Ship recognition for improved persistent tracking with descriptor localization and compact representations

Sebastiaan P. van den Broek [*], Henri Bouma, Richard J.M. den Hollander, Henny E.T. Veerman, Koen W. Benoist, Piet B.W. Schwering

TNO, Oude Waalsdorperweg 63, 2597 AK The Hague, The Netherlands

ABSTRACT

For maritime situational awareness, it is important to identify currently observed ships as earlier encounters. For example, past location and behavior analysis are useful to determine whether a ship is of interest in case of piracy and smuggling. It is beneficial to verify this with cameras at a distance, to avoid the costs of bringing an own asset closer to the ship. The focus of this paper is on ship recognition from electro-optical imagery. The main contribution is an analysis of the effect of using the combination of descriptor localization and compact representations. An evaluation is performed to assess the usefulness in persistent tracking, especially for larger intervals (i.e. re-identification of ships). From the evaluation on recordings of imagery, it is estimated how well the system discriminates between different ships.

Keywords: track recognition, persistent tracking, maritime situational awareness, infrared, Fisher vector, SIFT.

1. INTRODUCTION

Recognition of a ship can make it easier to determine if it is of interest in naval operations, such as piracy prevention, by using historical information and deriving a long term description of its behavior. For example, knowing whether a skiff has been in a certain area for a longer time, may make it possible to distinguish a fisher boat from a possible pirate, especially if at an earlier encounter it was determined that it was of no interest. When intent of ships can be predicted, it may help planning the use of assets such as UAVs to more efficiently cover a (larger) surveillance area.

This paper describes research on determining the effective use of features obtained from electro-optical systems, in recognition of small ships. In earlier work, the scale-invariant feature transform (SIFT) descriptors in infrared recordings were found to be useful, even with low resolution imagery of ships. The use of localization of key-points in comparing images caused an improvement in recognition, and Fisher vectors were introduced as a way of presenting the information of a varying number of key-point descriptors in a fixed length vector [10]. In this paper, we examine the addition of descriptor location into the Fisher vector representation. For this new evaluation, we compare results to our previous methods, to a bag-of-words approach – a more common way of representing a varying number of key points – and to a human observer. Evaluation is done on a subset of the earlier used dataset of infrared images, where unreliable annotations and low quality images were removed. The usefulness in persistent tracking is illustrated using tracking results on simulated ship movements.

The outline of the paper is as follows. The different recognition methods are presented in Section 2. The experiments and results are shown in Section 3. Results are discussed in Section 4 and conclusions and recommendations are summarized in Section 5.

2. RECOGNITION METHOD

The recognition method consists of the following steps. First, detection and segmentation is performed to localize the targets (Sec. 2.1). Secondly, the detections are preprocessed, which includes intensity stretching, centering and scaling (Sec. 2.2). The main part of the recognition method are the matching algorithms. Matching is performed of the detected

---

[*] bas.vandenbroek@tno.nl; phone +31 888 66 4086; http://www.tno.nl

target to a database of images of the different ships to identify the most similar type. Section 2.3 describes different matching algorithms, including those of a previous paper [10], a bag-of-words approach and an exploration to combine descriptor location into the Fisher vector description. Section 2.4 describes the ranking of results, and finally, Section 2.5 describes the simulation environment in which the effect of recognition is examined.

2.1 Detection and segmentation

In order to automatically extract ships from recorded imagery, we use the detection method described in [6] to separate foreground objects (e.g., ships) from background (sea, sky). In this method, the background is assumed to have properties that are constant per image line (corrected for rotation) and to vary only smoothly from line to line. This assumption is sufficiently accurate for various levels of sea state, in open ocean as well as in bay environments, as known from infrared statistical background analysis work by Schwering [23]. Pixels that are statistically unlikely to be background are clustered in a segmentation of possible ships. This results in a binary mask and bounding boxes around the target. This mask is used for computation of central moments and to select a window for the other description methods.

2.2 Preprocessing

The ships are extracted from the images based on automatic segmentation but with a manual check on completeness. The extracted images have 16-bit pixel depth but contain a much smaller range of intensity values. Therefore, the images are enhanced by histogram stretching using a 0,5% value at both the upper and lower range. After stretching the images are converted to 8-bit per pixel. As there is some spatial variation in the way the ship ROIs are extracted from the recordings, the alignment of the ships in the extracted images will not be precise. In order to compensate for this, more detection mask properties (e.g., center of mass) are used to improve the extraction of the ships. Since there are several possibilities for the extraction/cropping of the images and this affects feature computation and localization, multiple images will be extracted per ship to evaluate the influence on the recognition result. We used three types of extraction. Besides the original extracted image (type A), there are two versions that are centered at the mask's center of mass. Type B, contains only valid image/mask pixels and is cropped at the shortest sides of the mask. Type C, contains all of the mask area and is appended by pixel replication in order to retain the desired image center. All three versions of the extracted images are scaled to have a fixed width of 300 pixels. Scaling of the images is performed by means of bicubic interpolation. Examples of the extracted versions for a ship are shown in Figure 1.

2.3 Matching methods

We compare the newly proposed Fisher vector method to other approaches, some of which were also used in our previous paper [10]. Here all methods are shortly described.

Central moments

The first type of matching is based on normalized moments for the binary mask of detected pixels. These moments are statistical features that describe properties such as symmetry and skew. The normalized central moments [8][9] are scale invariant and computed around the center of mass of the binary mask. Six values are used, obtained from the nine combinations of different orders (0 to 2) for x and y, discarding the three values that are constant due to normalization. Comparison between images is based on a statistical distance between these values, computed as a Euclidean distance of the 6 value vector, where each value is divided by the standard deviations over all images.

SIFT features (with localization)

The second type of matching is based on the scale-invariant feature transform SIFT [17][18] and keypoint localization [19]. SIFT keypoints are a commonly used feature in image retrieval, where the match criterion is generally based on the number of keypoint matches. This generally ignores the fact that the spatial distribution of the keypoints over an object should be similar as well.

Type A: The original extracted image and the initial detection mask

Type B: Centered + Minimal dimensions

Type C: Centered + Maximal dimensions

Figure 1: The ship image is extracted in different ways (denoted Type A, B and C) in order to evaluate its effect on recognition. Type B is centered at the center of mass of the mask (red dot) and its dimensions are limited by the minimum width/height that is contained in the image (green rectangle). Type C is also centered but its dimensions are limited by the maximum width/height that is contained in the image. The shorter sides (corresponding to the gray area in the mask) are padded by pixel replication.

The relative location can be computed by comparing locations relative to the center-of-mass of the thresholded detection image (see description in Section 2.2), scaled by the width (to compensate for slight differences in aspect angle) and height of the detection (to compensate for observation distance). We use a somewhat different dissimilarity measure than in [19], namely:

$$1 - \frac{\#\text{matches}}{\#\text{keypoints}} \tag{1}$$

where only the number of keypoints in the observation image is used in the denominator. This dissimilarity lies in the range [0,1] since matching is performed for each of the 'observed' keypoints so there will be at maximum this number of matches. When using the keypoint location, this changes the number of matches, as matches are discarded if the location is different.

Bag of Words

Instead of a per-feature comparison as in the SIFT-matching method, there is also the possibility to create an aggregate description of all SIFT keys for the image. A popular method for making such an aggregate description is the Bag Of Words (BOW) approach [24]. Here the keypoint features are first clustered, creating a dictionary of 'Words'. All keypoints are subsequently assigned individually to one of these clusters (Words), producing a histogram of clusters per image. The clustering is performed by K-means clustering using a number of cluster centers and a Euclidean distance measure on the 128-element SIFT keys. After assigning all SIFT keys to the closest cluster, the (Euclidean) distances between the BOWs (histograms) are determined and used for training an SVM-classifier for each ship type. For recognition, SIFT keypoints in an image are again assigned to the clusters, and the SVM-classifier with the highest posterior probability yields the most likely match. The training and testing is performed using a leave-one-ship-out cross validation.

Fisher Vectors (with localization)

For each keypoint in the image, a SIFT descriptor of length 128 is computed. This results in a variable number of descriptors per image. One fixed-length vector is computationally less intensive for matching than a variable number of SIFT-descriptors. The BOW approach is one approach to obtain a fixed-length descriptor. A more advanced approach to

obtain a fixed-length compact representation is the Fisher Vector (FV) [14][15]. The BOW approach uses hard assignment to cluster centers whereas the FV is based on soft-assignment.

The method consists of the following steps: computation of SIFT descriptors, principal component analysis (PCA) on these descriptors, Gaussian mixture model (GMM) on the principal components of SIFT, computation of the Fisher Vector, normalization, and finally, computation of nearest neighbors (NN), which results in a dissimilarity distance. The method contains two important parameters: the number of Gaussians in the GMM (g), and the number of principal components to keep with PCA of the SIFT (p). The length of the Fisher Vector is $p*g$.

In this paper, we add localization information to the SIFT feature vector of 128 values and compare objects with a compact representation in the Fisher Vector. The localization is added by concatenation and weighted with a factor 'w', where the SIFT feature is multiplied by (w) and the location information by (1-w). An overview of the localization approaches is shown in Table 1. The simplest approach to add location information is by concatenating the absolute location. To make the information more invariant to dislocation of the detection or to make it scale invariant, one can concatenate the relative location information. To separate orientation information and distance information, a polar projection can be performed from the relative location to an angle and a radius. To avoid problems with discontinuities in orientation matching, we use two orthogonal components (sin + cos) to encode the angular information [21] that can be projected on the unit circle.

Table 1: Approaches to add localization information to the Fisher Vector.

Label	Size	Description
No localization	128+0	No localization information is concatenated to the SIFT feature.
Absolute	128+2	The absolute x,y location in the image is added.
Relative	128+2	The relative x,y location is added, by subtracting the center of mass and dividing by the maximum of the image width and image height.
Unit circle	128+2	The relative x,y locations are projected on a unit circle around the center of mass. This makes the location scale invariant and it encodes the orientation angle without a discontinuity.
Radius	128+1	The radius is the distance to the center of mass that is computed based on the relative x,y location, which is scaled by image size.
Unit + Radius	128+3	This is a concatenation of the unit circle and the radius.

2.4 Ranking

For most methods, the comparison is of one image to another, using some dissimilarity measure to determine how alike two images are. For each image, a rank-1 result is obtained that selects the image with smallest dissimilarity out of all other images. Although a rank-1 match is most similar to comparing one image to (few) previous images, a more robust assessment may be obtained by looking at more matches than only the first. For each ship, the threshold is determined for which 5% of the images of that ship have a dissimilarity below that value. The percentile-5 result, is then the ship that has the lowest threshold, i.e., 5% of its images are more alike the reference image than the 5% of images of another ship. A percentile value is used instead of counting images, as this allows interpolation for ships with few images. While for most methods, the percentile-5 performance is worse than when using rank-1 (percentile-0 value), in some cases it is slightly better; the best result will be reported.

2.5 Persistent tracking simulation

In the MSA (Maritime Situational Awareness) research program, persistent tracking is one of the topics of research. Other topics are intent estimation based on the observed tracks (and other information) [4] and asset planning, where assets (such as helicopters and UAVs controlled from a frigate) are planned in time based on updated risks from intent and commercial shipping. These three parts are also built into a simulation environment, where ship traffic is simulated and observations are created based on sensor ranges and vessel appearance. An impression of the simulated traffic is shown in Figure 2, where white marks indicate co-operative (commercial) vessels, and other marks are (smaller) vessels,

such as skiffs, dhows, trawlers, etc. The squares indicate the three observation areas, each covered by a frigate with helicopter and UAV.

Figure 2: Simulation environment.

The implementation of persistent tracking is as follows:

- For commercial (cooperative) shipping, it is assumed that positions are known. These ships are not taken into account in the further examination.
- As long as ships are in radar range, they are assumed to be tracked by the observing platform.
- Only limited kinematic tracking is implemented, based on last observations and a maximum speed.
- Recognition is simulated by an single value for each vessel, which is only observed by certain sensors and at limited range. A difference in this value is an indication of the possibility of the observation matching a track. Setting a maximum allowed difference is a way of indicating how probable a confusion between ships is. In other words, if values are allowed to differ by 10%, one in ten ships looks alike, and at 1% only one in a hundred.

3. EXPERIMENTS AND RESULTS

3.1 Description of the imagery

For assessing the different methods of matching, recordings of five small ships are used, recorded during the international SMARTEX trial, which took place over two days in June 2012 in co-operation with the US Coast Guard. Target location from GPS is available for several targets, which is used in annotating images and providing distance and aspect angle estimations. Five targets are used in these tests for recognition: a sailing boat, two 24-feet cabin boats (called 24A and 24B), a flat boat of 35 feet and a 74 feet flat boat. Examples are shown in Figure 3.

From the ship positions, a distance and aspect angle relative to the camera can be estimated, as well as expected pixel resolution on the ship. From this, images are automatically retrieved over a range of resolutions and aspect angles. In this automatically selected set, most images (95%) range in resolution from 0.1 to 0.5 meters per pixel, with a maximum of 1 meters per pixel (see examples in Figure 3). 80% of images are within 20 degrees of side view, with 10% being closer to frontal/back views. After applying the automatic detection and segmentation, images are manually checked to see if the ship was actually in the image, and detected. This annotation results in 186 recordings of the five different ships, including both mid-wave and long-wave infrared. This set was used in the previous paper, but was found to have some cases with aspect ratio or resolution outside the expected range, as well as cases where the correctness of the annotation could not be verified with ground-truth information. This latter is especially important for the two 24 feet boats, that look very much alike. Therefor a new set was created, with more consistent annotation, resulting in 128 images. Table 2 gives the number of images per ship, for the full and cleaned set. Since processing for some of the methods is not symmetric

(i.e., a ship does not match its mirrored image), horizontally flipped versions of the images are added to the set. In assessing the recognition, numbers are corrected for this, and an image is not compared to its mirrored version.
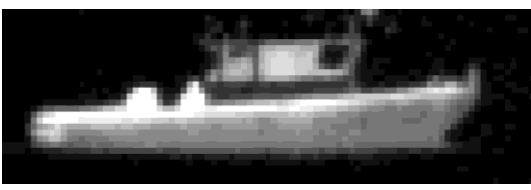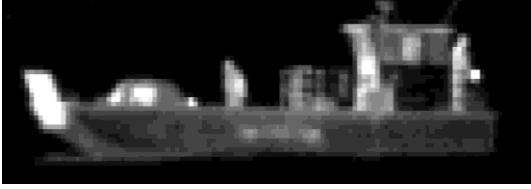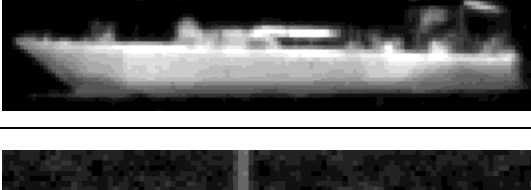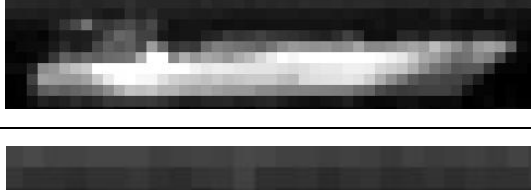
	Resolution	
	0.1 m/pixel	0.5 m/pixel
24A Cabin boat		
24B Cabin boat		
35 ft flat boat		
74 ft flat boat		
Sail boat		

Figure 3: Examples of IR images of ships in the data set, with high and low resolution.

The 35 and 74 (feet) targets are flat boats that are somewhat alike and viewed at a distance they are not dissimilar from the sail boat, of which its mast is often not seen in the automatic segmentation. The 74 flat boat only had 4 examples in this set, which makes the result for this target less accurate.

3.2 Performance measures

We used two ways to measure the accuracy of the matching methods. The first gives equal weight to each of the 128 images (accuracy by number) and the second gives equal weight to each of the five target classes (accuracy by percentage).

Table 2: Ships and numbers used in the experiment.

Ship label	Description	Full Set		Cleaned Set	
		Number of samples	Fraction	Number of samples	Fraction
24A	24-feet cabin boat	54	29%	37	29%
24B	24-feet cabin boat	67	36%	37	29%
35	35-feet flat boat	21	11%	14	11%
74	74 feet flat boat	4	2%	4	3%
Sail	sail boat	40	22%	36	28%
TOTAL		186	100%	128	100%

Combining the results of all images for each target results in a confusion matrix, containing on the diagonal the number of images that correctly have an image of the same ship as most similar, and off-diagonal the mismatched images. From these values, two overall accuracies are determined. The accuracy by number, is obtained by counting the percentage of correct images out of the total number of images:

$$Acc_{number} = \frac{\sum_{targets} N_{correct,target}}{\sum_{targets} N_{target}} * 100\% \tag{2}$$

where $N_{correct,target}$ is the number of correct matches for one of the five ships, out of N_{target} images of that ship. Although this presents a good overall value, it does not well represent the result of those ships classes that only have a few matches. Therefor a second value is used, the accuracy by percentage, which is obtained by taking the mean of the percentages obtained per target class:

$$Acc_{percentage} = \frac{\sum_{targets} \frac{N_{correct,target}}{N_{target}}}{N_{targets}} * 100\% \tag{3}$$

where $N_{targets}$ is the number of different ships. This value weighs all target classes more evenly, but is very sensitive for differences in the targets with a low number of images. For example, in our case with five target classes, an increase of one correct match of a target with only 4 images, causes a change in accuracy of 5 percent.

For the simulator, we used the average and the maximum number of track IDs given to each ship. Recognition values – indicated by confusion probability CP – are used to simulate how well ships can be distinguished from each other when a camera observation is available. In the association from observations to existing tracks, this is used to eliminate unlikely matches. A value of one indicates no recognition improvement, as each ship can appear the same as all others. Lower values indicate some exclusion is possible. For example, with CP=0.1, of possible wrong associations, 90% will be excluded.

3.3 Human recognition

As assessment of the data set and comparison to the automatic methods, recognition was also performed by a human observer on the cleaned set. For each of the images, a closest match among the rest was indicated, which is scored correctly if it was the same ship. To speed up this process, it was possible to group similar ships, for example all clear sailboats or all cabin boats with similar intensity distribution and no objects on the back. All images in such a group were scored as if the ship most common in the group was selected. The scores are shown in the table below. It is clear that to a human, the non-cabin boats are easily recognized by for example the mast, and shape of the hull. For the cabin boats, the human observer could not always identify whether an image is of one or the other, but in majority of cases, an image of the same ship that is similar can be found.

Table 3: Human recognition results.

		Predicted				
		24A	**24B**	**35**	**74**	**Sail**
GT	**24A**	**91.9**	8.1	0.0	0.0	0.0
	24B	13.5	**86.5**	0.0	0.0	0.0
	35	0.0	0.0	**100**	0.0	0.0
	74	0.0	0.0	0.0	**100**	0.0
	Sail	0.0	0.0	0.0	0.0	**100**

This corresponds to an accuracy on percentages of 95.7%. The accuracy on numbers is 93.8%. It was found that with a less precise assessment (with a faster decision instead of finding the closest match), the process was still quite slow, and the accuracy was much lower.

3.4 Recognition results of different methods

Table 4 lists the results of all the different methods, including the base methods (indicated by reference [10]), the human operator, bag-of-words with different number of clusters, and the different methods of incorporating the location in the Fisher vector description. The baseline Fisher method gives an accuracy on percentages of 86.5% for the preprocessed set A. Table 5 shows the confusion matrix of the best Fisher result on preprocessed set C which equals 92.2%.

Table 4: Comparison of the methods.

Preprocessing	Method	Accuracy on numbers	Accuracy on percentage
-	Human operator	93.8	95.7
-	Moments [10]	77.3	70.9
C	BOW, 250 clusters	73.4	61.1
C	BOW, 500 clusters	79.3	66.0
C	BOW, 1000 clusters	84.8	70.7
C	BOW, 2000 clusters	86.3	78.5
C	BOW, 5000 clusters	87.1	75.0
C	SIFT [10]	85.6	76.6
C	SIFT + localization [10]	87.9	84.7
A	Fisher + No localization [10]	83.5	86.5
B	Fisher + No localization	81.4	84.4
C	Fisher + No localization	84.5	88.5
C	Fisher + Absolute (w=0.1)	84.6	87.3
C	Fisher + Relative (w=0.1)	87.0	90.1
C	Fisher + Unit circle (w=0.1)	87.0	90.2
C	Fisher + Radius (w=0.1)	86.3	89.6
C	Fisher + Unit+Radius (w=0.1)	88.6	**92.2**

Table 5: Confusion matrix using Fisher + Unit+Radius in percent. The accuracy based on percentages is 92.2 ± 1.3.

		Predicted				
		24A	**24B**	**35**	**74**	**Sail**
GT	**24A**	**91.5**	3.9	0.0	4.6	0
	24B	16.9	**76.4**	1.1	4.9	0.8
	35	0.0	0.0	**99.6**	0.4	0.0
	74	0.0	0.0	0.0	**100.0**	0.0
	Sail	0.0	2.5	0.3	3.6	**93.6**

Among the non-Fisher methods, the SIFT + localization matching approach gives the highest recognition rates. Localization of the features provides additional discriminative power. The BOW approach does not employ localization information, but scores relatively well on the accuracy on numbers for large cluster numbers. This description in terms of a 'dictionary' is better than using SIFT matching alone.

For the BOW approach, it turns out that there is a clear difference between the accuracies based on numbers and percentages. This is caused by the fact that there is a very small number of samples of the 74 feet flat boat in the dataset, which is therefore not well represented by the generated clusters. The closest cluster centers will therefore be far off for this type of boat, making it very difficult to recognize properly.

This effect is not present in the Fisher vector approach, where not only the numbers of different clusters are used to describe the image but also the distances to these clusters. This produces a richer representation which seems to incorporate all the pros of the other methods. The results show that it is beneficial to add localization information to the Fisher Vector. The difference between the best FV without localization and the best FV with localization is almost 4%. The location encoding with a unit vector (indicating direction) and radius (indicating relative distance) appears to perform best at 92.2% accuracy on percentage.

3.5 Effect of recognition on tracking in simulations

In simulation, it is modeled that 286 smaller ships are observed (by sensors) in a 5 day period, in the limited area of operation of three frigates and their assets (see Figure 2). Recognition is only possible when multiple observations are done with a camera where the ship was out of view in between observations.

Table 6: Results of persistent tracking for a confusion probability (CP) value of 1 (only kinematics in the tracking algorithm), and for a ship type dependent recognition.

Confusion Probability	Measure of performance	
	Average tracks per ship	Max. tracks per ship
1 (no recognition)	32.2	132
Type specific, from 1/4 to 1/50	13.6	67

Table 6 shows the results of tracking the smaller ships over five simulated days. This results in an average of 32.2 track IDs per ship, without recognition. For recognition, CP is set at fixed values for different types of ships: 0.25 for skiffs, 0.05 for dhows and 0.02 for (less common) larger vessels. In that case, it appears that even this limited recognition of the smaller ships (skiffs, dhows) gives a major improvement (from 32.2 to 13.6 tracks per ship).

4. DISCUSSION

The assessment by a human observer shows that it is possible to distinguish the flatter boats (sailboat and flat boats) very well, based on details that may not be apparent automatically, especially at lower resolutions. Distinguishing between the two cabin boats, i.e., indicating if it is 24A or 24B, is not perfectly possible, however, given an image of one, it is often

possible to find a very similar image, based on small details that may be less visible in the automated recognition as well. However, human observation is slow, and may be very much helped by using the automated methods to make a first selection.

The keypoint matching, bag-of-words (BOW) and the Fisher vector methods all use the same SIFT keypoints. The results using Fisher vector (even when not using localization) are better when looking at the accuracy by percentage, as this shows results on the ships with fewer images more prominently. It was found that these ships were not represented well in the BOW clusters, unless many clusters were used, and even than the flat boats were mixed with the sail boat description. This may partly be the case as well for the GMM clusters in the Fisher vectors, but since a soft assignment is used in the latter case, differences for single images may be still be seen.

The different sets of cropped images around the center of mass (preprocessing A, B and C in Table 4), show best results for set C, where more space around the ship is taken. This may allow more accurate SIFT key point descriptions on the edges of the ships. Of different ways of incorporating the keypoint location in the Fisher Vector description, using a unit vector (indicating direction) and radius (indicating relative distance to the center) appeared to perform best. This localization makes the orientation information independent of scale and without angular discontinuities and it preserves distance information in an orthogonal component.

Compared to the baseline, the addition of location in the Fisher vector description clearly improves the recognition results. Since an increase is seen both in the accuracy by numbers (favoring the more common images) as well as accuracy by percentage (favoring the less common ships), this improvement seems to hold for all ships. The overall best result is for the Fisher vector with unit and radius localization, which shows recognition of over 92 percent for most ships, except for the 24B cabin boat, which is mainly confused with the (almost identical) 24A. This performance is close to that of the human observer.

5. CONCLUSIONS

In this paper, we presented an update of our research on the use of feature extraction and matching methods for recognizing ships from electro-optical imagery. An indication of discriminative power is obtained on infrared imagery of ships.

The overall performance using Fisher vectors with SIFT keypoints is improved by including keypoint localization in the representation. Compared to a more commonly used bag-of-words approach, using Fisher vectors allows for better recognition of ships with few example images, as the soft-assignment also uses likeness to other ships in the similarity measure. For the SIFT keypoints, results show that it is better to include more area around the ship to obtain good descriptors. The performance of our approach is close to that of the human observer, and the system may be helpful to assist a human in retrieving similar images of a target faster.

The simulation of tracking of small ships over five days, with small coverage of observation area, shows that recognition percentages as obtained with Fisher vectors (i.e., in the order of one in ten ships looks alike) already gives a significant increase in tracking accuracy in terms of track breaks, resulting in long-term descriptions of ships.

ACKNOWLEDGEMENT

REFERENCES

[1] Alves, J., Herman, J., Rowe, N., "Robust recognition of ship types from an infrared silhouette," Thesis Monterey Naval Postgraduate School California USA, (2004).

[2] Bouma, H., Lange, D.J. de, Broek, S.P. van den, Kemp, R., Schwering, P., "Automatic detection of small surface targets with electro-optical sensors in a harbor environment," Proc. SPIE 7114, (2008).

[3] Bouma, H., Dekker, R., Schoemaker, R., Mohamoud, A., "Segmentation and wake removal of seafaring vessels in optical satellite images," Proc. SPIE 8897, (2013).

[4] Broek, A.C. van den, Broek, S.P. van den, Heuvel, J.C. van den, Schwering, P., Heijningen, A.W. van, "A multi-sensor scenario for coastal surveillance," Proc. SPIE 6736, (2007).

[5] Broek, A.C., van den, Hanckmann, P., Smith, A., Bolderheij, F., "Vessel intent recognition in maritime security operations," Proc. NATO SCI 247, (2012).

[6] Broek, S.P. van den, Bakker, E.J., Lange, D.J.J. de and Theil, A., "Detection and classification of infrared decoys and small targets in a sea background," Proc. SPIE 4029, 70-80 (2000)

[7] Broek, S.P. van den, Schwering, P., Liem, K., Schleijpen, R., "Persistent maritime surveillance using multi-sensor feature association and classification," Proc. SPIE 8392, (2012).

[8] Broek, S.P. van den, Bouma, H., Degache, M., "Discriminating small extended targets at sea from clutter and other classes of boats in infrared and visual light imagery," Proc. SPIE 6969, (2008).

[9] Broek, S.P. van den, Bouma, H., Degache, M., Burghouts, G., "Discrimination of classes of ships for aided recognition in a coastal environment," Proc. SPIE 7335, (2009).

[10] Broek, S.P. van den, Bouma, H., Veerman, H., Benoist, K., Hollander, R. den, Schwering, P., "Recognition of ships for long-term tracking," Proc. SPIE 9091, (2014).

[11] Dekker, R., Bouma, H., Breejen, et. al., "Maritime situation awareness capabilities from satellite and terrestrial sensor systems," Proc. Maritime Systems and Technologies MAST Europe, (2013).

[12] Gray, G. J., Aouf, N., Richardson, M. A., Butters, B., Walmsley, R., Nicholls, E., "Feature-based recognition approaches for infrared anti-ship missile seekers," Imaging Science Journal 60(6), 305-320 (2012).

[13] Ergula, M., Alatana, A., "An automatic geo-spatial object recognition algorithm for high resolution satellite images," Proc. SPIE 8897, (2013).

[14] Jegou, H., Douze, M., Schmid, C. "Hamming embedding and weak geometry consistency for large scale image search," Proc. ECCV, (2008).

[15] Jegou, H., Perronnin, F., Douze, M. Sanchez, J. Perez, P., Schmid, C., "Aggregating local image descriptors into compact codes," IEEE Trans. Pattern Analysis and Machine Intelligence 34(9), 1704 - 1716 (2012).

[16] Kim, S., "Analysis of small infrared target features and learning-based false detection removal for infrared search and track," Pattern Analysis Appl., (2013).

[17] Lowe, D., "Object recognition from local scale-invariant features," IEEE ICCV, (1999).

[18] Mikolajczyk, K., Schmid, C., "A performance evaluation of local descriptors," IEEE Trans. Pattern Analysis and Machine Intelligence 27(10), 1615-1630 (2005).

[19] Mouthaan, M.M., Broek, S.P. van den, Hendriks, E.A., Schwering, P., "Region descriptors for automatic classification of small sea targets in infrared video," Optical Engineering 50(3), (2011).

[20] Perronnin, F., Sánchez, J., Mensink, T., "Improving the fisher kernel for large-scale image classification," ECCV, 143-156 (2010).

[21] Rieger, B., Vliet, L. van, "A systematic approach to nD orientation representation," Image and Vision Computing 22, 453-459 (2004).

[22] Schwering, P., Lensen, H., Broek, S.P. van den, Hollander, R. den, Mark, W. van der, Bouma, H., Kemp, R., "Application of heterogeneous multiple camera system with panoramic capabilities in a harbor environment," Proc. SPIE 7481, (2009).

[23] Schwering, P.B.W., Bezuidenhout, D.F., Gunter, W.H., le Roux, F.P.J., Sieberhagen R.H., "IRST infrared background analysis of bay environments," Proc. SPIE 6940, (2008).

[24] Sivic, J., Zisserman, A., "Video Google: A text retrieval approach to object matching in videos," ICCV, (2003).

[25] Steinvall, O., Elmqvist, M., Karlsson, K., Larsson, H., Axelsson, M., "Laser imaging of small surface vessels and people at sea," Proc. SPIE 7684, (2010).