SHIP RECOGNITION USING OPTICAL IMAGERY FOR HARBOR SURVEILLANCE

Dr. Patricia A. Feineigle, Dr. Daniel D. Morris, and Dr. Franklin D. Snyder
General Dynamics Robotic Systems, 412-473-2159 (phone), 412-473-2190 (fax)
fsnyder@gdrs.com, http://www.gdrs.com

ABSTRACT

Many military and homeland defense missions require automated situation awareness in maritime environments. A major element of these missions is automatic detection, tracking, and recognition of ships as they transit harbors. We advocate the use of optical sensors in an Automated Target Recognition (ATR) system to accomplish these missions. This paper reports on the development of maritime optical ATR systems that incorporate a new capability for recognition of known ships, using a database of previously acquired imagery. The approach investigated here uses the local interest point detector and descriptor known as SIFT (Scale Invariant Feature Transform) features. The SIFT interest point detector locates extrema in scale space of Difference-of-Gaussian functions, generating a set of distinctive image regions; the keypoint descriptor measures the orientation of local gradients in the region. The features are normalized, making them invariant to image scaling and rotation and partially invariant to changes in illumination and viewpoint. SIFT features are used in object recognition by matching features extracted from test images with those previously measured in database images. Following feature matching, a geometric verification process is used to eliminate false matches. This paper describes criteria developed to recognize ships using SIFT features and strategies employed to handle changes in camera viewpoint and cluttered backgrounds common in maritime environments.

1 INTRODUCTION

A focal point of global maritime security is harbor situation awareness, including reliable classification and identification of ships entering and leaving U.S. and foreign ports. Since signatures from AIS, ELINT, COMINT, and acoustic contacts can be spoofed or noisy, it is recognized that Maritime Domain Awareness (MDA) can benefit from independent optical verification of ships. Hence, video-based automatic target recognition (ATR) software has been identified recently as an important component of automated maritime situation awareness. This paper discusses an approach to ship recognition using optical sensors and a database of previously acquired images of ships of interest.

Much work has been done recently on object recognition in digital imagery for applications ranging from on-line image search engines to security monitoring systems [4, 5]. An approach that is gaining acceptance for these tasks is based on the detection of unique interest points in objects to be identified. Rather than recognition of an object as a whole, recognition occurs by matching a set of interest points, via descriptors that characterize the region surrounding an interest point. This type of approach provides some immunity to distortion and occlusion, problems frequently encountered in optical images. This feature is very important for MDA due to the cluttered environment in most ports.

Several different interest point detectors and descriptors have been described in the literature [1, 2]. To test this type of approach for maritime ship recognition, we chose to use the Scale Invariant Feature Transform (SIFT) [3] since it has performed well in independent comparisons of similar methods. The remainder of this paper will describe briefly the basic SIFT detector/descriptor and how we applied this method to ship recognition, with illustrations drawn from our test results.

2 APPROACH

In general, object recognition from optical imagery requires a method that is insensitive to object position, orientation (viewpoint), scale (zoom level), and intensity (lighting). The method also must be tolerant to changes in the object's surroundings and partial occlusion. These requirements give interest point detection techniques several advantages over whole object-based methods, such as contour matching, including: a) image segmentation is not a pre-requisite for object recognition; b) the local nature of interest points gives immunity to clutter, occlusion, position, scale, and 2D orientation; and c) illumination changes can be compensated partially by the local region descriptor. Object recognition with interest points involves three major steps: detection of distinctive points in an image, characterization of each distinctive point with a region descriptor, and matching of the descriptor to those stored in a database. If a sufficient number of descriptors are matched to the database, and the spatial configuration of the interest points is consistent with the database information, an object can be declared as recognized.

The interest point detector used by Lowe [3] for his SIFT features finds image regions with large gradient changes at a sequence of image scales. This is achieved by filtering the image with a Difference-of-Gaussian (DoG) filter at multiple image resolutions. Interest points are selected from the filter outputs at positions that are maxima or minima over both spatial position and scale. Each interest point is identified by its (x,y) image position, the scale at which it was detected, and the principal orientation(s) of the gradients in the region surrounding the point; the orientation is determined by locating peak(s) in the histogram of the local gradients. Figure 1 shows an example of the interest points detected by this method. The origin of each arrow shows the location of an interest point, with the arrow direction indicating the principal orientation of the feature and the length representing the scale at which the feature was detected.

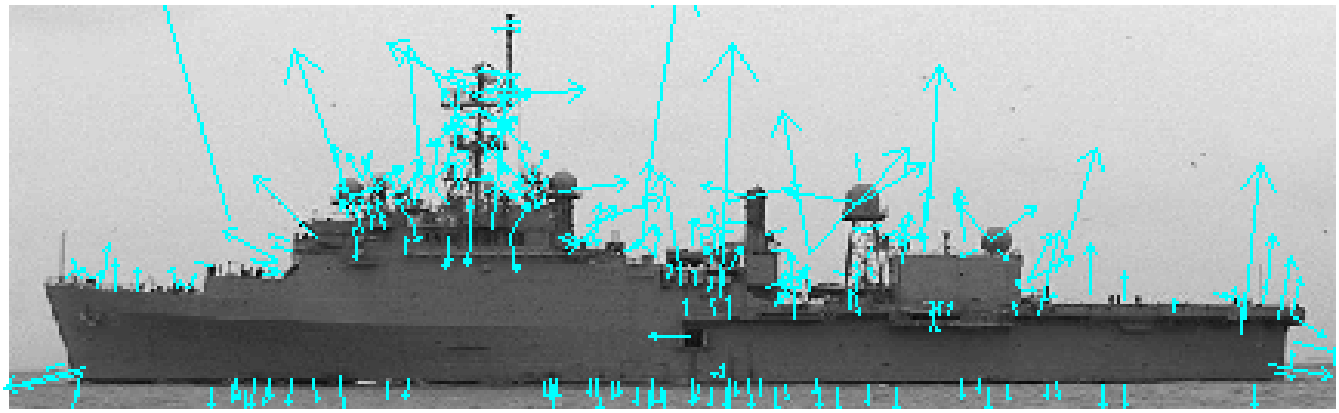


Figure 1. Example of interest points extracted from a ship image.

The region descriptor used by the SIFT method consists of a 128 element vector that represents the orientations of local gradients in the region centered at the interest point location. The vector consists of sixteen, eight-bin (spanning 360 degrees) orientation histograms; each of the histograms corresponds to one block of the 4x4 grid overlaid on the image region, with each block consisting of 4x4 pixels. The histograms are weighted by the gradient magnitudes and are computed from the DoG filter outputs at the scale of the interest point, with the data rotated relative to the orientation of the interest point, providing in-plane rotation invariance to the descriptor. The feature vector is normalized to unit length to provide some degree of image intensity invariance.

Once interest point descriptors have been extracted from a database image, the descriptors, along with their corresponding position, scale, and orientation information, are stored in the keypoint database. Since we are interested in determining whether a particular ship is present in a new image, we store the interest points for each database image separately, rather than in one large database as was done originally by Lowe; this practice eliminates the needless comparison of test image features with database features extracted from other ship types. Feature vectors can be compared using any standard vector distance measure, such as Euclidean distance or angle between vectors. Each test image feature is

compared to each database feature to find the best matching database feature; if the distance between this pair of feature vectors is sufficiently small, a matched feature can be declared. We have added an additional constraint to the matching procedure to reduce the likelihood of false feature matches for similar image regions. We require that the "matched" database feature also selects the "paired" feature from the test image as its best match, from among all other features in the test image, in order to qualify as a matched feature; we term this requirement the reciprocal match test.

A second stage in the feature matching process is implemented to verify that an appropriate geometric relationship exists between the matched keypoints. This stage uses the actual position, orientation, and scale information of the detected interest points (in image space), rather than the feature descriptors (feature space data). For each pair of matched features, the change in position, orientation, and scale is computed between the test and database image. An iterative, least-squares fitting procedure is applied to the set of translation and rotation values to determine if a consistent affine transformation relates the matched interest points in the test and database images. Keypoint pairs that appear to produce large errors in the geometric fit are eliminated from the matched set. Iterations continue until no more keypoint pairs need to be removed from the matched set. If the final fitting error is sufficiently small, and the number of remaining interest point matches is above the threshold value, a ship can be declared as "recognized". Figure 2 shows an example of the benefit of the geometric fit process. In this test, two images of a ship undergoing a large translation were compared; matching keypoints for several features on the ship were found, as well as a matching keypoint on a stationary buoy (on the right side of the image). The geometric fit identified the consistent translation of the ship-based points (shown by blue lines in the figure) but eliminated the buoy point due to its lack of a similar movement.

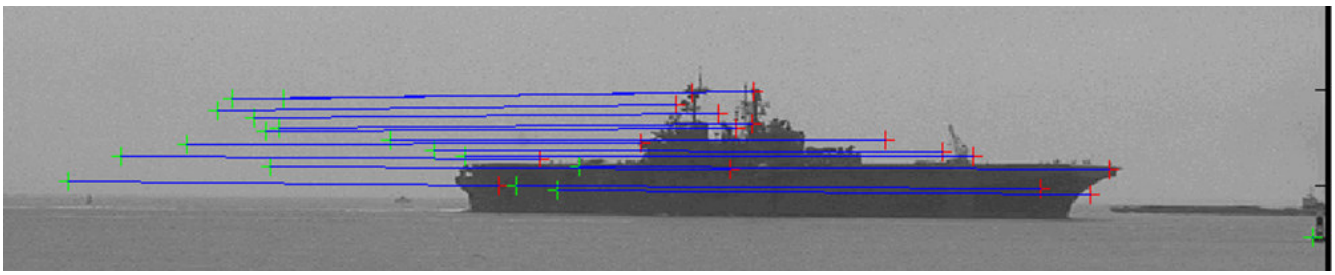


Figure 2. Demonstration of feature matching for a transiting ship.

As it was designed, the SIFT method for interest point detection and description is invariant to in-plane (2D) rotation and partially invariant to out-of-plane rotation. Lowe estimates that a non-planar object can undergo a twenty degree out-of-plane rotation and still generate SIFT features that match those produced by the non-rotated object [3]. This characteristic is important when devising a ship recognition system since ship appearance varies greatly with viewpoint. Our system requires a set of images for each ship to be recognized; each image set consists of target ship views spaced at twenty to thirty degrees in azimuth, covering the expected viewpoint range. The database images should be recorded at an elevation angle and zoom level that are consistent with those achievable by the surveillance system. During operation, the sensor system would compare an acquired image to each database image for the expected target ship until a recognition decision is made. A positive decision can be made more quickly than a negative decision, in most cases, since a negative decision requires comparison to the full set of images.

In its most basic form, a recognition decision can be made based solely on the number of SIFT feature pairs that remain after the geometric verification process is performed. In this case, a ship is recognized if the number of matched keypoints exceeds the threshold and not recognized otherwise.

Our test results show that additional requirements can be implemented to improve the correct decision rate. First, a large root mean square error from the geometric fitting process indicates that the computed transformation likely is not valid and the ship should not be recognized. Second, if the

matched keypoints subtend only a small area of the test image, the evidence for ship recognition is reduced because a local feature, such as a ship porthole, may have produced all of the matching keypoints. In isolation, there are many ship components, including portholes, antennas, masts, and winches, that look similar on different types of ships; however, the combination of such components and their physical arrangement are generally sufficient to distinguish ships. Thus, we require that the area subtended by the matched features exceeds an area threshold in order to recognize a ship. Finally, due to the non-uniqueness of local, small scale features, some keypoints match frequently on similar types of ships and may subtend a large area of the ship (such as a series of portholes). Hence, some ship features are more distinctive than others and are weighted more highly in our recognition system. This weighting is determined by calculating the frequency with which each keypoint matches those extracted from a set of images in the false alarm (FA) database; keypoints that generate many false matches are penalized more heavily than those that do not. The scale factor for this weighting is determined by a trade-off between the number of true and false recognitions for a given ship. The use of these three additional test criteria has reduced the false recognition rate, without a decrease in correct recognitions, in our test sets.

3 TEST RESULTS

Extensive testing has been done to evaluate the feasibility of the SIFT feature-based approach for use in ship recognition by an automated harbor surveillance system. The images used for these tests were acquired by an automated, land-based camera system set-up in Norfolk, VA to capture harbor traffic that emit an AIS message (required of all large ships). The AIS messages provided unique ship identification (MMSI) and real-time location. This information was used to adjust a camera's pan, tilt, and zoom settings to capture images of the ship once per second as it moved through the harbor. These images, along with others, provide a wide variety of ship types and viewpoints, including military and

commercial vessels, as well as pleasure craft. Repeated appearance of certain unique ships, such as specific tugboats, enabled recognition testing under varying weather and lighting conditions.

The many tugboats that operate in Norfolk tend to follow similar paths as they transit the harbor. Our camera system was able to capture similar sequences of a group of tugboats on multiple days; some of these tugboats look quite similar while others show distinct differences. These images were used to form several databases for testing detection and false alarm rates (that is, true and false recognition decisions). One such test used a database of seven different views of a tugboat named the Gulf Coast, captured on several different days. Each of these images was compared to a sequence of nearly two hundred images of the same boat, captured on a different day from the database images; the number of matched features (after completion of the geometric verification process) from each comparison was recorded. Figure 3(a) shows a plot of the maximum number of matches, taken across the seven database images, for each of the images in the test sequence. It can be noted that eight or more features were matched in all but four of the test images; the reduced matching in the remaining images was due to occlusion of the tugboat by a pole in the camera's field of view. In contrast, Figure 3(b) shows the matching results for the Gulf Coast database compared to images of 20 different tugboats. The maximum number of matched features for these boats was seven, with a mean of 1.7 matches. Test results such as these indicate that a lower limit of eight matches is a reasonable threshold for making a positive recognition decision for a ship with visual detail comparable to a tugboat and imaged at a high resolution. Smaller or less complex ships, or a lower resolution sensor system, would produce fewer interest points and, therefore, fewer matches.

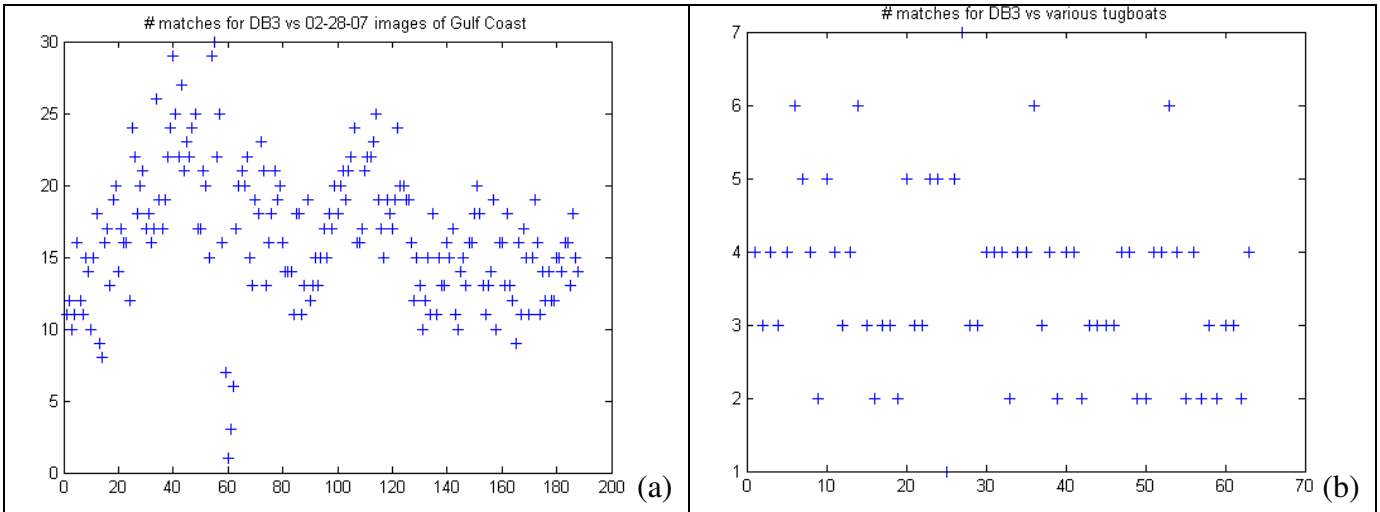


Figure 3. Plot of the number of matched features for the Gulf Coast when compared to (a) a long video sequence of the Gulf Coast and (b) a set of images of different tugboats.

Tugboat images were used to evaluate the use of keypoint match frequencies in the recognition decision computations. A single image of the Gulf Coast was selected as a database image; this image was compared to approximately 1100 images collected of this tugboat on several days. The frequency, with which each feature from the database image was matched in the test set, was counted. Figure 4(b) shows this frequency distribution as a gray-scale image, while the locations of the most frequently matched features are circled in Figure 4(a) (overlaid on the database image). In contrast, the database image was compared to all images of other tugboats collected on several days (a total of 6859 images) to determine which features were likely to match in non-identical ships. Figure 4(d) shows the frequency distribution for the false matches, with the locations of features that matched in at least ten images circled in Figure 4(c). For this test, if a recognition decision was based solely on a count of the number of matched features, with a threshold setting of nine matches, ten of the 6859 false alarm images would "alarm". Using the match frequency-weighting described previously, the number of alarms would be reduced to six or to one, depending on the scale factor selected. The images generating the false recognitions were from a tugboat, called the New England Coast, which looks very similar to the Gulf Coast; this tugboat is shown in Figure 4(e).

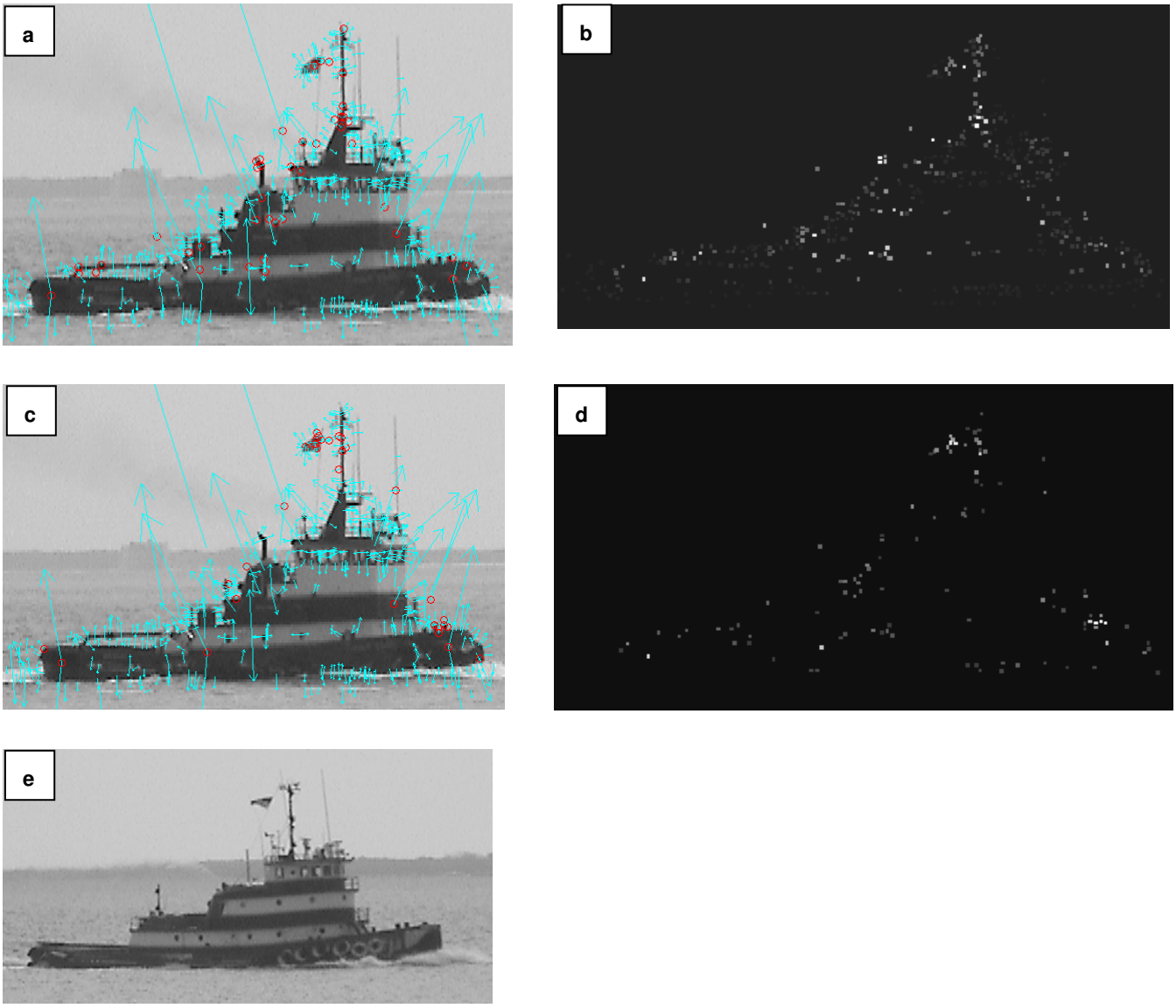


Figure 4. Results of matching tests for the Gulf Coast. (a) Locations of the SIFT interest points, shown at the detected scales and orientations, with red circles indicating the most frequently matched features (>= 50) for the same tugboat. (b) Correct match frequencies of the SIFT features displayed as a gray scale image (scaled to a maximum of 203). (c) Locations of the most frequently matched features (>= 10) in images of different tugboats. (d) False match frequencies of the SIFT features displayed as a gray scale image (scaled to a maximum of 235). (e) The New England Coast.

Image sequences from a wide variety of cargo carriers also were used to test the recognition capabilities of this SIFT-based algorithm. These tests demonstrated that even ships which look similar to the human eye are distinguished easily with this approach. Matching results showed that recognition was based on features located on ship structures, rather than on cargo containers, providing robustness to appearance changes common in this vessel class. Another test verified that identical ships (with different hull numbers) can be recognized as such. These results are illustrated in Figure 5 which shows images of the Nele Maersk, the Nicolene Maersk, and the Maersk Nevada. By eye, the Maersk Nevada looks similar to the Nicolene Maersk, but the Nele Maersk looks identical to the Nicolene (exclusive of the name and the cargo). An image subset, including different ship viewpoints, was extracted from video sequences in which each of these vessels moved in similar trajectories. The images of the Nicolene Maersk were compared to those for the other ships. The results for the Nicolene versus Nele comparison showed that image pairs, in which the ships were oriented similarly, consistently found ten or more matching features, with a maximum of 39 matches. In contrast, the comparison of Nicolene with Nevada showed a maximum of 8 matches, with an average of only 1.5 matched features. These match counts are shown as grayscale matrices in Figures 5(d) and (e); the width of the peak in 5(d) is dependent on the duration over which the ships were oriented in a similar direction.

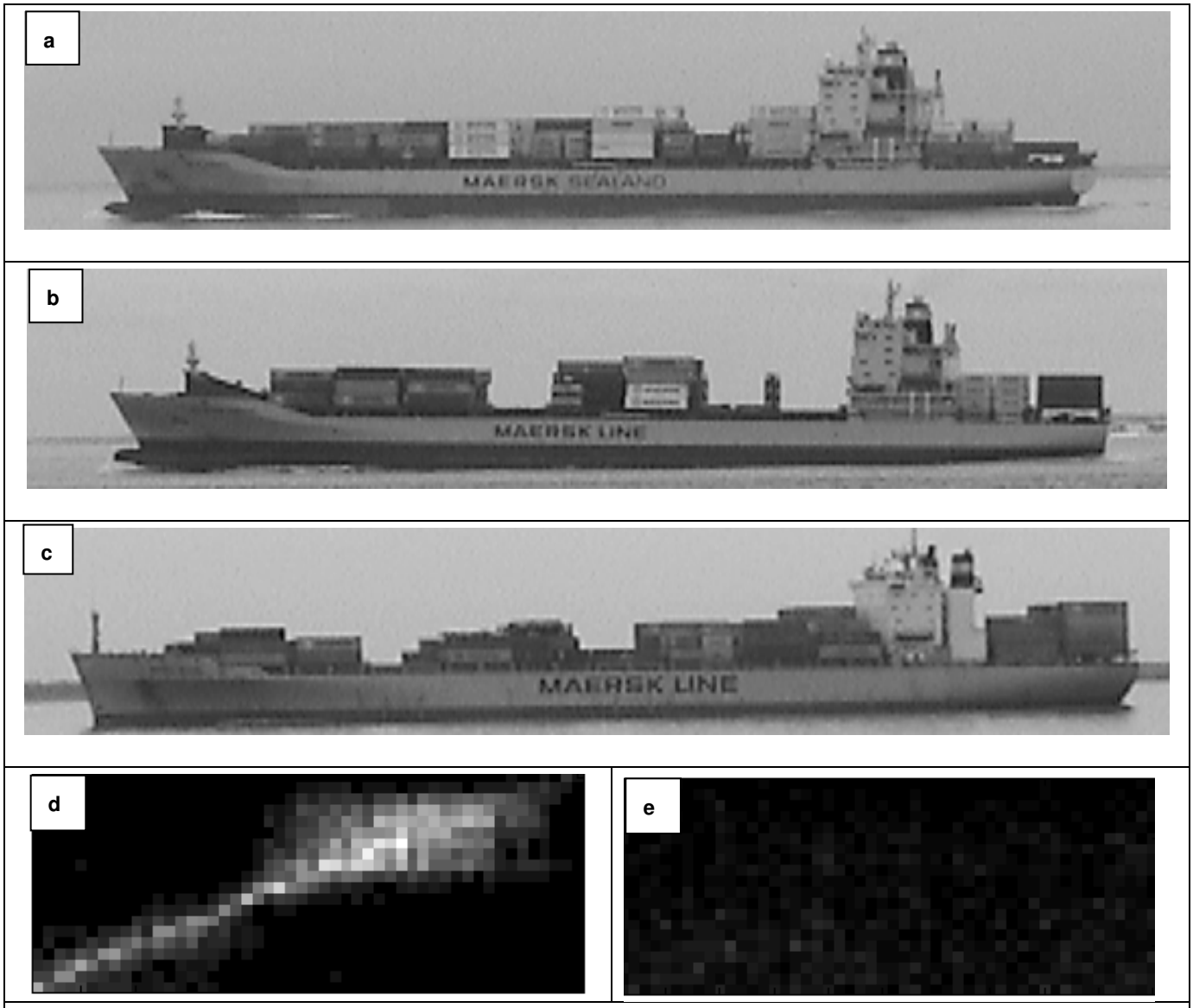
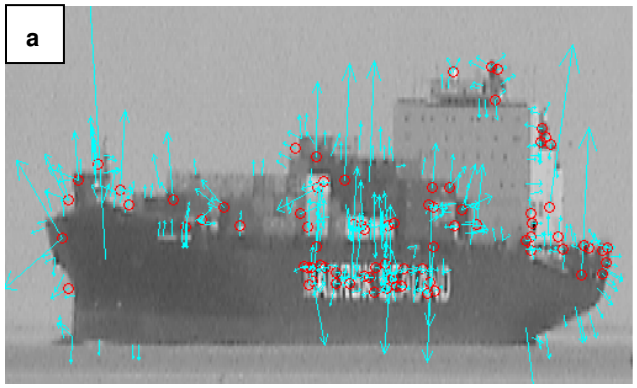


Figure 5. Matching results for three similar cargo carriers. (a) Nele Maersk. (b) Nicolene Maersk. (c) Maersk Nevada. (b) and (a) are the same ship type while (b) and (c) are different. (d) Number of matched feature pairs from a comparison of videos for Nele and Nicolene (scaled to a max of 39). The width of the bright band reflects the rate at which the orientation of the vessels changed throughout the sequences. Differences in the number and arrangement of containers loaded on the ships did not prevent recognition of these same ship types. (e) Number of matched feature pairs from a comparison of Nicolene and Nevada (scaled as in (d)), indicating that these are indeed different ships.

Additional tests addressed issues regarding the effect of background clutter and occlusion on the recognition capabilities of this approach. We observed that clutter bordering a ship's perimeter can alter the large scale features that characterize the transition region between the ship and its background; however, smaller scale, internal ship features are not changed. As long as sufficient internal ship features are present in an image, recognition can occur regardless of background clutter. The effect of occlusion is similar. Occluding objects primarily alter the local features in the occluded region. If a sufficient number of features can be matched in the visible portion of the ship, recognition can still occur. Figure 6 illustrates these findings. An image of a cargo carrier, named the Sealand Pride, was compared to a sequence in which the vessel was occluded by another ship being pulled by a tugboat; the number of matched features throughout the sequence was plotted, as shown in Figure 6(c). Despite the occlusion, the database image was able to match a large number of features in each frame. The dip at the beginning of the plot is due to differing viewpoints of the test images relative to the single database image; the dips at the end of the sequence are due to the tugboat and then the larger ship passing in front of the Sealand Pride, primarily occluding the painted name on the side of the ship. Features extracted from identifiers painted on vessels have proven to be quite useful in this recognition algorithm, as shown by this example.



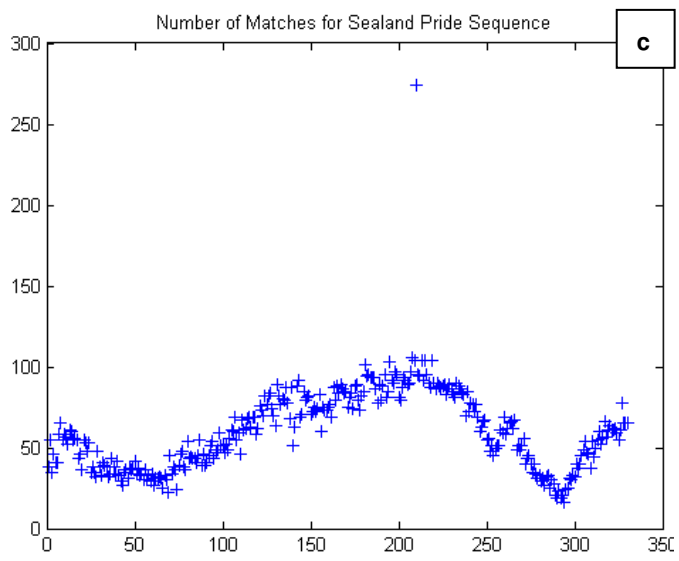


Figure 6. Matching results for a single image of the Sealand Pride against the remainder of the sequence. (a) Locations of the SIFT interest points, shown at the detected scales and orientations, with red circles indicating the most frequently matched features. (b) Image of the cargo carrier being occluded by another ship and tugboat. (c) Number of features matched in each image of the sequence. The dips in the curve are due to differing viewpoints and occlusion; the peak is due to the selected database image.

There are two aspects of the SIFT feature descriptor that require special consideration when this approach is used in harbor surveillance. While the SIFT descriptor is tolerant of minor changes in lighting (brightness and contrast), it is not invariant to drastic, non-linear lighting differences between database and sensor images. To accommodate this problem, we include images recorded under different lighting conditions in the image database; basic image tests can determine whether good or unfavorable lighting exists in an acquired image, and then the appropriate database images can be selected for comparison. The other issue is image scale; SIFT features can be matched reliably in images with scale changes up to a factor of two. We use the zoom capability of our camera system and an estimate of the range to a target ship to obtain a consistent image resolution for matching with database images.

4 SUMMARY

We have described a method developed for using local features extracted from optical imagery for automated ship recognition in harbor surveillance. The advantage of using local features is that they are tolerant of clutter typically present in maritime environments. With appropriate selection of the algorithm which detects and describes the local interest points, feature matching can be performed reliably despite moderate changes in image content, lighting, zoom, focus, and orientation. The SIFT technique, used to demonstrate local feature matching, has performed well in this application. The methods we have developed for incorporating multiple views of target ships in the image database and for verifying the geometric relationships of the matched features have been key to achieving excellent recognition rates in our test sets. Further work will evaluate the performance of this approach in a fielded surveillance system.

ACKNOWLEDGEMENTS

This work was performed under the sponsorship of ONR FNC-AO-IA, contract #N00421-03-C-0027, Marc Steinberg, PM. The automated data collection system was developed by Mr. Brian Colonna.

REFERENCES

1. Mikolajczyk K. and Schmid C., “A Performance Evaluation of Local Descriptors”, IEEE PAMI, Vol. 27, No. 10, pp. 1615-1630, October 2005.

2. Mikolajczyk K., Tuytelaars T., Schmid C., Zisserman A., Matas J., Schaffalitzky F., Kadir T., Van Gool L., “A Comparison of Affine Region Detectors”, IJCV, Vol. 65, No. 1-2, pp. 43-72, November 2005.

3. Lowe D.G., “Distinctive Image Features from Scale-Invariant Keypoints”, IJCV, Vol. 60, No. 2, pp. 91-110, November 2004.

4. Sivic, J. and Zisserman, A., “Video Google: A Text Retrieval Approach to Object Matching in Videos”, Proceedings of ICCV 2003, Vol. 2, pp. 1470-1477, October 2003.

5. Nister, D. and Stewenius H., “Scalable Recognition with a Vocabulary Tree”, Proceedings of CVPR 2006, Vol. 2, pp. 2161-2168, 2006.