

Discrimination of classes of ships for aided recognition in a coastal environment

Sebastiaan P. van den Broek*, Henri Bouma, Marianne A.C. Degache, Gertjan Burghouts
TNO Defence, Security and Safety, P.O. Box 96864, 2509 JG, The Hague, The Netherlands

*bas.vandenbroek@tno.nl; phone +31 70 3740442; fax +31 70 3740654

ABSTRACT

For naval operations in a coastal environment, detection of boats is not sufficient. When doing surveillance near a supposedly friendly coast, or self protection in a harbor, it is important to find the one object that means harm, among many others that do not. For this, it is necessary to obtain information on the many observed targets, which in this scenario are typically small vessels. Determining the exact type of ship is not enough to declare it a threat. However, in the whole process from (multi-sensor) detection to the decision to act, classification of a ship into a more general class is already of great help, when this information is combined with other data to assist an operator.

We investigated several aspects of the use of electro-optical systems. As for classification, this paper concentrates on discriminating classes of small vessels with different electro-optical systems (visual and infrared) as part of the larger process involving an operator. It addresses both selection of features (based on shape and texture) and ways of using these in a system to assess threats. Results are presented on data recorded in coastal and harbor environments for several small targets.

Keywords: Classification, target recognition, small targets, boats, jet-ski, coastal environment

1. INTRODUCTION

Tasks of naval forces are more frequently taking place in coastal environments. Such environments introduce specific challenges in building an operational picture for situational awareness. Finding ships that are of interest for an operational task (such as intercepting smugglers or the ships that could pose a threat) is a challenging task. Performing those tasks is often done by a human operator, based on radar information and using electro-optical sensors for final identification. The operator may be helped by a system that adds information from these sensors automatically, assisting him in selecting targets for further inspection, and helping in classifying ships as targets of interest.

Aided target recognition is a process providing object information to an operator in order to enable him to perform his tasks better, which means:

1. Earlier: recognize objects of interest at an earlier time by providing information at a larger distance.
2. Better: provide more detailed information or make classification more specific, which allows people with less expertise or specific training to perform the tasks.
3. More: a larger area or more sensors can be used by the same operator, by presenting relevant information, which prevents operators from being overloaded with data and allows them to do their task for longer periods.

In a coastal scenario, an operator will be helped with partial information, as long as not too much is presented, and confidence in the information is high enough. This also means classification does not need to be perfect, as long as the number of false alarms is acceptable. For example, simple features, together with track information, may help to go over many ships more quickly.

TNO is examining several aspects of the use of electro-optical systems in the research programme "Electro-optical sensor systems" for the Dutch Ministry of Defence. The investigation includes ways of adding information in the process to improve time-critical situational awareness and to assist an operator in classifying ships, with an emphasis on small targets in a coastal environment. Previous work describes the overview of a processing chain [18] and the extraction of simple features (e.g., size) and moments for the discrimination between different types of small targets, such as cabin

boats, rhibs and jet-skis [3]. The previous work showed that different boats of only a few pixels high could be distinguished. However, the work also showed that the aspect angle may have a major influence on classification results.

This paper shows an improvement in the classification algorithms, and their effect on the aspect-angle dependence. To improve time-critical situational awareness, discrimination between several classes is now implemented in a near-realtime framework for data acquisition and detection. New local features are examined that describe an observed ship using keypoints, instead of using global features describing the whole ship. Keypoint features describe local shape, changes or texture. Other features that are looked into describe global shape of vessels, and may be used when higher resolution data is obtained.

Prevalence of a class and misclassification-cost can be more important than an improvement of the classification performance of a single class. For example, when other classes have a higher prevalence, their correct classification is more important to avoid an over-shadowing effect of the class under investigation. The effect of using prior probabilities is shortly examined as well.

The structure of this paper is as follows. Section 2 describes the methods. The goal of the methods is to perform accurate and fast classification. Section 3 shows experiments and results to evaluate the performance of different classification methods. The experiments were done on representative data, acquired in a harbor scenario, containing jet-skis. Section 4 discusses the influence of multiple measurements and prior probabilities. Finally, Section 5 presents the conclusion.

2. METHOD

In this section different types of methods (based on different features) are described, at different stages of implementation. In subsection 2.1, the processing chain is described, continuing previous work towards a real-time implementation. Subsection 2.2 shows the work on local descriptors. In subsection 2.3, a few global shape features that are examined qualitatively are described.

2.1 Processing chain

Global descriptors are computed on the whole region of interest that is defined by a detection. These descriptors include normalized central moments, which are scale invariant [3,9]. These descriptors can be used to distinguish between clutter and surface targets, and the descriptors can be used to discriminate different types of targets. Classification is based on a binary decision tree and a mixture of Gaussians from PRTools [5].

A framework was generated that allows us to easily perform the following steps, as shown in figure 1.

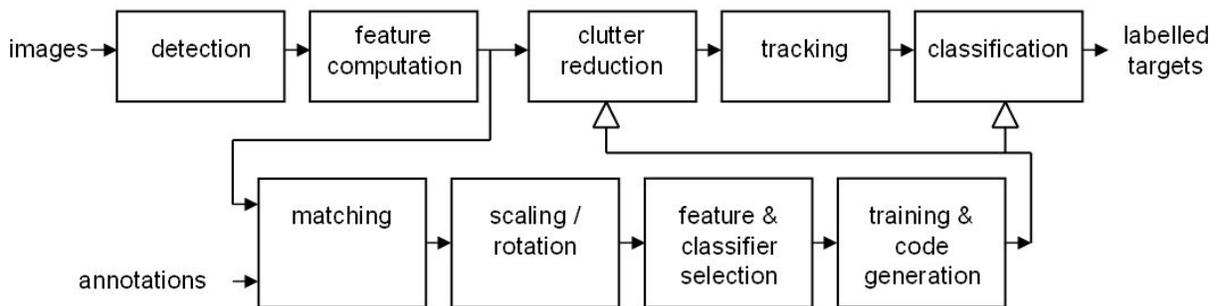


Fig. 1. Schematic overview of the processing chain. The top line is the near-realtime implementation. The bottom shows the offline processing performed to obtain the trained classifier.

The first step is real-time automatic detection of targets in the camera sequence. A detection algorithm provides us with bounding boxes in the image encapsulating surface targets, using robust estimation of the background intensity [1]. Ships are detected at various orientations, sizes and distances. After detection, the ‘global’ descriptors are computed in real-time for each detection.

Ground-truth labeling is performed manually (but aided) in the camera sequence. Assignment of class labels is needed to allow supervised learning. Boats are annotated in key frames and the labels are interpolated for intermediate frames. The labels are matched with the automatic detections to generate labeled detections.

Appropriate scaling and rotation of features is performed, e.g., by principal component analysis. For example, the performance of the nearest-neighbor classifier depends on the scaling between the different features, and the binary-decision-tree classifier depends on the orientation of the features.

Cross validation is used for performance estimation of the classifier. In our case, we used 70% of the data for training and 30% for testing, and we repeated this 20 times.

Feature selection and classifier selection are based on cross validation. Feature selection was performed for each classifier, because an optimal combination for one classifier is not necessarily an optimal set for another classifier. The classifier that performs best with its own optimal features is selected for classification. We used several classifiers from PRTTools [5]. The tree classifier handles the clutter-reduction step well, with the unbalanced class sizes. Mixture-of-Gaussians (MoG) classifier appeared to perform better than others in the discrimination between the different targets.

Finally, C++ code is generated of the trained classifier (the tree classifier and the MoG-classifier), which can be used in a real-time application for clutter reduction and discrimination between different targets.

2.2 Local descriptors

A detection algorithm provides us with bounding boxes in the image encapsulating a ship. For such a region of interest, the image part is extracted. Ships are detected at various sizes and distances. First the image is scaled to a pre-defined number of pixels to achieve scale invariance. The ship image is subsequently described by computing features.

Ships can be seen under arbitrary viewpoints. Hence, to limit the number of examples that are necessary to obtain a good classification, the description of the ship images needs to be robust to small affine transformations. The approach taken in this paper is to describe the ship image as a set of localized features [16]. Each of the features is localized, which makes it more feasible to achieve some robustness at the level of the smaller image patches. Furthermore, the description by multiple localized features makes the approach more robust when only a part of the image is observed. This is beneficial if some parts of the ship have low contrast due to transmission effects, clutter or partial occlusion. From a technical perspective, the local features have the advantage that no full segmentation (i.e. a very complex task) is required.

The ship image is described by a set of features. The goal here is to describe the ship as a collection of particular shapes and textures. To arrive at descriptive points in the image, interest points are determined. Here Harris points are considered [6], which detect prominent corners and edges. An illustration is given in figure 2. The interest points (diamonds) attach to structures of the ship.

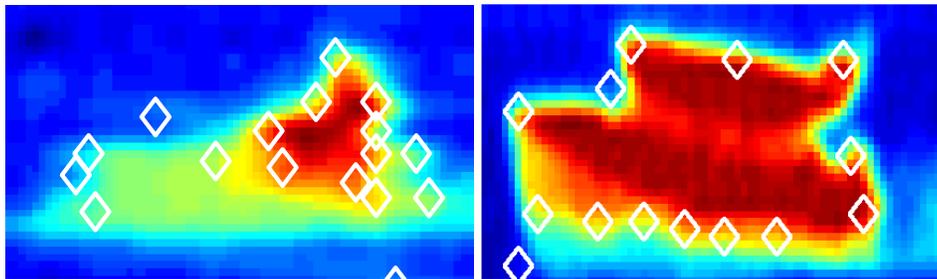


Fig. 2. Jet-ski (left) and water taxi (right) and detected interest points.

In the figure, it becomes very clear that the shapes of the two ships are significantly different. The jet-ski image contains curved contours and a gradient of image intensity towards the central part (the person on the jet-ski). Contrary, the water taxi contains lines that are straight and end up in corner points, with the image intensity being more or less constant. Hence, the aim is to encode the local image contents around the detected interest points, describing corners, edges, curves, etc.

For description, various feature types are considered:

- Moments: statistical features that describe implicitly properties such as symmetry and skew [7]. This is informative of the local intensity distribution. Three types of moments are considered in our setup: Global, central moment: one feature computed from the total image (no interest points are used; only included as a reference to standard performance). Local, central moment: the same feature as above, but centered around the interest points. Local, Hu-moments: the same as above, but with additional invariant properties, such as rotation invariance.
- Gaussian image derivatives: filter-based gradients of the image, indicative of edges and curves [11].
- Gabor texture measurements: Gabor filters perform a local Fourier analysis, hence indicative of local texture [2].
- Statistics: for four orthogonal orientations (left, right, up, down) the numerical matrix gradients from the image part are collected. Inspired by SIFT [15] but a simpler feature hence computationally more efficient. For each orientation, corresponding gradients are extracted which are weighed by their gradient magnitude. Hence, only statistics of strong edges are gathered. For instance, if an image part contains a white bar above a dark bar, the magnitudes will be predominantly oriented upwards. A 4-valued feature vector of orientations in a local neighborhood is thus obtained; it can be considered as a simplified SIFT-feature.

Note that the features describe the image contents differently. Hence they can be considered complementary to some extent. This is illustrated in figure 3. The figure shows on the left the jet-ski in gray values, and detected interest points in various colors. For each of the colored points, the following features have been computed: moments, Gaussian derivatives and Gabor measurements. These features are shown in the right panel of Figure 2. In this feature panel, four blocks are shown. The left block is the original image patch, whereas the consecutive three blocks depict moments, gauss and Gabor features. The feature values are significantly different for each patch, indicating complementary qualities.

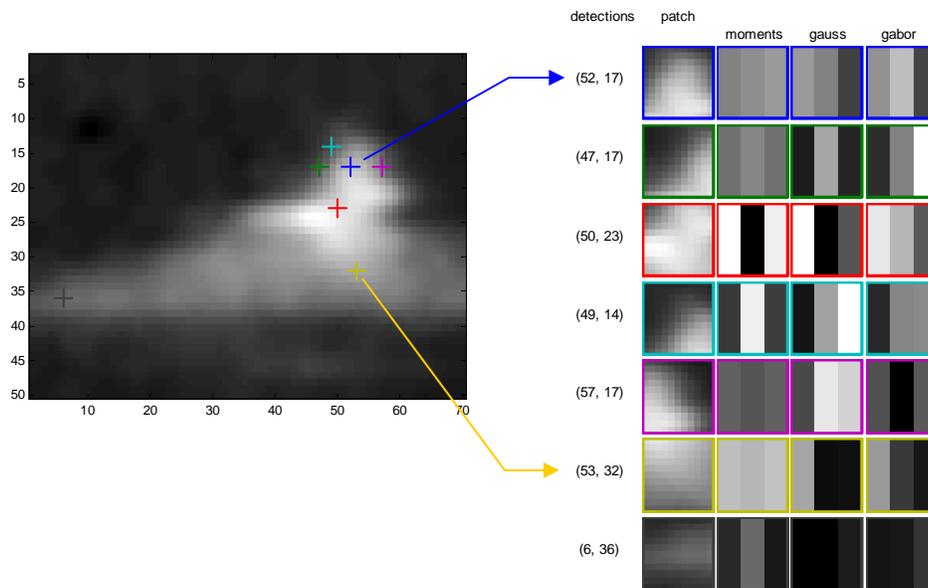


Fig. 3. Illustration of different features (right panel) for various interest points (left). The plus-signs on the jet-ski in the left panel are the interest points. For each of these, we obtain a description by a local feature, as shown in the right panel. This panel depicts from left to right: the image patch, its description by a local central moment, Gaussian derivatives, and Gabor filters. The different feature values indicate that these features are complementary, which opens up the way for combinations of the features e.g. by fusion.

To compare two ship images to each other, a similarity measure is needed from one set of features to another. In our application, the number of interest points may vary and thus the number of features. Hence, the similarity function should be able to compare two sets of features that are not of the same size. The Earth Mover's Distance (EMD) is able to express dissimilarity between two sets of different sizes, based on individual feature-to-feature dissimilarities [17]. The EMD dissimilarity is the result of a minimization of feature-to-feature assignment, resulting in larger dissimilarities if the features from the two sets are less similar. This approach has been shown to be very discriminative in conjunction with feature vectors, see for example [14].

Before individual features are compared to each other, their statistics are taken into account. Feature values may have a very different range, and may also correlate significantly, hence the feature values are whitened to obtain similar ranges and to achieve decorrelation. The whitened feature values are then used as input to the EMD comparison module.

Since the considered features are complementary to some extent, a data fusion module is incorporated. Two fusion schemes are considered, early and late fusion. Early fusion: combination of feature vectors at the lowest level by simply concatenating them. Late fusion: combination of EMD distances, hence the fusion takes place at a high level. The fused features are also evaluated in the experiments.

A ship image is classified by comparing its features to a database of features. The database is built from example images; these are referred to as the training images. The training images are encoded into feature values and put into the database before the algorithm operates. A new ship image is classified by assigning it to the label of the most similar feature set. This effectuates a nearest-neighbor classifier with an EMD-based distance metric.

2.3 Global shape descriptors

Other descriptors related to global shape were examined, especially to discriminate (larger) boats that are seen in more detail (see also [19]). For this, both real data (from infrared recordings) as well as simulated data is being used. The simulated infrared imagery is created using the EOSTAR software developed at TNO [12] that generates realistic images from 3D models and signatures, taking atmospheric and camera effects into account.

From these images, segmentations are obtained using the same detection algorithms used in the other experiments. In previous work [3] it was shown that using descriptions of contours obtained from these segmentations is hampered by the fact that they are strongly depending on aspect angle. On the other hand, it was shown that most dependence is near head on views. Figure 4 shows examples of silhouettes of a tanker (from real imagery) and a frigate (simulated).

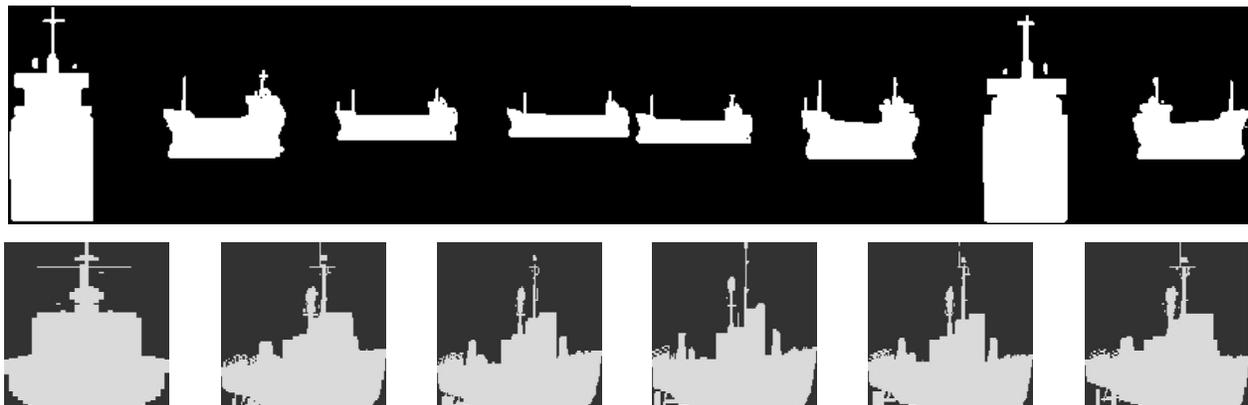


Fig. 4. Silhouettes of a tanker (top, from real imagery) and a frigate (bottom, simulated) at different aspect angles. Most changes occur near head on view.

One feature that was examined is ‘compactness’:

$$c = \frac{\sum_{i < n, j < m} \{p_{i,j}\}}{m * n}$$

where p_{ij} is the pixel intensity in the masked image. Figure 5 shows the compactness for three different types of ships (from simulated data) at aspect angles from 0 (head on) to 340 degrees. This shows a large dependence on aspect angle, but also clear differences between ships, especially in side-views. This makes compactness a good feature in addition to other more complex features, and may even help indicate (range of) observed aspect angles to assist in classification by limiting the possible matches.

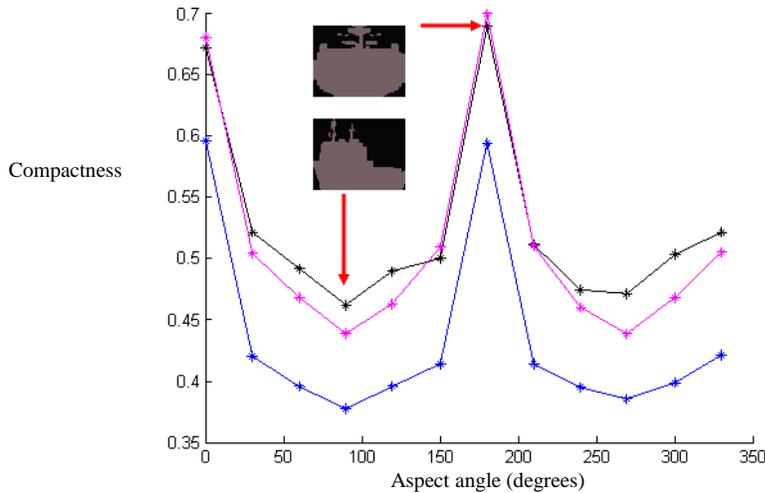


Fig. 5. Compactness of silhouettes of three simulated military ships at different aspect angles. There are large differences between ships in side views, and between side and head on views.

Another description of the shape is by finding parts defining a ship. By making a histogram of the heights of the segmented image, often the hull of the ship may be found, using the most common height. Structures on top of this can then be determined, to define the ship as different parts, that can be matched with a description in a database. Figure 6 shows an example of a height histogram for the tanker, where the histogram shows groups indicating hull, superstructure and masts.

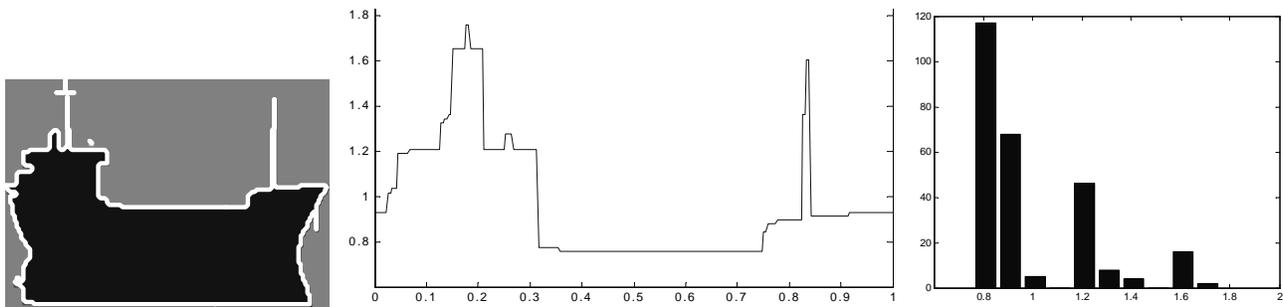


Fig. 6. The middle graph shows the height profile of the silhouette of the ship shown on the left. On the right the histogram of height values of this profile is shown. Groups in the histogram indicate hull, superstructure and masts.

3. RESULTS

For the experiments, data was used recorded in Rotterdam [10] using a long-wave infrared camera. Three types of ships are used in a harbor scenario: a water taxi (similar to a cabin boat), a rhib and a jet-ski. The three chosen classes are relevant types, which are different in operational context, but not too dissimilar in size and appearance to make discriminating them too trivial. Although the used data set is small, care is taken to use different training and testing sets. Even when those are similar in some aspects, general results do indicate the usefulness of methods, and especially dependences that are still observable are relevant.

3.1 Automatic detections and global descriptors

In the first experiment, we have classified ships using different global features. The classification procedure is repeated 20 times, with random selection of images to train and to test on. Obviously, the train and test images are disjunct in each repetition.

We used 44,266 automatically generated detections for training and testing, which consisted of 1049 labeled targets: 370 water taxis, 292 rhibs and 387 jet skis. The other 43,217 detections contain clutter -- such as waves -- and other boats. The three small surface targets were observed at different orientations (ranging from 0 to 360 degrees) and at different distances (ranging from 0.5 to 1.5 kilometers).

An initial clutter reduction step based on a bounding box in the feature space that was generated by the tree classifier using a feature based on the relative intensity resulted in a reduction of the number of false detections by 37,606 (from 43,217 to 5611).

Classification is based on the Multiple-of-Gaussians (MoG) classifier. Feature selection showed that an optimal separation of the classes is obtained using five features based on normalized central moments. The receiver-operator characteristic (ROC) curve is shown in figure 7, and the confusion matrix is shown in table 1. The matrix shows that 87% of the clutter is removed by the MoG classifier, reducing the number of false alarms to 741. Note that water taxis and rhibs are often confused (9% of the water taxis and 10% of the rhibs are classified as the other). Furthermore, clutter is often classified as a jet-ski (9%). Since the prior probability of clutter is much larger than that of jet skis, most detections that are classified as jet-ski, are caused by clutter.

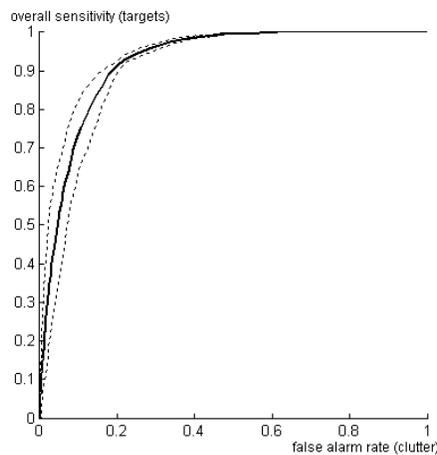


Fig. 7. ROC curve based on MoG classification based on four classes. The solid curves indicate the average performance, and the dashed curves indicate the confidence band.

Table 1. Confusion matrix based on MoG-classification (4 classes)

	Clutter (predicted)	Water taxi (predicted)	Rhib (predicted)	Jet ski (predicted)
Clutter (actual)	0.8679	0.0305	0.0121	0.0895
Water taxi (actual)	0.0351	0.8297	0.0892	0.0459
Rhib (actual)	0.0034	0.0959	0.8767	0.0240
Jet ski (actual)	0.0543	0.0052	0.0026	0.9380

Figure 8 shows a frame of a sequence that was processed using the whole processing chain, including detection, clutter reduction, tracking and the classification. Labels indicating the predicted classes are correct for most of the frames.



Fig. 8. Frame from a sequence processed by the processing chain described in section 2.1, including classification based on global descriptors. Color labels indicate classified type. Water taxi (red, left), jet-ski (blue, middle) and rhib (green, right) are all classified correctly.

3.2 Local descriptors

In the experiments, ships are classified using different features and their fused combinations (both early and late fusion). The classification procedure is repeated 20 times, with random selection of images to train and to test on. Obviously, the train and test images are disjunct in each repetition. For each class, five train and five test images are collected in this manner. These images have been taken from recordings in the harbor of Rotterdam, see for examples figures 2 and 8. The ship types under consideration are jet-skis, rhibs and water taxis, the latter being very common in the harbor environment. The ship orientations under consideration are $\{0, 15, 30, 45, 60\}$ degrees. For clarity, for each orientation ten images were collected.

The scale of the Harris interest point detector is varied over $\{0.8, 1.0, 1.4\}$ pixels. The size of the image part around an interest point (which is to be described) is varied over $\{4, 8, 12, 16, 20\}$ pixels. This will be referred to as the ‘feature size’ (sometimes in literature also called support region). It will be mentioned which parameter settings are optimal, and how well the ship images are classified using these settings.

Figure 9 shows the results for various feature sizes, where the depicted classification rates are averaged over ship orientations from 0 to 60 degrees (where 0 is a side view). It becomes clear that there is an optimum around 16 pixels. This implies that for the size of the image patches there is a trade-off. Smaller patches have a disadvantage. They have less discriminative power, because they encode a small neighborhood, e.g. an edge. Larger patches encode a larger area with the advantage that it is more descriptive. This is demonstrated in figure 9, where the performance increases from 4 to 16 pixels. The disadvantage of very large patches is that the appearance changes significantly if the ship orientation changes. This effect causes the trade-off between the larger sizes of the image patches and the robustness. This causes the classification performance to be optimal for an intermediate patch size. The conclusion from this experiment is that 16 pixels is a good image patch size, hence this parameter is fixed during the subsequent experiments.

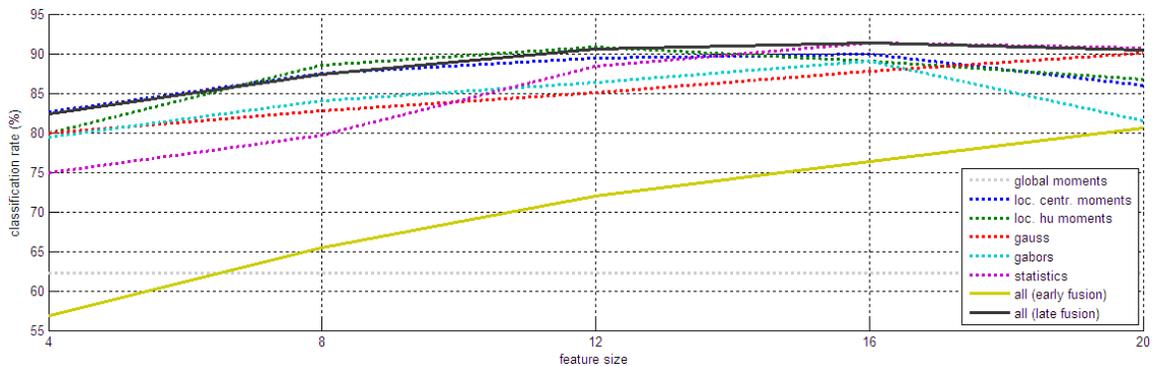


Fig. 9. Performance of ship classification for various sizes of the image patches (averaged over orientations of 0 to 60 degrees). There is an optimum around a patch size of 16 pixels. Dotted lines are specific features. Solid lines indicate fusion, with late fusion (top) giving better results.

The solid lines in figures 9 and 10 show the classification performance when all features are fused. Two fusion schemes have been tested: early and late fusion. Early fusion performs worse than each of the features individually. The conclusion is that the whitening fails, because the combination of different features does not constitute a multidimensional Gaussian distribution. The late fusion performs best on average, that is, for various sizes of the image patches. Figure 10 shows the performance across various ship orientations. This is an interesting experiment, because it evaluates to which extent the features are robust. For instance, the global image feature, i.e. global moments, is not robust. This can be expected: one feature to encode a total image changes significantly when the ship changes orientation. Individual features perform well (>90%) either when the ship orientation changes up to 45 degrees (e.g. Gaussian features), or when the change is large (statistics). Clearly, better performance is achieved by the local features compared to the global moments. Late fusion is the most robust feature on average, that is, across all tested ship orientations.

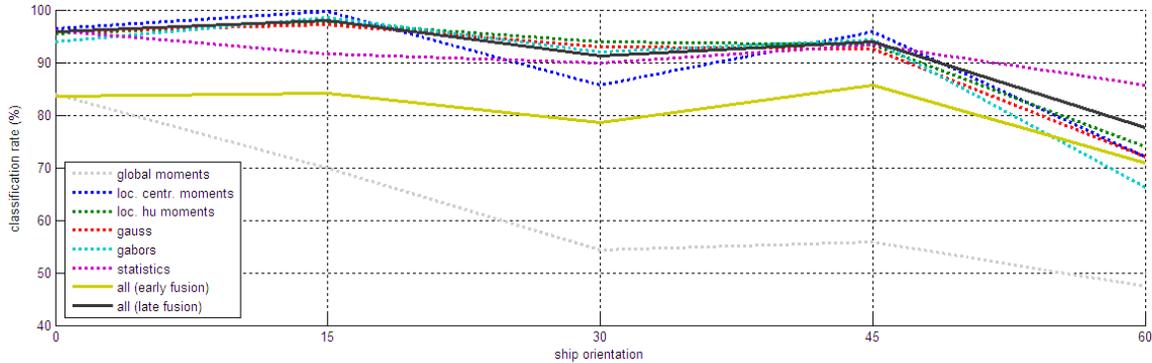


Fig. 10. Performance of classification for various ship orientations. Individual features (colored dotted lines) perform well (>90%) when the ship orientation changes up to 45 degrees from side view (e.g., Gaussian features)

The computation time of the algorithm with local features has been considered. Here, a distinction is made between computational costs for the image processing part on the one hand, and time needed for classification on the other. For the image processing part of our algorithm, the statistics feature is most expensive, for image patches of 16 pixels 1.4 seconds. This means that each 3 seconds, 2 images can be computed. For local centralized moments, Gabor and Gauss features the computation time is 0.6 seconds, or per 3 seconds 5 images. This means a speed-up by a factor of 2.5. If computation time is important, the Gauss features may be preferred, as they are computationally efficient and achieve reasonable classification performance. With respect to the time needed to classify an image, this is obviously very dependent on the chosen representation. The non-fused features are cheap to classify, below 40 ms. per image. For fused features the classification time increases significantly, up to 170 ms. for late fusion (increase of classification time by a factor of 4). Again, if time to classify images is of importance, Gauss features are to be preferred. As given processing times are for a not optimized implementation, significant improvements can still be expected.

In conclusion, for the task of ship classification various features have been evaluated for the representation of ship images. Evaluation has been based on classification accuracy and computational efficiency. If ships are expected to be seen at orientations of 0 – 45 degrees, Gaussian features are to be preferred. They are cheap and perform well. If more training examples are included in our database, say of 0, 45 and 90 degrees, the Gaussian features are expected to work well. That is, a flip-invariant representation has been considered, so only one quadrant of orientations has to be trained (0 and 180 degrees are equal, so are 45, 135, 225 and 315 degrees). If less learning images are included in the database, the late fusion algorithm, which combines all features in an appropriate weighing scheme, is to be preferred as it is most robust. However, it is an expensive method. If computation time is an issue, Gaussian features are to be used. Most time of classifying one image is spend on computing the image features. Gaussian features are cheap and provide the most significant speed-up.

4. DISCUSSION ON PROBABILITIES AND MULTIPLE MEASUREMENTS

Often when classification results are presented, focus is on the probability a class is classified correctly. When one method gives a better percentage, this is seen as a relevant improvement. It is however important to note, that prior

probabilities (i.e., how often this and other classes are observed) and cost (how important it is not to miss this class) are sometimes more relevant. When other classes are more prevalent, even when they are only rarely misclassified as the most important class, they may cause the majority of predicted classifications of this class. A limited improvement in the classification of a rare type (e.g., from 80% to 90%) does not improve the overall result much. However, a slight reduction of misclassification of a common type may have major impact.

This does not mean such results for limited data sets and limited sets of classes are not useable. They do indicate features and methods that are useful in discriminating between classes. When applying such methods for an operational use, care should be taken to define the correct classes, and spend effort in correctly classifying the more common ones, as well as the ones with highest cost when they are missed. Bayes' formula shows the influence of prevalence of classes:

$$P(A_i | O_j) = \frac{P(O_j | A_i) \cdot P(A_i)}{P(O_j)}$$

where $P(A_i)$ is the prior probability of the actual class A_i and $P(O_j)$ is the prior probability of the predicted class O_j , $P(O_j | A_i)$ the conditional probability of predicted class O_j given actual class A_i , and $P(A_i | O_j)$ is the posterior probability of the actual class A_i given the predicted class O_j . $P(O_j)$ can be determined from:

$$P(O_j) = \sum_i P(O_j | A_i) \cdot P(A_i)$$

Thus, the effect of improving classification of actual class A_i given by $P(O_j | A_i)$, is weighed by the prevalence of this class, $P(A_i)$.

Some simulations were done for a situation with three classes and a rest class. A confusion matrix – which contains the conditional probabilities $P(O_j | A_i)$ – defines how well a class is recognized as such (diagonal values) and how classes are misclassified as others (off-diagonal). Together with the prior probability $P(A_i)$, this allows to compute posterior probabilities for all classes given an observation.

The prior probability describes what is known about a class without an observation. The posterior probability is taking the measurement into account, so after one measurement the probability of an object being of a given class changes from the prior probability to the posterior probability. When doing another measurement – assuming they are uncorrelated – the posterior probability obtained after the last measurement can be used as prior probability for the next, giving a new probability, and so on. More detailed descriptions of methods can be found in [8] for a “possibilistic” approach to use these probabilities when prior knowledge is inaccurate.

Some simple simulations were done using example probabilities, to see the effect of not taking all information into account. Three target classes are assumed to be trained for classification, with a fourth class containing all other ships. A correct classification of 90% is assumed for the three classes, and some misclassification among the three. The fourth class is always misclassified as one of the other three, resulting in the following confusion matrix:

Table 2: Confusion matrix with probabilities for a simple simulation.

	Class 1 (predicted)	Class 2 (predicted)	Class 3 (predicted)	Rest (predicted)
Class 1 (actual)	0.9	0.01	0.09	0
Class 2 (actual)	0.01	0.9	0.09	0
Class 3 (actual)	0.05	0.05	0.9	0
Rest (actual)	0.333	0.333	0.333	0

The prior probabilities for the four classes are chosen such that the fourth class occurs in 85% of the cases and the other three classes occur in 5% of the cases. These values can now be used to compute the posterior probability of an object being any of the classes, given an observation, using Bayes formula. For example, when observing an object as the first class, the posterior probabilities are 0.136, 0.002, 0.008 and 0.855 for the four classes respectively. The value 0.136 is obtained using

$$P(A_1 | O_1) = \frac{P(O_1 | A_1) \cdot P(A_1)}{P(O_1)} = \frac{0.9 \cdot 0.05}{(0.9, 0.01, 0.05, 0.333) \cdot (0.05, 0.05, 0.05, 0.85)^T} = 0.136$$

where in general $P(O_j | A_i)$ is found in the j^{th} column of table 2, on the i^{th} row, and $P(O_j)$ is the inner product of the j^{th} column with the prior probabilities. This means a single observation of class 1 only gives a 13.6% posterior probability of it actually being of this class. It is most probably a member of the fourth (rest) class.

Suppose this is used in a scenario with ten ships, one of each in the classes 1, 2 and 3, and seven in the remaining class. Using the probabilities defined above, a first measurement (applying the probabilities in the confusion matrix as a prediction without prior) might be: three of class 1, three of class 2, four of class 3 and none of the remaining class. However, when using the computed posterior probability for this same situation (with priors), and choosing the class with highest probability for each classification, the following is obtained as most probable situation: none of the classes 1, 2 and 3 and ten of the remaining class. This shows that no useful classification can be done from a single measurement. However, more (uncorrelated) measurements can be combined, by updating the probabilities with each measurement. As repeated indications of an object being of class 1 makes it continuously more improbable that it is class 4, at some point, class 1 will be the only probable class. If the predictions would be correct each time, after six measurements, the most probable classifications match the actual situation.

In practice, measurements in an image sequence (even when taken at some interval) will not be uncorrelated, due to aspect angles being similar and other influences such as weather not changing. When combining measurements, this can be taken into account by not updating the probabilities completely, but using some update factor to slowly change them based on repeated measurements, as described in [4].

It is important to note that a major improvement of the results can be obtained by improving the classification of the remaining class. If this is not possible, combining measurements with prior information means that even with imperfect classification still useful results can be obtained.

5. CONCLUSION

In this paper, results were presented on ongoing work on discriminating classes of small vessels recorded with electro-optical systems. A method using global descriptors of segmented objects is implemented in software that combines data-acquisition, detection, clutter reduction, feature extraction and tracking with this classification, resulting in tracks labeled as one of several classes of small vessels. This implementation allows near-realtime processing of data, and is applied to real infrared recordings of small vessels such as rhib, jet-ski and water taxi (cabin boat).

Various local descriptors (based on keypoints) have been evaluated for the representations of ships. Evaluation was based on classification accuracy and computational efficiency. If ships are expected to be seen at orientations within 45 degrees of side views, Gaussian features are to be preferred. They are cheap in processing demand and perform well. When more learning examples are included in our database, the Gaussian features work well. If less training images are included in the database, the late fusion algorithm, which combines all features in an appropriate weighing scheme, is to be preferred as it is most robust. However, it is a more expensive method. Most time of classifying one image is spend on computing the image features. Gaussian features are cheap and provide the most significant speed-up.

A single measurement does not allow declaring an object to be of a certain class, even with high classification probability for that class, if prevalence of different classes is not taken into account. To make classification useful for an operator, false positives need to be reduced by improving correct classification of the more common classes. Another way is to combine several observations in time, reducing the probability of a classification as being a false positive.

6. ACKNOWLEDGEMENTS

The work for this paper was supported by the Netherlands MoD under the programme V602 "Electro-Optical sensor systems".

REFERENCES

- [1] Bouma, H., Lange, D.J.J. de, Broek, S.P. van den, Kemp, R.A.W. and Schwering, P.B.W., "Automatic detection of small surface targets with electro-optical sensors in a harbor environment," Proc. SPIE 7114, 711402 (2008)
- [2] Bovik, A.C., Clark, M. and Geisler, W.S., "Multichannel texture analysis using localized spatial filters," IEEE Transactions on Pattern Analysis and Machine Intelligence 12, 55-73 (1990)
- [3] Broek, S.P. van den, Bouma, H. and Degache, M.A.C., "Discriminating small extended targets at sea from clutter and other classes of boats in infrared and visual light imagery," Proc. SPIE 6969, 69690B (2008)
- [4] Broek, A.C. van den, Broek, S.P. van den, Heuvel, J.C. van den, Schwering, P.B.W. and Heijningen, A.W.P. van, "A Multi-sensor scenario for coastal surveillance," Proc. SPIE 6736, 67360P (2007)
- [5] Duin, R.P.W., Juszczak, P., Paclik, P., Pekalska, E., de Ridder, D. and Tax, D.M.J., "PRTools4, A Matlab Toolbox for Pattern Recognition," <http://prtools.org>, Delft University of Technology (2004)
- [6] Harris, C. and Stephens, M., "A combined corner and edge detector," Proceedings of the 4th Alvey Vision Conference Manchester, 189-192 (1988)
- [7] Hu, M.K., "Visual pattern recognition by moment invariants," IRE Transactions on Information Theory 8, 179-187 (1962)
- [8] Huizing, A.G. and Groen, F.C.A., "A possibilistic approach to target classification," in: [Fuzzy systems and soft Computing in nuclear engineering], edited by D. Ruan, Springer Verlag (1999)
- [9] Hupkens, T.M., "Invariant Legendre and Zernike Moments," Recent research developments in pattern recognition 3, 35-52 (2002)
- [10] Kemp, R.A.W, de Groot, J.F., Broek, S.P. van den, Lange, D.J.J. de, Dijk, J. and Schwering, P.B.W., "Results of Optical Detection Trials in Harbour Environment," SPIE Proc. 6943, 69430Y (2008)
- [11] Koenderink, J.J., "The structure of images," Biological Cybernetics 50, 363-370 (1984)
- [12] Kunz, G.J., Degache, M.A.C., Moerman, M.M., Eijk, A.M.J. van, Neele, F.P., Doss-Hammel S.M. and Tsintikidis, D., "Status and Developments in EOSTAR, a model to predict IR sensor performance in the marine environment," Proc. SPIE 5572, 101-111 (2004)
- [13] Lauberts, A., Karlsson, M., Näsström, F. and Aljasmí, R., "Ground Target Classification Using Combined Radar and IR with Simulated Data," Proc. of the 7th international Conference on Information Fusion 2, 1088-1095 (1998)
- [14] Lazebnik, S., Schmid, C., and Ponce, J., "A sparse texture representation using local regions," IEEE Transactions on Pattern Analysis & Machine Intelligence 27(8), 1265-1278 (2005)
- [15] Lowe, D.G., "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision 60(2), 91-110 (2004)
- [16] Mikolajczyk, K. and Schmid, C., "A performance evaluation of local descriptors," IEEE Trans. Pattern Analysis and Machine Intell. 27(10), 1615-1630 (2005)
- [17] Rubner, Y., Tomasi, C. and Guibas, L.J., "The earth mover's distance as a metric for image retrieval," Int. J. Computer Vision 40(2), 99-121 (2000)
- [18] Schwering, P.B.W, Broek, S.P. van den and Iersel, M. van, "EO System Concepts in the Littoral," Proc. SPIE 6542, 654230 (2007)
- [19] Withagen, P., Schutte, K., Vossepoel, A.M. and Breuers, M., "Automatic classification of ships from infrared (FLIR) images," Proc. SPIE 3720, 180-187 (1999)