from only one semantic region. For each segment, a pre-defined amount of pixels is randomly chosen and classified as water or non-water, using a pre-trained SVM. The segments are labeled as water if the amount of pixels classified as water gives a percentage that is above a certain threshold.

After the water detection, we construct our context model based on the following two considerations. We define \( C_{\text{cand}} \) as a region that potentially contains a ship and \( C_{\text{water}} \) as a water region. Firstly, we assume that ships in the image are represented by regions \( C_{\text{cand}} \) which are fully or partially surrounded by water. Secondly, the number of pixels in \( C_{\text{cand}} \) should be reasonably large for visualization and limited by the size of the whole water region. This requirement imposes a scale context in our context model, specified by:

\[
600 < |C_{\text{cand}}| < 0.5 \times \sum_{i=1}^{W} |C_{\text{water}}|,
\]

where \(|C|\) denotes the number of pixels in the corresponding region, and \(W\) is the number of segments labeled as \(C_{\text{water}}\). The lower bound of 600 pixels is empirically set and depends on the video capturing devices.
2.2 Context-based Motion Similarity and Saliency Analysis

Motion Similarity Analysis. In order to avoid over-segmentation from previous step, we group the non-water segments into semantic regions based on our context model by examining the motion similarity. Each ship has a particular region-level motion, which distinguishes it from other ships or non-ship objects. We first calculate the pixel-wise motion using optical flow [12]. Derived from SRM [9], we define a merging predicate \( P(C_i, C_j) \) to determine whether two regions \( C_i \) and \( C_j \) are from the same statistical region \( (i \neq j) \). Instead of using color features as in [9], we use the the criterion for motion features and add an additional constraint on the segments being of the same region type. We use an average flow vector, which is the average of flow vectors of all pixels in a segment, to represent the motion of a segment. Based on the over-segmentation results, we create a motion map by calculating the values of magnitude \( MAG \) and angle \( ANG \) for each average flow vector. The values of \( MAG \) belongs to the set \( \{1, 2, ..., g_{MAG}\} \) and \( ANG \) to \( \{1, 2, ..., g_{ANG}\} \) (here, \( g_{MAG} = 60 \) and \( g_{ANG} = 360 \)). Each segment in the motion map is assumed to be described by a set of distributions. In the motion map, the meaningful regions representing objects should have a common homogeneity property in two ways: (1) in a certain statistical region, each statistical segment has the same expectation in both \( MAG \) and \( ANG \); (2) for two adjacent statistical regions, the expectations differ from each other in either \( MAG \) or \( ANG \) values.

Context in Motion Similarity Analysis. The context model (Section 2.1) imposes the above additional constraint: each statistical region should contain segments with the same label. Suppose \( MAG \) and \( ANG \) are represented by a set of \( Q \) independent random variables and the any possible sum of those variables belong to \( \{1, 2, ..., g\} \), the merging predicate can be defined as:

\[
P(C_i, C_j) = \begin{cases} 
\text{true} & \text{if } \forall k \in \{MAG, ANG\} \text{ it holds that } \frac{(C_j(k) - C_i(k))}{\sqrt{b^2(C_i, k) + b^2(C_j, k)}} \leq \delta \land (L(C_i) = L(C_j)), \\
\text{false} & \text{otherwise,}
\end{cases}
\]

(2)

where \( b(C_i) \) is equal to (index \( j \) may also be used)

\[
b(C_i, k) = g_k \sqrt{1/(2Q|C_i|) \ln(|C_i|/\delta)},
\]

(3)

where \( \delta \) is the probability error, \( L(\cdot) \) is the label of the segment \( (0 = \text{water}, 1 = \text{non-water}) \) and \( C_{|C_i|} \) is the set of regions with \( |C_i| \) pixels. The parameter \( Q \) indicates a user-controlled parameter to guide the level of segmentation merging. Since the contextual information already gives a prediction of the object type (water or non-water), we guide the region merging with a small value for \( Q \), which imposes a strong merging trend on the segmentation results. The scale context also provides a size constraint for the final region size. After the SRM merging, the non-water segments are finally grouped into regions with a particular regional motion, while distinguishing themselves from adjacent non-water regions with a different motion. These merged non-water regions are potential ship regions \( C_{\text{cand}} \). By distinguishing the regional motion, we can separate individual ships.

Motion Saliency Analysis. After the motion similarity analysis, the main challenge is in distinguishing ships from other non-water objects, such as vegetation. We solve this by utilizing another important motion feature of ships: their motion is more significant compared to the surroundings. Therefore, the motion of \( C_{\text{cand}} \) and the motion of the surrounding background are calculated and compared. Since the inner parts of ships are often painted in a uniform color which deteriorates motion estimation, we employ morphological operations to extract the ROIs – the outer part of ship \( C_{\text{ship}} \) and the surrounding local background \( C_{bg} \) [8].

The motion of water can considerably change with time and this causes difficulties in measuring motion saliency. To limit the influence, we use the relative motion in \( C_{\text{ship}} \) and \( C_{bg} \). Using the extracted water region, the relative motions \( rv_{\text{ship}} \) of \( C_{\text{ship}} \) and \( rv_{bg} \) of \( C_{bg} \) are calculated by subtracting the motion of the water region from the motion of corresponding ROIs. To define the motion saliency, we need to analyze the motion contrast between the \( C_{\text{ship}} \) and \( C_{bg} \). For ships, the contrast should be large enough. We calculate the difference between relative motions of \( C_{\text{ship}} \) and \( C_{bg} \) and set a threshold to measure the contrast:

\[
\frac{|rv_{\text{ship}} - rv_{bg}|}{|rv_{\text{ship}}|} > T_1.
\]

(4)
In the equation, the motion difference between the ship and the local background is normalized by $rv_{ship}$ for a robust motion saliency definition, which is independent from the camera zooming factor and the actual speed of the moving ships. Motion contrast itself is inadequate to define motion saliency in the sense that small distracting motion can result in false saliency detections. For example, consider non-ship objects (e.g. floating buoy) in a static water region, even though the motion of the object is not visually salient, the motion contrast can be high enough. Therefore, a constraint should be set for the absolute value of the motion as another measurement defining motion saliency. We compute the difference of the magnitude of motion between the $C_{ship}$ and $C_{bg}$ and the motion of ship should be sufficiently larger than the motion of the non-ship, so that:

$$|rv_{ship}| - |rv_{bg}| > T_2.$$  

(5)

The motion saliency is analyzed for each candidate ship region $C_{cand}$ to detect the moving ships based on the above two motion saliency measurements.

3. EXPERIMENTAL RESULTS

To evaluate the performance of our moving ship detection, especially the quality of handling occlusions, we choose 4 “difficult” video sequences recorded in the harbor of Rotterdam, the Netherlands. All the videos are recorded during day time and have an SD resolution of 720 $\times$ 576 pixels, with the zoom range of 1–35x and with tilting angles of $-45^\circ$–$0^\circ$. In these videos, a large number of occlusions occur among ships and vegetation. Ships are of various types, including container ships, cruise ships and speed boats. The video sequences contain various lighting conditions and different background. In our experiments, we have set $Q = 4$ and thresholds $T_1 = 0.1$ and $T_2 = 0.1$. We also compare the performance of our approach with two recent algorithms: “ship-motion” [8] and “cabin-detector” [6]. Figure 2 shows a visual comparison between the detection results of them.

![Figure 2](image)

Figure 2. Visual comparison of our approach, “ship-motion” and “cabin-detector”: The first column shows the results for our ship detection approach; the second and third columns are the corresponding results of the ship-water method and cabin detector.

When there is occlusion between a cruise ship and vegetation (first row), our approach successfully detects the ship while the “ship-motion” method includes the vegetation area as part of the ship. The “cabin-detector” generates two detections because the local features have a limited ability to describe ships. The second row shows another typical occlusion scenario when two container ships are overlapped. Our approach again detects the two ships individually. Since the “ship-motion” method treats the two ships as one object, the opposite motion in
two ships make the motion of the united region not salient and detection fails. Although the “cabin-detector” correctly finds the two cabins, the algorithm cannot detect the entire bodies of the ships.

Table 1. Ship detection results of our approach, “ship-motion” and “cabin-detector”. TP+FN = manually marked ships, TP+FP = detected ships, TP = correctly detected ships, TOC = total occlusions, SOC = solved occlusion cases.

<table>
<thead>
<tr>
<th>Methods</th>
<th>TP+FN</th>
<th>TP+FP</th>
<th>TP</th>
<th>Precision (%)</th>
<th>Recall (%)</th>
<th>TOC</th>
<th>SOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>our approach</td>
<td>785</td>
<td>757</td>
<td>739</td>
<td>97.6</td>
<td>89.4</td>
<td>203</td>
<td>169</td>
</tr>
<tr>
<td>“ship-motion”</td>
<td>785</td>
<td>656</td>
<td>609</td>
<td>92.8</td>
<td>82.8</td>
<td>203</td>
<td>11</td>
</tr>
<tr>
<td>“cabin-detector”</td>
<td>785</td>
<td>829</td>
<td>627</td>
<td>75.6</td>
<td>79.9</td>
<td>203</td>
<td>194</td>
</tr>
</tbody>
</table>

The numerical results of the 3 methods are shown in Table 1. Our approach outperforms the “ship-motion” in all measurements. Although the “cabin-detector” can separate more ships in case of occlusions, it tends to generate multiple detections for the same ship, which leads to a significantly lower precision. Moreover, the algorithm fails to detect a large amount of ships due to the lack of local features for those ships.

4. CONCLUSIONS

In this paper, we have presented our moving ship detection approach based on context-based analysis of motion similarity and saliency. A context model is first constructed by determining water/non-water for over-segmentation results of the frame and exploring the spatial and size constraints between them. Then, non-water segments are merged into candidate ship regions, which have statistically similar motion. Finally, motion saliency is defined and applied to each candidate region to detect moving ships. In the motion analysis, a context model is employed to filter the candidate ship regions and provide improved robustness for saliency measurements. The experiments show that our algorithm is robust and accurate in detecting various types of moving ships. The main contribution of our approach is that it is able to handle occlusions among ships and vegetation, benefiting from the comprehensive analysis of context and motion similarity/saliency.

REFERENCES