Learning and Leveraging Context for Maritime Threat Analysis: Vessel Classification using Exemplar-SVM

Brendan Morris[1], David W. Aha[2], Bryan Auslander[3], and Kalyan Gupta[3]

[1]University of Nevada, Las Vegas
[2]Navy Center for Applied Research in Artificial Intelligence;
Naval Research Laboratory (Code 5514); Washington, DC
[3]Knexus Research Corporation; Springfield, VA
brendan.morris@unlv.edu, david.aha@nrl.navy.mil,
{bryan.auslander, kalyan.gupta}@knexusresearch.com

September 27, 2012

Abstract

Modern fleet security requires accurate threat analysis in real-time, which relies on a range of contextual information (e.g., vessel size, speed, heading, etc.). Rich contextualization may be possible using imaging systems if the images can be used to detect and classify maritime vessels and track their movements. In this work, the effectiveness of the ensemble of Exemplar-SVMs (E-SVM) object detection scheme is evaluated for maritime data where targets are small and have low inter-class variation due to its scalability and ability to learn from limited training examples. Experimental evaluation shows average precision for Annapolis Harbor vessel data is lower than the general 20-category PASCAL VOC challenge due to confusion between boat types.

1 Introduction

Early comprehension and prediction of threats is critical for modern fleet security. Threat detection and assessment is a challenging problem in the complex maritime environment because of wide coverage areas, large number of simultaneous vessel activities, and the need for real-time analysis. Currently, maritime threats are assessed by watchstanders, officers designated for lookout, with assistance from automated video surveillance systems. These systems are designed reduce overload when monitoring many concurrent contacts. However, most of these systems perform simplistic perimeter-based behavior analysis. Threat is only considered a function of proximity rather than they myriad of features, characteristics, and relations among maritime vessels.

Behavioral analysis techniques have been applied to maritime data to learn typical patterns of activity and detect anomalies. A comparison between global and local anomaly detection algorithms found that global density based algorithms outperformed the local ones in terms of both speed and accuracy [1]. A dynamic dual-hierarchical Dirichlet process probabilistic model has recently been proposed for using the activity models as priors for incrementally updating harbor models over time for long-term surveillance [2]. Another study comparing probabilistic sequence models, hidden Markov models, conditional random fields, and Markov logic nets (MLNs), with the rule-based perimeter defense technique found the MLN was better able to detect attack situations on

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
27 SEP 2012		**00-00-2012 to 00-00-2012**

4. TITLE AND SUBTITLE

Learning and Leveraging Context for Maritime Threat Analysis: Vessel Classi cation using Exemplar-SVM

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory,Navy Center for Applied Research in Arti cial Intelligence,Code 5514,Washington,DC,20375

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR’S ACRONYM(S)

11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Modern eet security requires accurate threat analysis in real-time, which relies on a range of contextual information (e.g., vessel size, speed, heading, etc.). Rich contextualization may be possible using imaging systems if the images can be used to detect and classify maritime vessels and track their movements. In this work, the e ectiveness of the ensemble of Exemplar-SVMs (E-SVM) object detection scheme is evaluated for maritime data where targets are small and have low inter-class variation due to its scalability and ability to learn from limited training examples. Experimental evaluation shows average precision for Annapolis Harbor vessel data is lower than the general 20-category PASCAL VOC challenge due to confusion between boat types.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**28**	

real mission data where unmanned sea surface vehicles were tasked with protecting a high value unit [3].

Often the data used for behavior analysis comes from the Automated Identification System (AIS) which is able to track surrounding vehicles by data exchange. However, a true threat detection system can not rely on accurate AIS information as this could be manipulated by an attacker or simply not be available (e.g. as with smaller vessels). Vision systems are one technique for tracking in the local area with the advantage of being able to distinguish the vessel type as well. The vessel type is another major factor for maritime threat assessment crucial for mission planning and execution [4]. Smaller vessels, which are not required to use AIS like larger ones, have high maneuverability which increases their potential danger as a small motorized attack vessel. Researchers have recognized the importance of automated systems for visual classification and tried to improve simple classifiers with the use of contextual cues and relationships [5] but an in depth study of visual classification performance has not yet been presented in literature.

The computer vision field has been actively studying object classification for many years. This activity has lead to the organization of the PASCAL Visual Object Classes (VOC) Challenge [6] as a benchmark in visual object category recognition and detection. Each year, the very best classification systems compete in the 20 object class challenge to push the state-of-art. In 2010, Malisiewicz et al. [11] proposed the ensemble exemplar-SVM (E-SVM) classification system, which had high performance with a conceptually simple generalized nearest neighbor classifier. Using E-SVM, strong classifiers could be developed with few positive examples.

This work examines the effectiveness of the ensemble E-SVM system on maritime vessel classification where targets are small and have low inter-class variation. This work addresses a number of research questions:

- Can distinct vessel types with low inter-class variability be reliably discerned?
- How well does E-SVM perform when the object categories have high intra-class variability?
- How should exemplars be selected from a dataset?

Preliminary results on a small number of hand selected exemplars for four distinct types of vessels from the Annapolis Harbor dataset suggest the E-SVM technique can be a viable maritime classification scheme that can also be used as a detector for tracking. Further studies need to be conducted to fully characterize the performance on a complete set of vessel types and compare with other top performing VOC algorithms.

2 Annapolis Dataset

In order to analyze the performance of object detection and classification algorithms in maritime applications, a new marine vessel dataset was collected from the Annapolis Harbor. Video overlooking the harbor was captured from webcams placed by the Annapolis Yacht Club [7]. Images of the "Spa Creek" view were grabbed from the public video streaming website in 1 second intervals. The data was collected over the course of a week from 19:40 Friday August 13, 2010 through 03:00 Saturday August 21, 2010, which resulted in 186 hours of monitoring.

2.1 Harbor Vessels

There is a wide variety of vessels in the harbor and a complex classification and relational hierarchy can be constructed based on hull features and size [8, 5]. However, the classes selected in this work relate to the visual appearance rather than strict marine identifiers. Visual categories were selected


(a) Water Taxi


(b) Cabin Cruiser


(c) Motorboat


(d) Raft (Rigid-Hulled Inflatable)


(e) Kayak


(f) Rowboat


(g) Canoe


(h) Paddleboard


(i) Sailboat

Figure 1: Annapolis Harbor Marine Vessels

to provide consistency with the computer vision literature to ensure comparison was applicable. In addition, absolute vessel size information, critical for marine classification, is not generated since the camera system was not calibrated. The omission of size information makes the proposed detection system more scalable since it is possible to directly apply the vessel detectors to alternative views. However, this also makes the classification task more difficult because vessels of different sizes can be confused. (In this case, scale can be addressed during the detection stage by selecting the appropriate scale for a particular boat).

The Annapolis Harbor dataset consists of nine vessel categories, which include water taxi, cabin cruiser, motorboat, raft, kayak, rowboat, canoe, paddleboard, and sailboat, presented in Fig. 1. The nine vessel types were chosen to be general enough for wide applicability at many U.S. coastal regions. However, there are a few classes which are specific to Annapolis. Water taxis have a unique canopied top and are a favorite method to travel around the harbor. Paddleboards, while not unique to Annapolis, are a popular recreational water activity and were specifically included because they appeared often in the video collected. Complete definitions and example images from the Annapolis dataset for each of the vessel classes can be found in Appendix A.

2.2 Annotation Scheme

The Annapolis data was annotated to create training and test datasets. The annotations included:

- vessel type {`cabin_cruiser`, `canoe`, `kayak`, `motorboat`, `paddleboard`, `raft`, `rowboat`, `sailboat`, `water_taxi`}
- bounding box $[x, y, width, height]$
- occlusion type {`none`, `masts`, `partial`, `full`}
- lighting type {`day`, `night`, `tough`}
- and optional notes.

The occlusion field is used to indicate the difficulty of the recognition task. The `none` value meant there was little to no occlusion. The special `masts` value was used to denote when only sailboat masts (foreground of harbor view) occluded the view, resulting in thin strips of occlusion. The `partial` occlusion tag indicated when 25 - 60% of the vessel was covered while `full` occlusion meant the example could not be easily determined without contextual clues.

The lighting field was used to make distinctions between the different times of the day. The `tough` label was applied to boat instances that suffered from severe lighting effects typically associated with sunrise and sunset conditions. These could include deep shadows or saturation which made visual identification extremely difficult for the human labeler.

The optional notes field was left open for any additional characteristics of importance. This was included to allow for supplementary data for further classification refinement. An example of an optional note is the term "`double`" to indicate a two-person kayak.

The annotation was performed manually by a single participant. There were no strict rules (or measurements of occlusion percentage) on any of the annotation fields, but only guidelines that represented the labeler's best judgment.

Appendix B presents an alternative annotation that accounts for the vessel parts (e.g., hull, sail, etc.), size, as well as its action based on the United States Coast Guard Navigation Rules [8]. Although this alternative annotation scheme is consistent with maritime categorization and terminology, it was not utilized in this study because of its dependence on size information. Since vessel size was not known, it would not be meaningful to require distinction from the visual classifier and would prevent comparison with the object classification literature.

2.3 Vessel Exemplars

Exemplar images were manually selected from Annapolis dataset for four of the 9 vessel classes in this preliminary study. Every effort was made to select a wide range of colors and orientations (direction of travel) to try to span the vessel appearance space. Exemplar selection preference was given toward instances that were larger and more in focus. Due to the Annapolis webcam optics, the center of the image was slightly out of focus while the left-hand side was sharp. Therefore, the majority of exemplars were selected from instances that were located in the bottom left of the image (bottom to be closer to the camera for higher resolution). Figures 2 – 5 show all 95 exemplars for each of the four vessel types.

2.4 Training

The Annapolis annotations were divided into training and test sets. The training set was made up of hour sets 11–17 which all correspond to a single Saturday morning from 06:30 – 13:30. Only a small fraction of the training data was selected as exemplars, as previously mentioned, while the vast majority was used for validation of the E-SVMs. (The large number of training samples are for future studies that require many positive examples).

One of the advantages of the E-SVM architecture is that only a single positive training exemplar is required. A large negative training set is required for generalization. The negative dataset consisted of images taken from the SUN Database [9] as well as frames of Annapolis video with no boats present. The full negative dataset consists of 17,506 images with 17,425 coming from SUN and the remaining 81 from Annapolis. These negative images provide the millions of negative sub-images required for SVM training. (Future work should expand on the number of Annapolis specific negatives for better performance).

A full characterization of the training database and distribution of data between classes is presented in Table 1. 10,290 out of a total of 23,815 frames during the training hours contained at least a single example of one of the nine Annapolis vessel types. Almost 2000 training instances

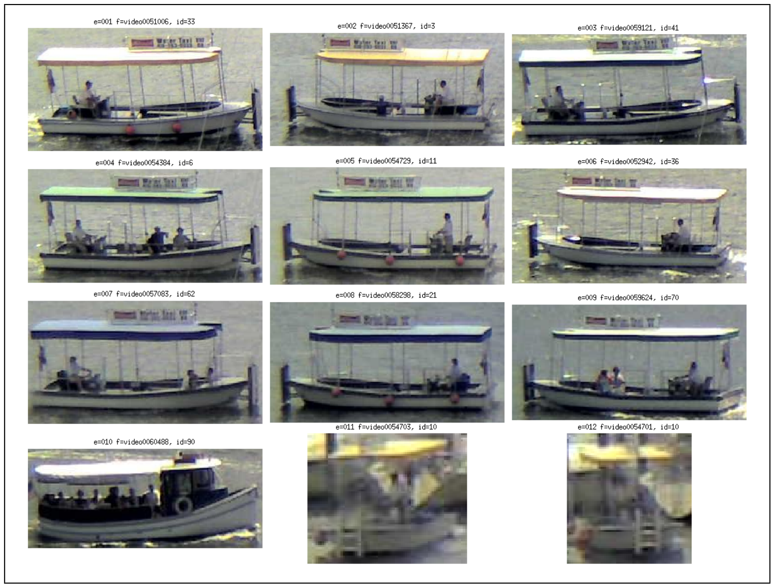


Figure 2: Water Taxi (12 Exemplars): The examplars highlight the characteristic canopy above the boat with a taxi sign. The fourth row, first column shows a water taxi variant that occurred only one time. The last two images show the boat approaching the camera. These were across the harbor and therefore have low resolution and quite a bit of clutter in the background.

Figure 3: Cabin Cruiser (12 Exemplars): Cabin cruisers are large vessels that have a second level cabin. It is possible that the top left image was mislabeled as multi-level cabin cruiser rather than a motorboat with a canopy. This may contribute to erroneous detections.

came from frames with more than a single vessel present. The *Exemplars* column indicates the number of exemplars that were manually selected. The *Frames* column gives the number of frames that had at least one example of the vessel. The *Instances* column gives the number of unique occurrences of the vessel. This is the number of times a boat of the given type appeared in a frame. (There are more instances than frames since multiple unique boat instances can be in a

Figure 4: Sailboat (35 Exemplars): While in the harbor, the sails were not typically opened but rather lowered in a distinct horizontal bar. Effort was made to select exemplars both from the profile as well as at other orientations. There were no examples approaching the camera.

single frame). The *Truncated* column counts the number of instances that were partially out of the screen. The *Occlusion* columns count the number of instances with a given occlusion label. The last column, *Best*, are examples that were not truncated and had no occlusion. These were the only samples used for validation and calibration of the E-SVM models.

2.5 Testing

The test set was constructed by selectively sampling different hours of the Annapolis dataset. The test set was annotated to cover the same Saturday as training as well as an hour from every other day during the week of data collection. The Saturday test data contained 3 hours between 13:30 – 16:30 and data between 17:30 – 18:30 hours of subsequent days, where lighting was generally favorable. The testing set also included the 12:30 – 13:30 hour of Sunday when it was raining. The test data had every 10th frame annotated (as opposed to each frame as in the training set) for a total of 3455 images. The distribution of data between the classes in the testing dataset is presented in Table 2. The test set contained a total 5048 labeled instances.

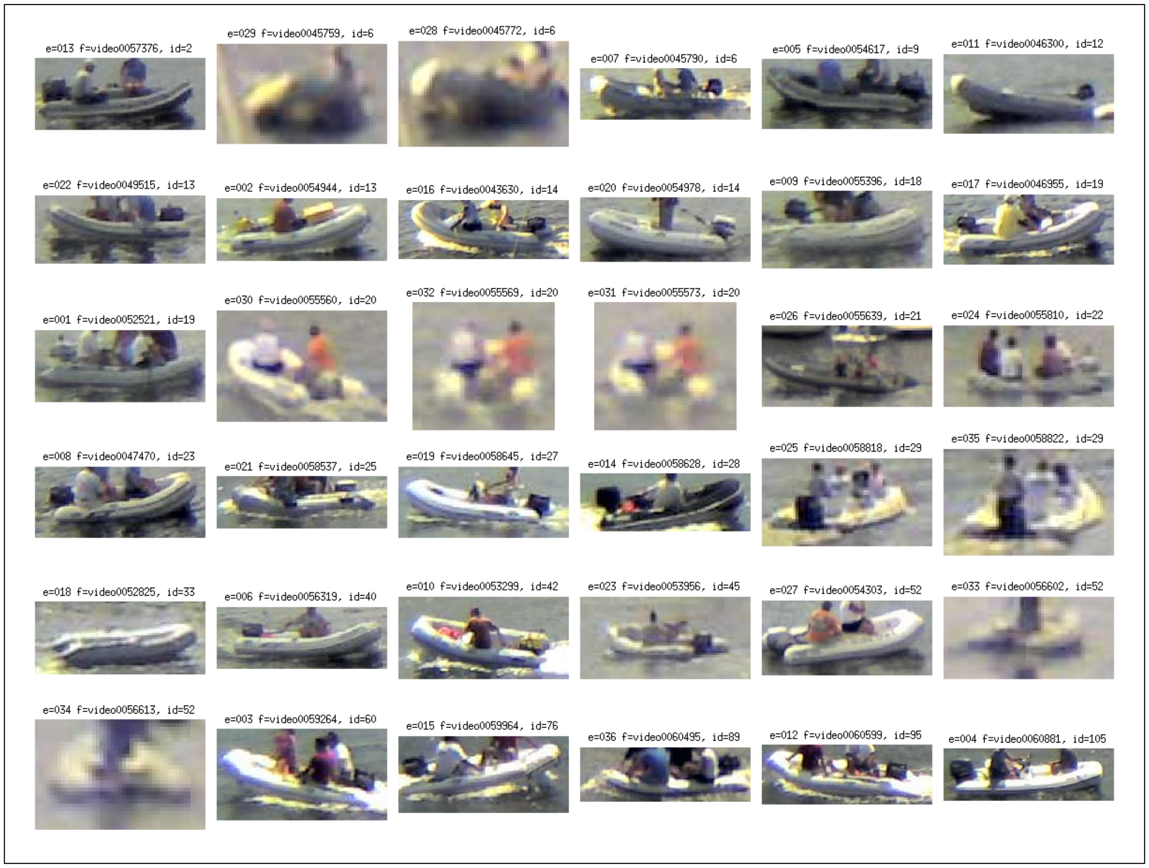


Figure 5: Raft (36 Exemplars): Rafts are small personal vehicles that could have various number of passengers. The raft area was selected to contain only the rubber hull and not the people. Notice that orientations were typically left-right and the up-down directions were of much lower resolution.

3 Results

The following section presents results of the E-SVM object classification system on the Annapolis Harbor data. Four (water_taxi, cabin_cruiser, sailboat, and raft) of the nine vessel types have been used in the evaluation, which examines the effectiveness of both localization and classification.

3.1 Evaluation Criteria

The performance of the Exemplar-SVM was evaluated on the Annapolis Harbor test set using the Pascal VOC detection protocol based on the precision/recall curve. Precision is the fraction of retrieved instances that are relevant, whereas recall is the fraction of relevant instances that are retrieved.

$$\text{Precision} = \frac{tp}{tp + fp} \qquad\qquad \text{Recall} = \frac{tp}{tp + fn}, \tag{1}$$

where tp is a true positive (correct classification), fp is a false positive (incorrect classification), and fn is a false negative (example that was missed).

The quantitative measure of performance is the average precision (AP), which is the area under the precision/recall curve. A detection is considered a true positive if the area of overlap a_o with

Table 1: Training Data Statistics

Class	Exemplars	Frames	Instances	Truncated	Occlusion none	Occlusion mast	Occlusion partial	Occlusion full	Best
cabin_cruiser	12	876	907	154	466	95	340	6	342
canoe	0	229	229	8	121	36	49	23	117
kayak	0	1979	3028	119	1792	190	710	336	1716
motorboat	0	3695	4274	341	2641	412	1163	58	2442
paddleboard	0	292	346	8	278	12	28	28	272
raft	36	1198	1256	41	864	49	252	91	832
rowboat	0	202	210	13	143	15	43	9	137
sailboat	35	1534	1615	177	1265	65	281	4	1097
water_taxi	12	285	292	38	171	29	92	0	150
Total	95	10290	12157	899	7741	903	2958	555	7105

Table 2: Testing Data Statistics

Class	Exemplars	Frames	Instances	Truncated	Occlusion none	Occlusion mast	Occlusion partial	Occlusion full	Best
cabin_cruiser	12	103	103	20	68	6	29	0	58
canoe	0	10	10	1	9	0	1	0	8
kayak	0	384	585	20	408	29	110	38	397
motorboat	0	1340	1986	145	1326	141	500	19	1233
paddleboard	0	64	80	2	52	5	15	8	51
raft	36	213	227	4	169	12	28	18	166
rowboat	0	18	18	1	13	0	5	0	12
sailboat	35	1202	1702	101	1165	91	427	19	1084
water_taxi	12	334	337	46	221	26	89	1	197
Total	95	3668	5048	340	3431	310	1204	103	3206

the ground truth bounding box is greater than 50%.

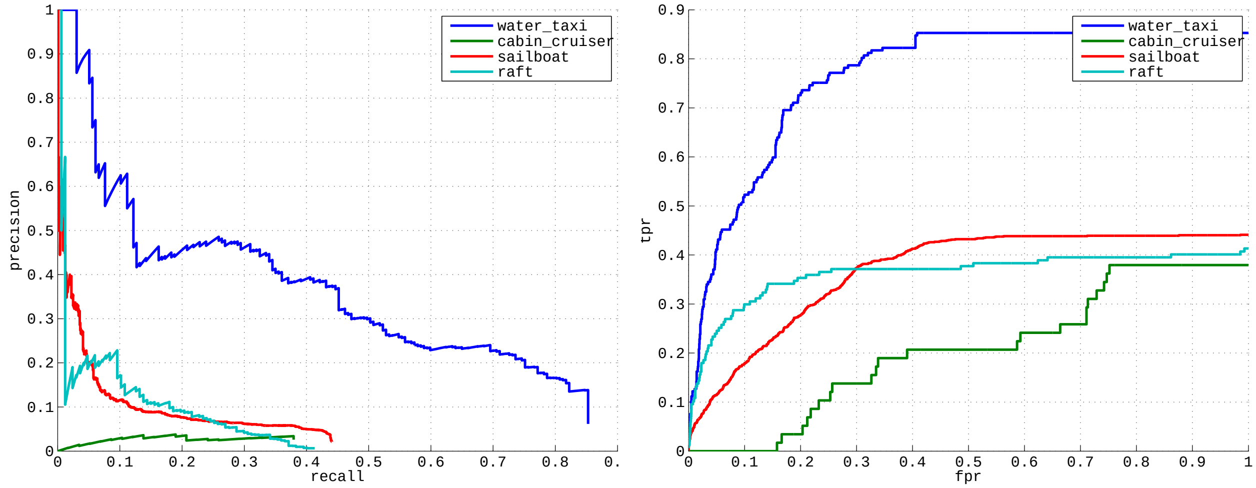
$$a_o = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} \tag{2}$$

where the predicted bounding box is B_p and the ground truth bounding box is B_{gt}.

Detections are generated using the sliding-window approach to visit each pixel in the image. Since nearby pixels will have a similar detector response, a non-maximum suppression (NMS) technique must be used to thin out these multiple responses as well as suppress spurious responses. However, instead of using standard NMS to average overlapping bounding boxes [10], an exemplar co-occurrence pooling method is used. A context feature is generated for each detection to combine the SVM scores of nearby (overlapping) detections by weighted sum and generate the final NMS detection score [11].

3.2 Annapolis Harbor Performance

The precision/recall and ROC curves for each of the classes is presented in Fig. 6 and the final AP results for the Annapolis vessel types are shown in Table 3. The 0.0853 mean AP (mAP) of the four Annapolis vessels types is significantly lower than the 0.227 value obtained by Malisiewicz et



(a) (b)

Figure 6: (a) Annapolis Precision/Recall Curves and (b) ROC Curves

Table 3: Vessel Detection Results (Average Precision)

	water_taxi	cabin_cruiser	sailboat	raft	mAP
AP	0.3362	0.0136	0.0502	0.0515	0.0853

al. [11] for the 20 PASCAL VOC categories. However, the water_taxi had a similar AP value of 0.3362 as the top performing classes in the VOC challenge. The performance on the sailboat and raft categories was similar to the VOC dog and cat categories. The cabin_cruiser instances had the lowest performance of the Annapolis classes with an AP of 0.0136 because the E-SVM was not able to distinguish them from the other three vessel types.

Many of the errors came not in the detection of a boat but the classification of the type of vessel. The cabin_cruiser often triggered on motorboats or sailboats. The similarity between object classes causes confusion and makes Annapolis vessel classification a difficult task. Marine vessels have a number of similar components or parts, such as the hull, cabin, sail, etc., which are shared among different classes and may cause confusion. The evaluation protocol in this work looked at each individual vessel classifier in isolation, meaning the results presented do not account for multiple detectors firing on the same object (e.g., motorboat and cabin_cruiser detectors firing on the same detection). An operational system would need to make a final decision after pooling and examining the results from each of the individual vessel detectors. The final performance can be improved at the pooling stage by utilizing a hierarchical classification scheme that accounts for strong similarities between object categories (such as cabin_cruiser and motorboat both being powered vessels) [1].

Another source of the poor classification performance was false detector firings on the water surface. The texture caused by small waves on the surface of the water caused the detectors to inadvertently fire. Even worse, the detections caused by the waves had very high detection scores at times. This indicates the classifiers may need to be re-trained with the inclusion of more negative wave examples. Negative samples should be taken from the Annapolis Harbor videos to include both

- Annapolis training frames that contain no vessels (to expand water surface negatives) and

- frames that have no examples of the particular class (to use all vessels as negatives - e.g. motorboats are negatives for cabin_cruiser).

3.3 Performance with Occlusion

Classification accuracy was also evaluated at various levels of occlusion due to its affect on tracking performance. Each vessel example in the Annapolis dataset was given a label to indicate the amount of occlusion present in the particular frame as described in Section 2.2. The performance curves for the different occlusion levels of the water_taxi class are plotted in Fig. 7 in increasing difficulty. These curves are are representative of the other vessels as well. However, the smaller vessels have higher rates of occlusion because they are more easily occluded by scene elements. When examining the ROC curves, there is only a small degradation in performance due to `mast` occlusion. However, there is a much larger performance drop when the occlusion increases to `partial` or `full`. There is little difference in performance between `partial` and `full` occlusion because there are very few full occlusion instances.

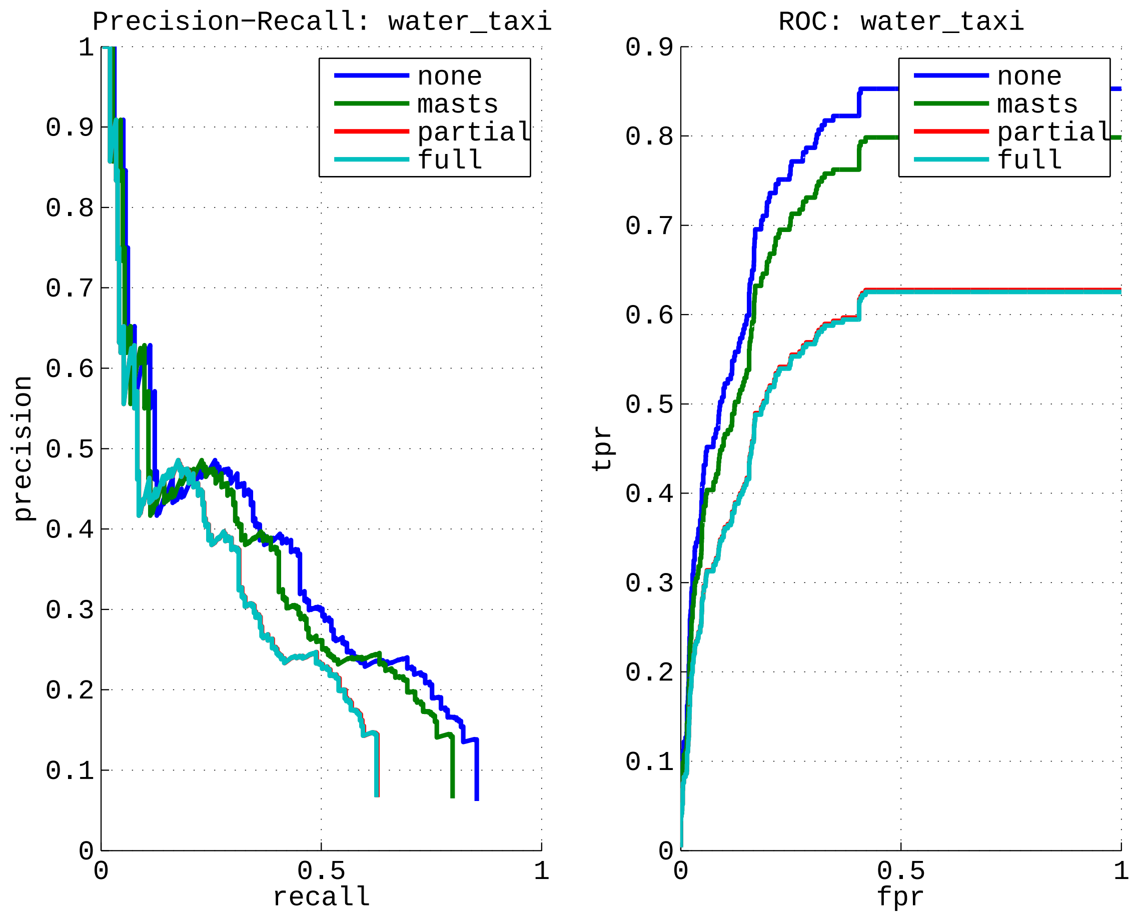


Figure 7: Water Taxi performance curves for different occlusion situations show the degradation in performance as the classification problem increases in difficulty.

4 Future Work

This report presents only the initial evaluation of the E-SVM for maritime vessel classification. A more complete characterization of the performance and difficulty of the Annapolis Harbor data will be possible after evaluating the five remaining vessel types. The motorboat and kayak classes make up 60% of the entire training set, and their inclusion in the evaluation will greatly influence the average precision.

4.1 Exemplar Influence

It will be important to study how the selection and the number of exemplars affects classifier performance. In essence, the E-SVM is a nearest neighbor classifier, which means there should be many examples (enough to adequately map out all the appearance variations). In this preliminary study, only a very small number of exemplars were hand selected for training. The assumption was that the vessels would not be dramatically different in appearance. This seems to hold true for water_taxi which has strong performance with only six exemplars, but not for the other three classes that were evaluated. More automated methods, such as random selection, would decrease human dependency and involvement and improve usability.

In addition, the impact of specific exemplars will need to be studied. After validation and calibration of the E-SVM during the local score pooling step of training, the final detection scores seem to be disproportional with some being quite large. The mismatch has major implications for detection threshold selection and maximizing performance. In these experiments, the top scoring detections of a particular class tended to come from just a few exemplars (e.g., exemplar 6 was selected as the the best match for 10 out of the top 20 detections). This may indicate a strong exemplar or that there is a problem in the normalization of E-SVM scores within a vessel type.

4.2 Testing Protocol

The testing protocol should be adjusted to evaluate every frame in the test set rather than only the frames that had a specific vessel present. The results presented only tested frames that contained an example of one of the four trained models for speed considerations. Therefore, the performance of the system when operating in real-time for tracking can be expected to be lower because of the inclusion of false positives in the negative frames.

4.3 Functional-Mapping of Vessels

However, the final detection/classification performance can be improved under real-time operation by managing the pooling between vessel classifiers by utilizing a hierarchical classification scheme that accounts for strong similarities between object categories (such as cabin_cruiser and motorboat both being powered vessels) [1]. While not providing clear visual distinction between classes, this would allow for more functional characterization compatible with threat analysis.

5 Concluding Remarks

One critical component of fleet security is robust threat assessment. In order to accurately assess threat, it is crucial to know the enemy vessel types to assess their capabilities. This work presents a preliminary study of a vision-based vessel identification and tracking system. Marine data was collected from the Annapolis Harbor and models of four different vessels were trained using the

Exemplar-SVM method to learn with a small number of positive examples. Initial results indicate detection is sufficient for tracking purposes. However, classification performance between the boat types is quite variable with high average precision for water taxis but low precision for cabin cruisers. Further studies that examine all nine of the Annapolis vessel classes will need to be performed before completely characterizing the E-SVM system's performance as well as comparative analysis between other state-of-the-art object classification methods.

Acknowledgments

The authors would like to thank their colleagues at the NRL and Knexus Research for their support and advice. This work was performed through the Office of Naval Research (ONR) Summer Faculty Research Program (SFRP) administered through the American Society for Engineering Education (ASEE). Thanks to ONR 31 for their support of this research.

References

[1] B. Auslander, K. M. Gupta, and D. W. Aha, "A comparative evaluation of anomaly detection algorithms for maritime video surveillance," in *Proc. Society of Photographic Instrumentation Engineers*, 2011.

[2] X. Wang, K. T. Ma, G.-W. Ng, and W. E. L. Grimson, "Trajectory analysis and semantic region modeling using nonparametric hierarchical bayeian models," *International Journal of Computer Vision*, vol. 95, no. 3, pp. 287–312, 2011.

[3] B. Auslander, K. M. Gupta, and D. W. Aha, "Maritime threat detection using probabilistic relational networks," in *Florida Artificial Intelligence Research Society Conf.*, 2012.

[4] D. W. A. A. M. Bryan Auslander, Matthew Molineaux and Q. Pizzini, "Towards research on goal reasoning with the TAO Sandbox," Naval Research Laboratory, Tech. Rep. AIC-09-155, 2009.

[5] K. M. Gupta, D. W. Aha, and P. Moore, "Case-based collective inference for maritime object classification," in *Proc. Inter. Conf. on Case-Based Reasoning*, 2009, pp. 443–449.

[6] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (VOC) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

[7] (2012) Spa creek webcam. Annapolis Yacht Club. [Online]. Available: http://www.annapolisyc.com/default.aspx?p=DynamicModule&pageid=281973&ssid=157918&vnf=1

[8] (2012) Navigation rules online. United States Coast Guard. [Online]. Available: http://www.navcen.uscg.gov/?pageName=navRulesContent

[9] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torrabla, "SUN database: Large-scale scene recognition from abbey to zoo," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2010, pp. 3485–3492.

[10] P. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, Dec. 2001, pp. 511–518.

[11] T. Malisiewicz, A. Gupta, and A. A. Efros, "Ensemble of exemplar-SVMs for object detection and beyond," in *Proc. IEEE Inter. Conf. Computer Vision*, 2010, pp. 89–96.

[12] (2012) Types of boats. Boats Depot. [Online]. Available: http://www.boatsdepot.org/boats/types/

A Visual Maritime Categories

The vessels in the Annapolis Harbor dataset were placed into 9 different visual categories. While the categories were designed to respect maritime distinctions, the appearance of a vessel was the major cue for determining its class. The 9 most often occurring vessels types were retained for the classification categories. Definitions of the vessel types comes from `http://www.boatsdepot.org` [12].

A.1 Water Taxi

A water taxi is a popular way to enjoy Annapolis during the warmer months. These provide transportation service through the Harbor (Downtown to Eastport) and avoids the need to find parking. The taxis can be boarded at City Dock and other public and commercial shoreline destinations such as Spa and Back Creeks. This transport vessel can carry a number of passengers and typically has a canopy for shade with a lighted taxi sign.

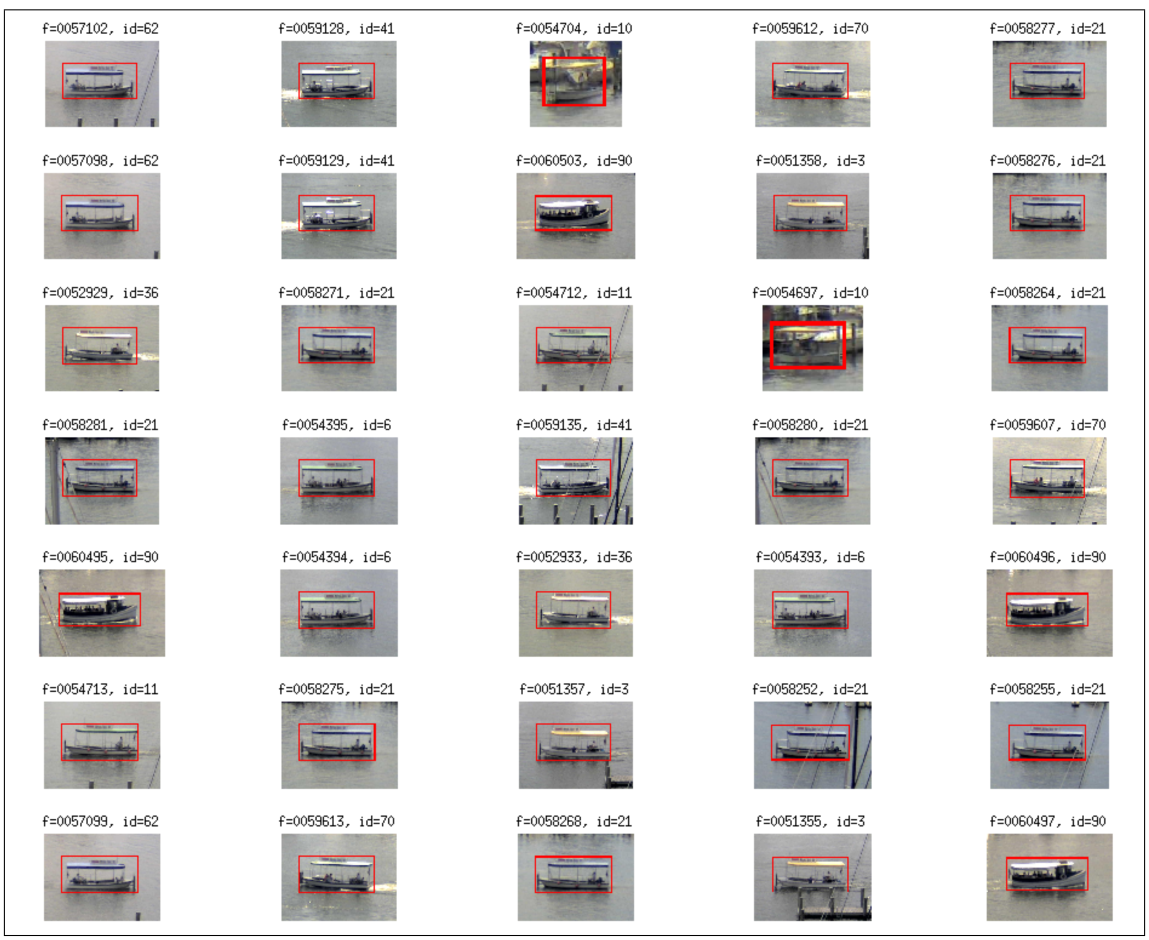


Figure 8: Water Taxi (12 unique vessels)

A.2 Cabin Cruiser

A large motorboat that has a cabin and plumbing and other conveniences necessary for living on board. They generally feature 2 to 3 sleeping areas, a galley and a separate head with shower. Cabin cruisers were distinguished by their secondary (or greater) compartments above the main deck.

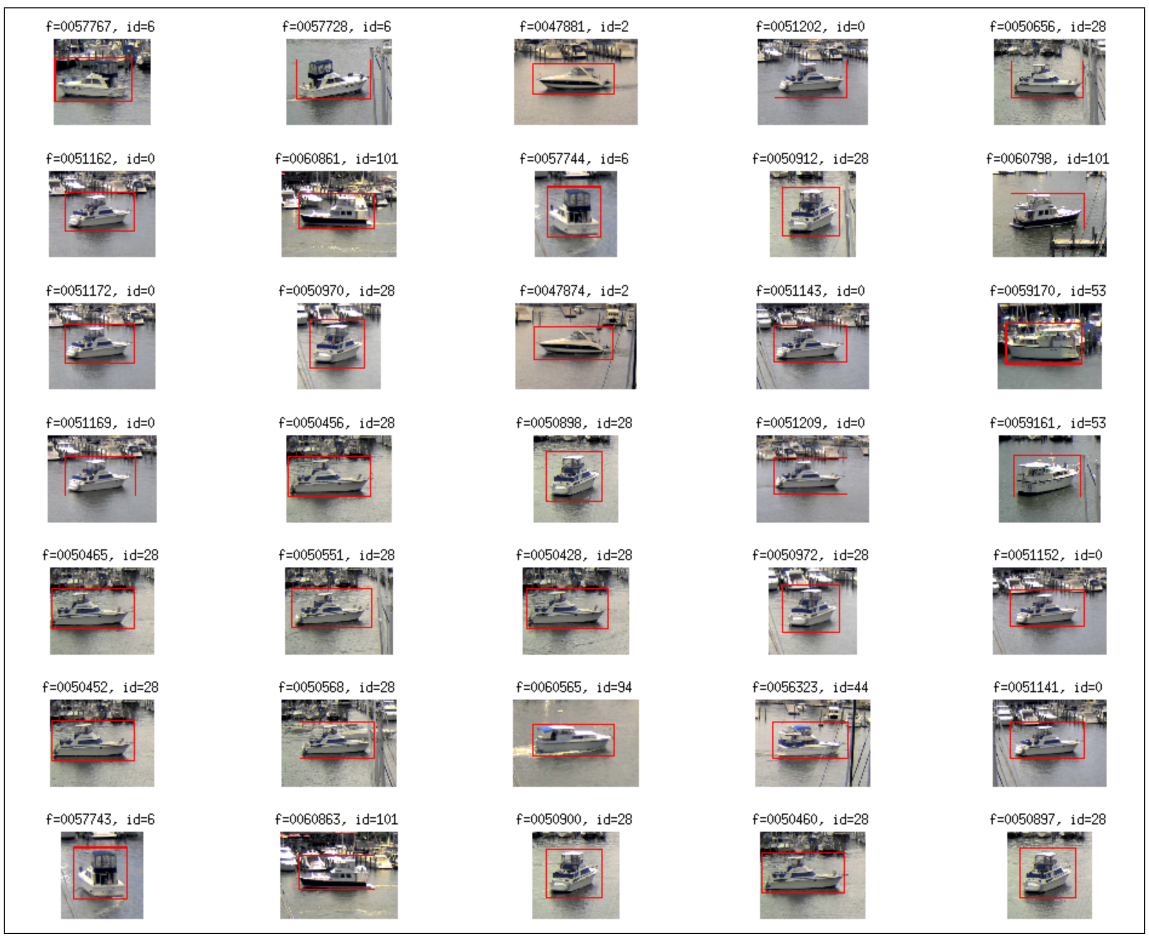


Figure 9: Cabin Cruiser (10 unique vessels)

A.3 Motorboat

Motorboat is the general label applied to a large class of vessels. These include small, medium, and larger vessels of multiple purposes but for the most part seem to be personal recreation vehicles such as fishing boats and speedboats used for towing. This is the most diverse vessel category in the Annapolis dataset and has the widest variety of appearance.

A motorboat in general is a vessel other than a sailboat or personal watercraft, propelled by an internal combustion engine driving a jet or a propeller.

A speedboat is a small motorboat designed to move quickly, used in races, for pulling water skiers, as patrol boats, and as fast-moving armed attack vessels by the military.

Figure 10: Motorboat (129 unique vessels)

A.4 Raft

A raft is a special type of boat, distinguished by the absence of a hull. Rafts are kept afloat either by buoyant materials such as wood, or by inflated containers. A rigid-hulled inflatable boat or RHIB is a light-weight but high performance and high capacity boat constructed with a solid, shaped hull and flexible tubes at the gunwale which makes it look very similar to a raft.

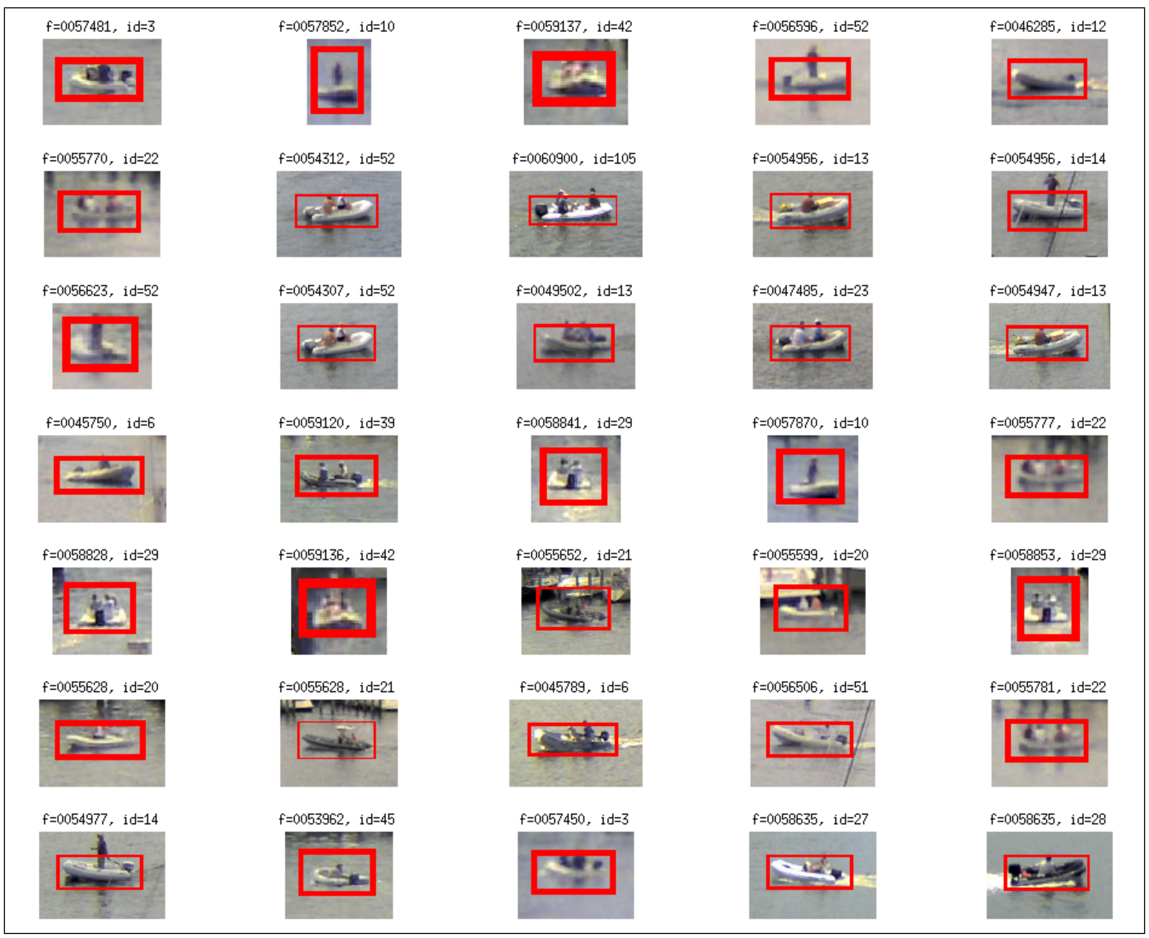


Figure 11: Raft (40 unique vessels)

A.5 Kayak

The term *kayak* is derived from usage in arctic languages where it describes a long narrow hunting boat propelled by single or double paddles. Today *kayak* has been most closely associated with boat designs derived from indigenous Greenland types. Greenland type kayaks are almost invariably propelled by a single paddler using a double bladed paddle.

This dataset contains both single and double occupancy kayaks which typically travel on tours in groups.

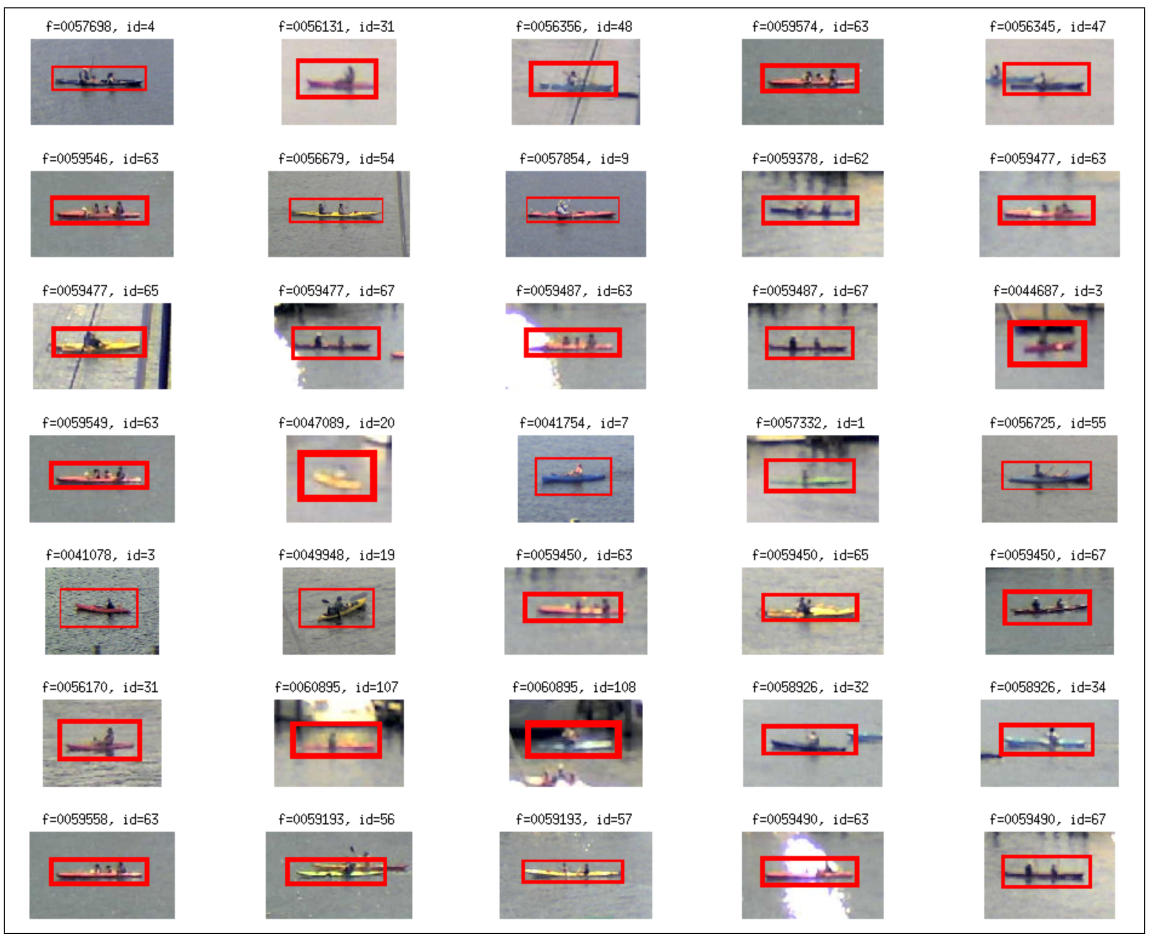


Figure 12: Kayak (65 unique vessels)

A.6 Rowboat

A small boat of shallow draft with cross thwarts for seats and rowlocks for oars with which it is propelled.

Figure 13: Rowboat (6 unique vessels)

A.7 Canoe

The canoe class is closely related to rowboats because it is usually propelled by means of paddles. However, this small, light, and maneuverable boat has a distinct appearance because it is pointed at both ends.

Figure 14: Canoe (4 unique vessels)

A.8 Paddleboard

Stand-up paddle boarding is a derivative of surfing where the rider uses large outrigger-type paddles to propel himself while standing on a long board.

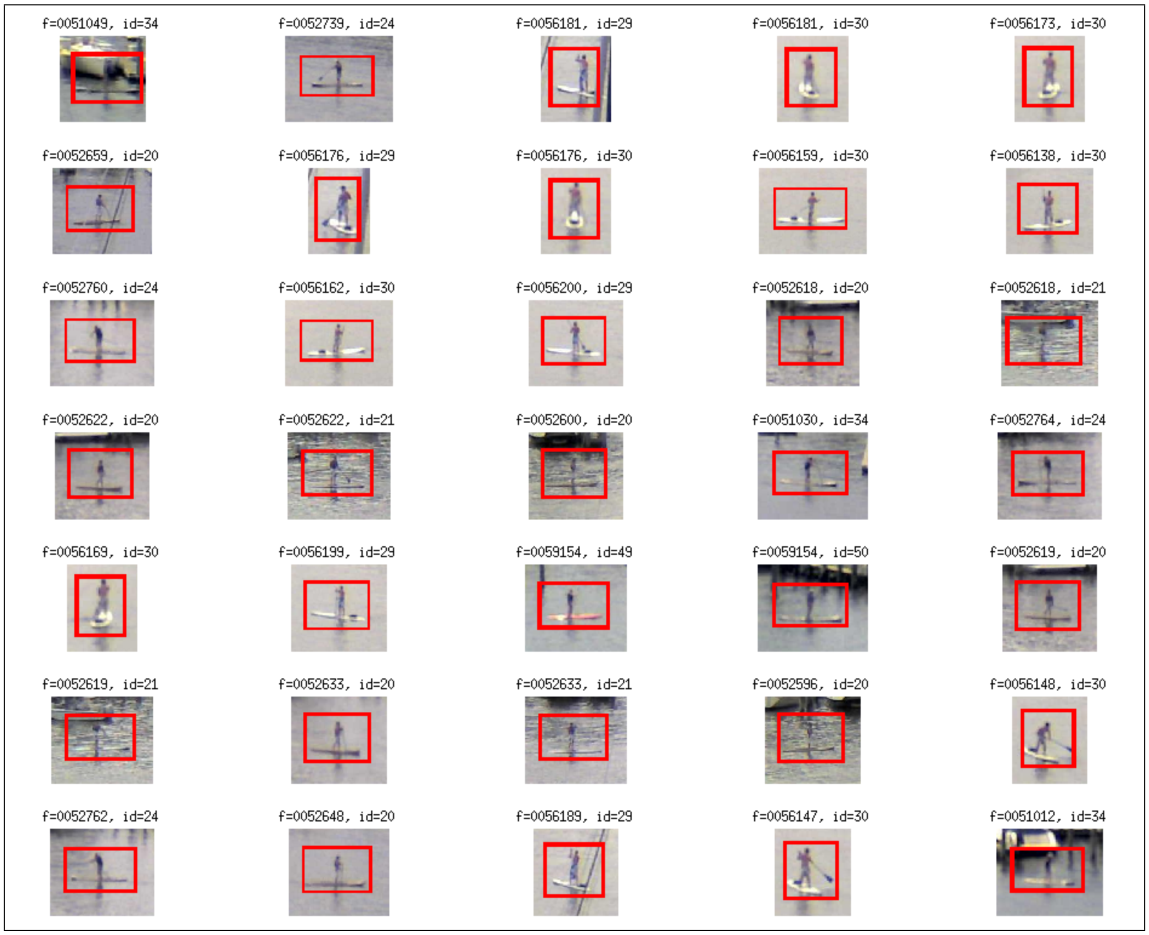


Figure 15: Paddleboard (4 unique vessels)

A.9 Sailboat

A sailboat is a relatively small wind-driven vessel used primarily for sports and personal purposes.

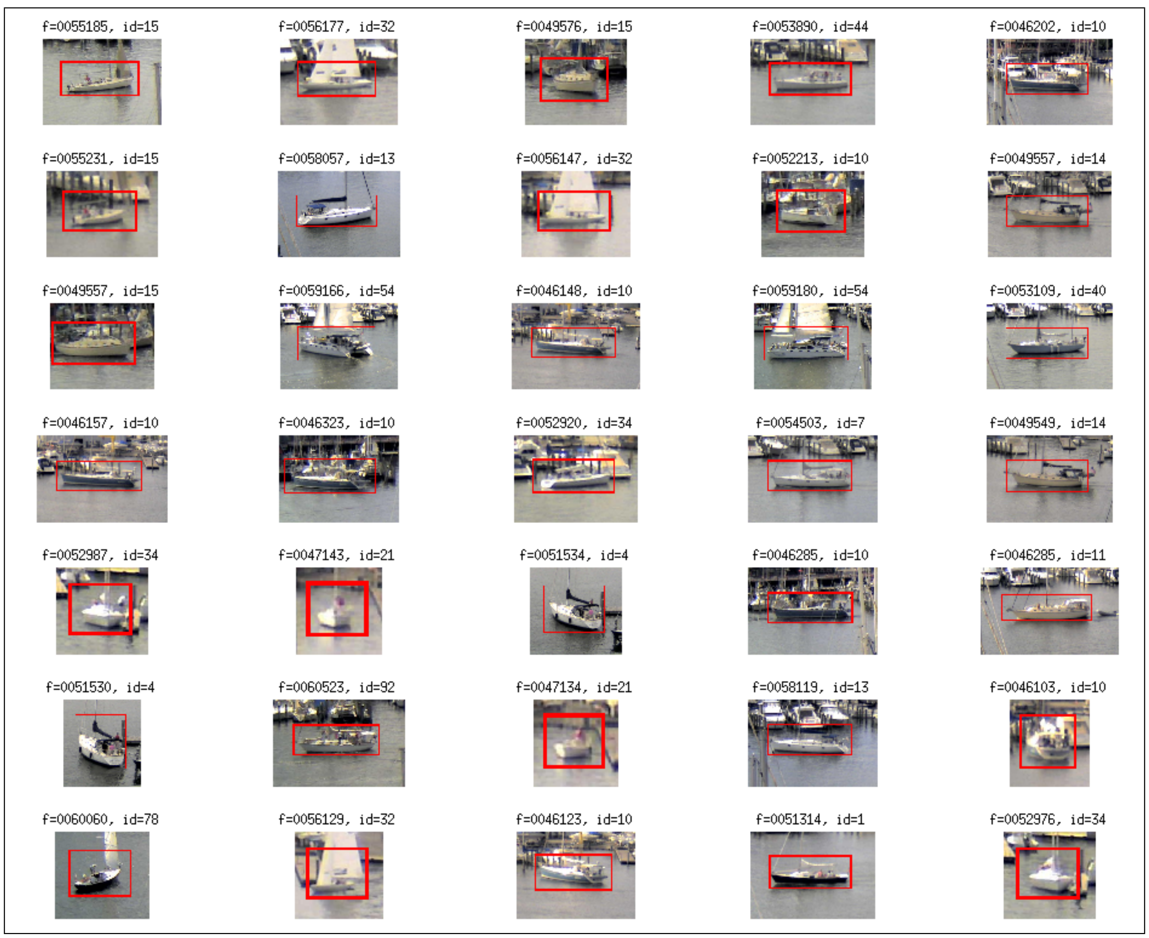


Figure 16: Sailboat (33 unique vessels)

B Categories based on U.S. Coast Guard Navigation Rules

Below are examples of annotations from the first day of the Annapolis dataset generated by Bryan Auslander. The annotations are based on a maritime ontology extracted from the United States Coast Guard Navigation Rules [8] which provide specification for lights based on vessel size and action. The class name is the union of 4 different labels. First is the type of ship {Personal, Commercial, Public, Cannopy, Sail, Mast, Crane}, second is size {small, medium, large, NA}, third is the propulsion type {sail, motor, manual, NA}, and the last is the speed {slow, medium, fast}. In addition (though not indicated in the class label), each instance is given an action code such as {Cruising, Docking, Unloading, Towing, etc.}.

The segments shown in red have been padded to provide more context. Therefore, the canopy class shown in Fig. 17 only contains the small center portion (e.g. blue cover in row 1 columns 1-3). In this labeling, the hull of the boat was the a main defining characteristic. An additional portion (e.g. canopy, sail, mast, etc.) could be linked to a hull.

There was difficulty using these annotations for the visual classification task because its primary purpose was as a size and action based hierarchy. A number of the labeled examples were occluded (e.g. extending outside of the image frame) or not even visible due to severe occlusion.

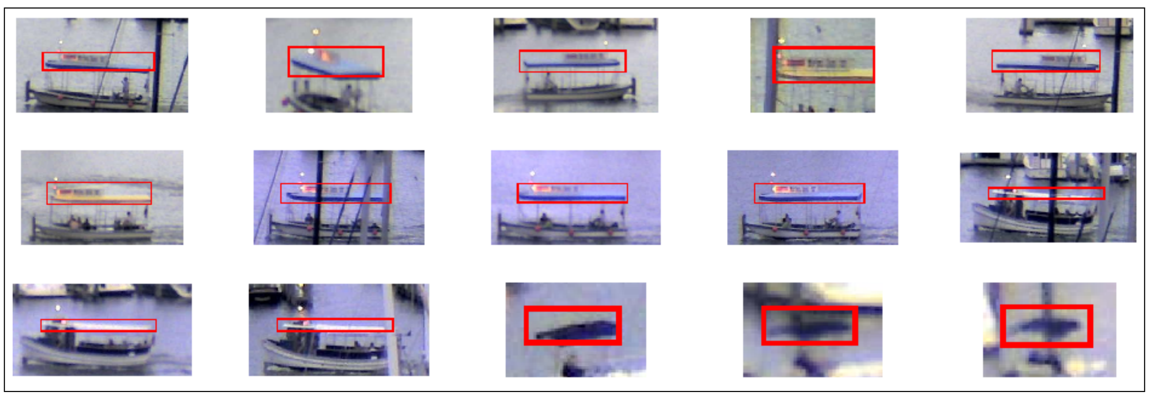


Figure 17: CannopyNANANA: Notice a large number of canopy images seem to come from the same boat (rows 2 and 3). They also are of low resolution and seem to have plenty of difficult background.

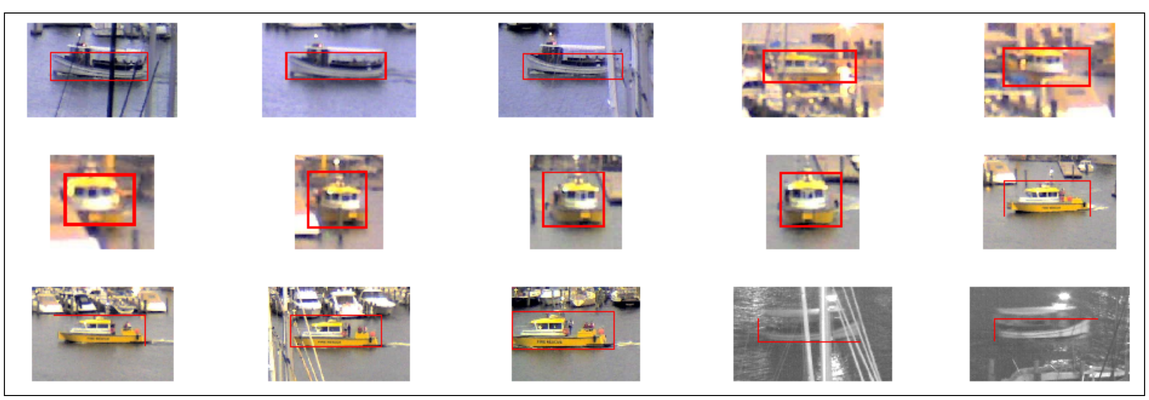


Figure 18: Commerciallargemotormedium: Notice a number of examples were obtained at night. There is blooming both from ship lights as well as reflections off the water surface.

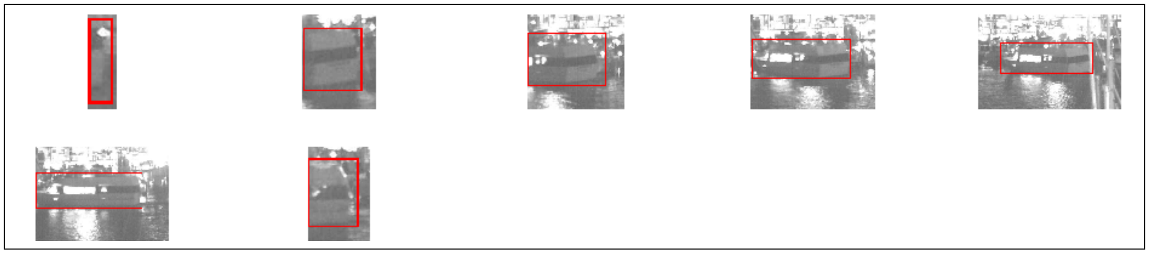


Figure 19: Commerciallargemotorslow

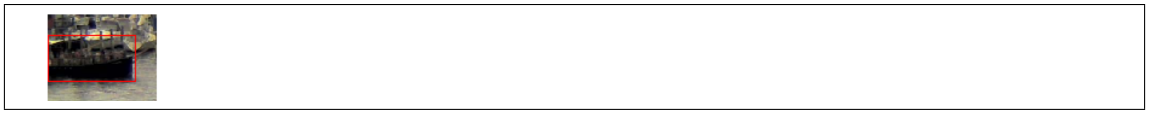


Figure 20: Commerciallargesailmedium

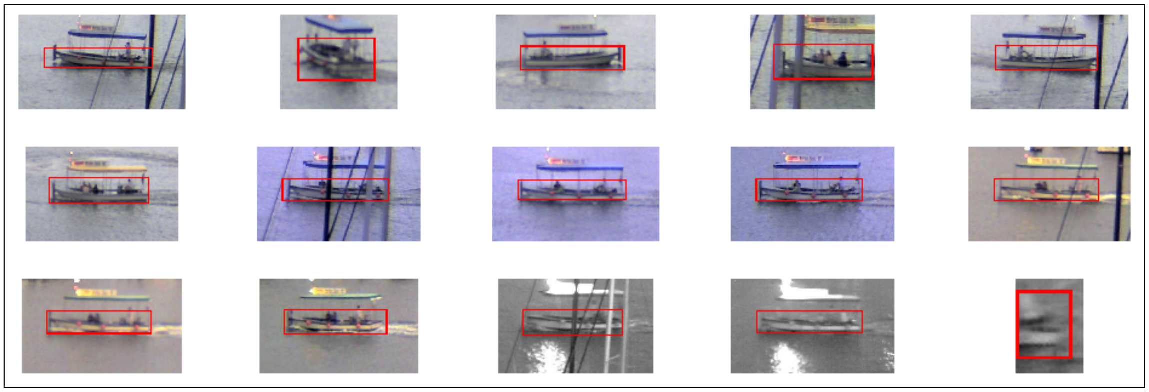


Figure 21: Commercialmediummotorfast

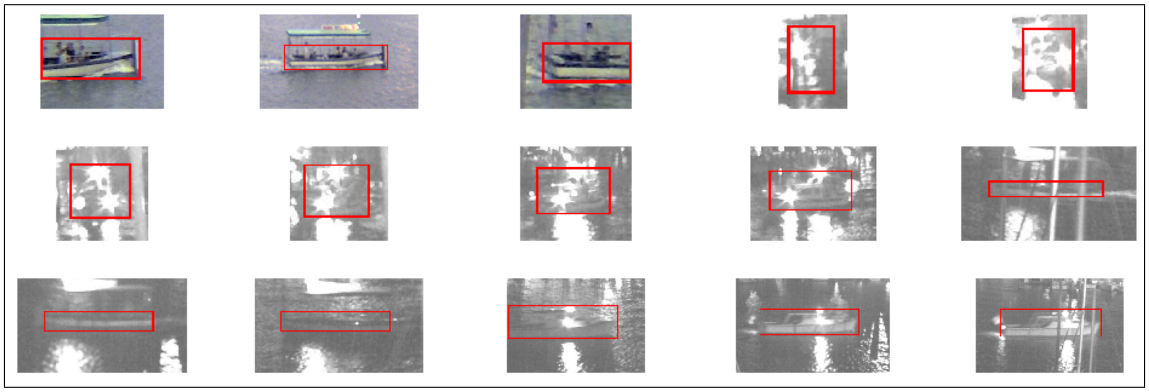


Figure 22: Commercialmediummotormedium

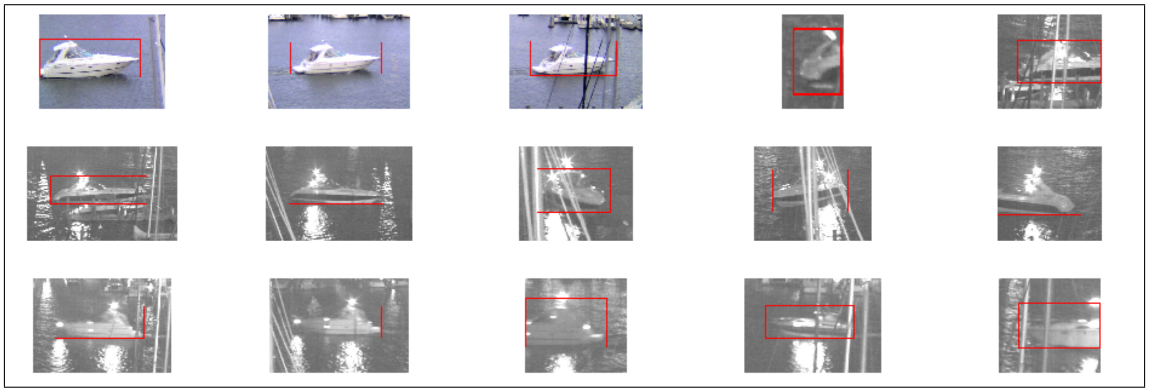


Figure 23: Personallargemotorfast

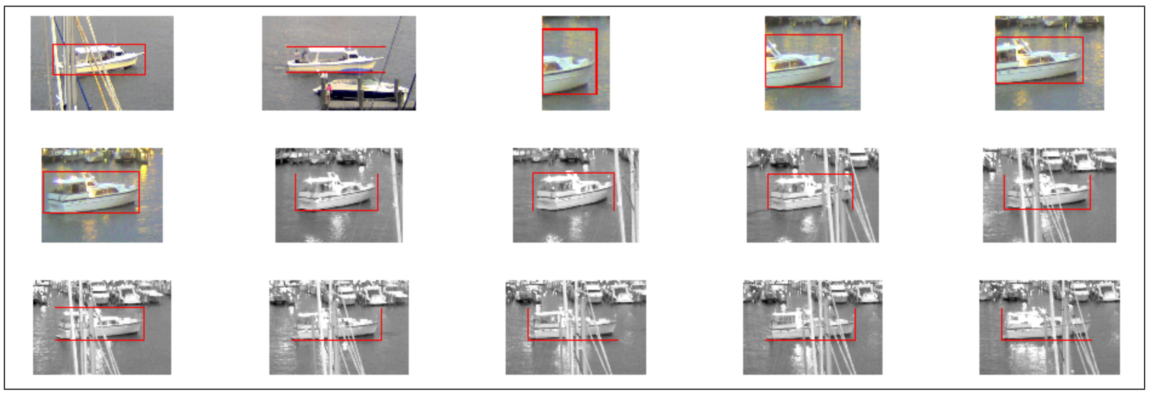


Figure 24: Personallargemotormedium

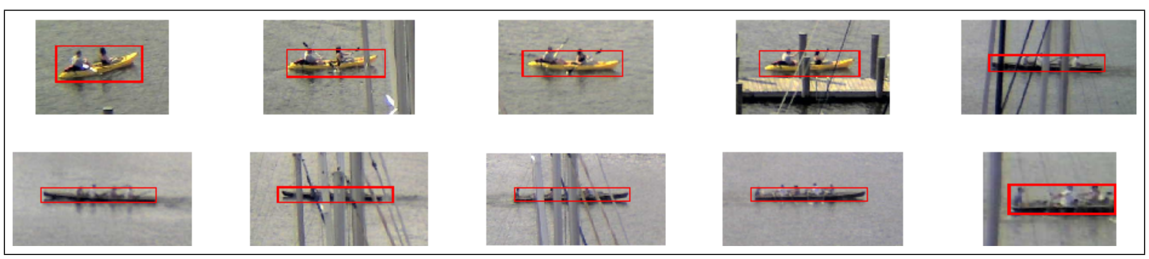


Figure 25: Personalmediummanualslow

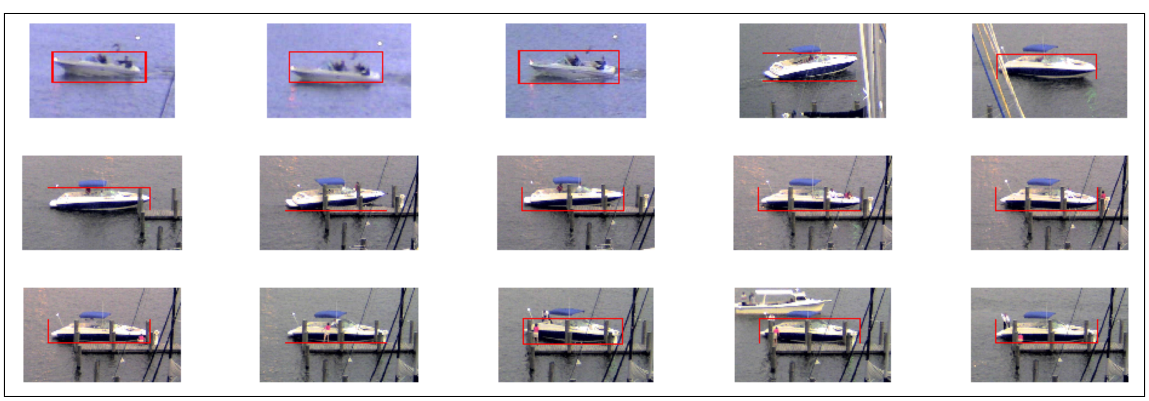


Figure 26: Personalmediummotorfast

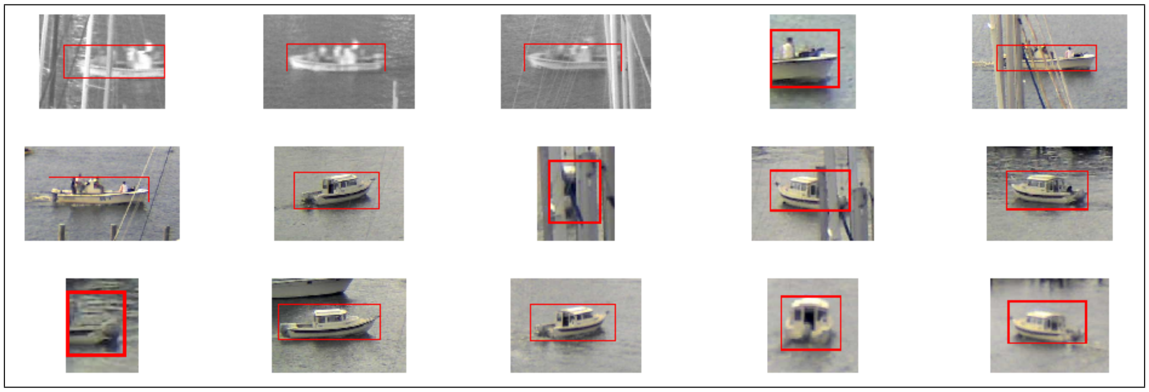


Figure 27: Personalmediummotormedium

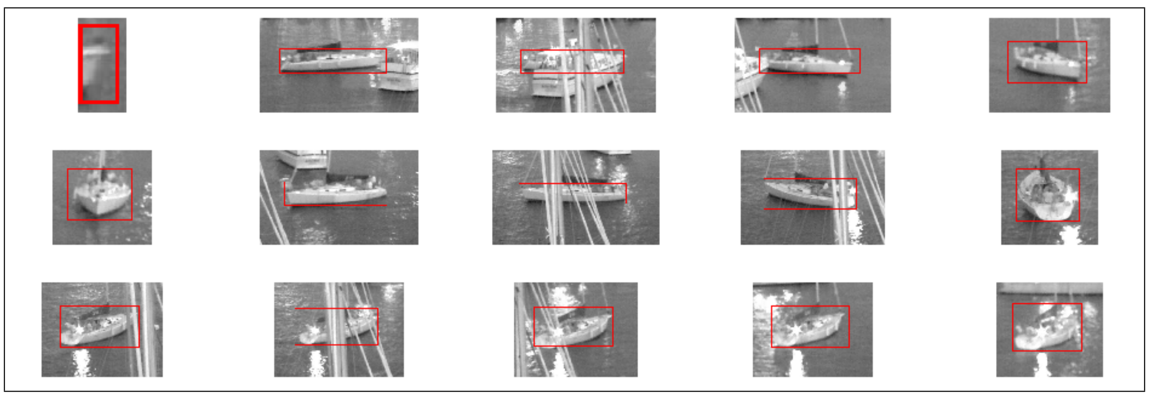


Figure 28: Personalmediumsailmedium

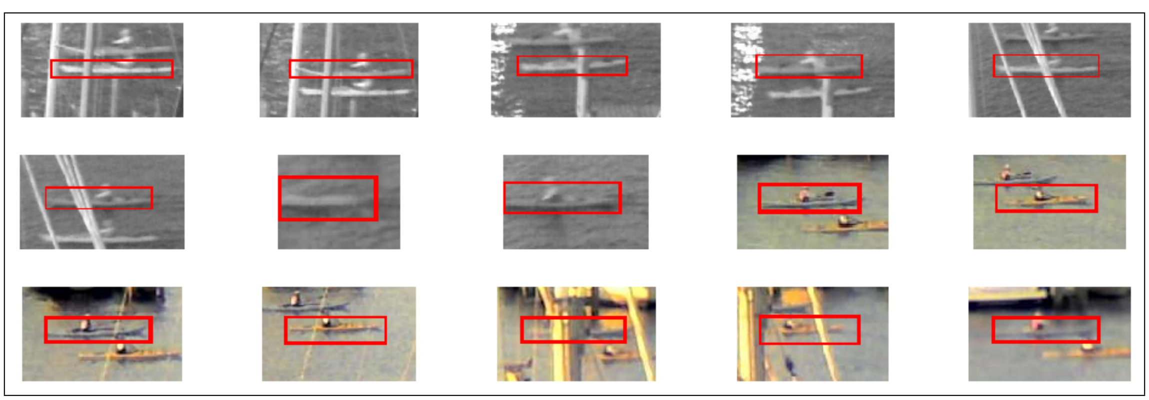


Figure 29: Personalsmallmanualslow

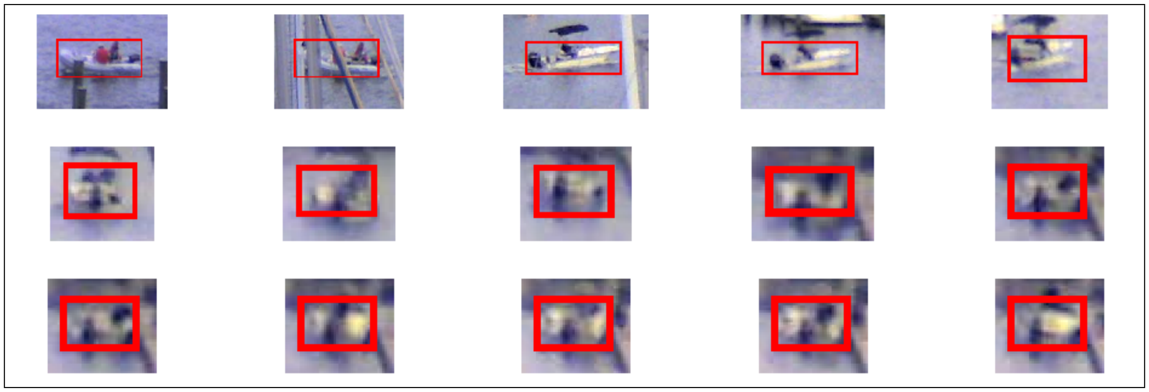


Figure 30: Personalsmallmotorfast

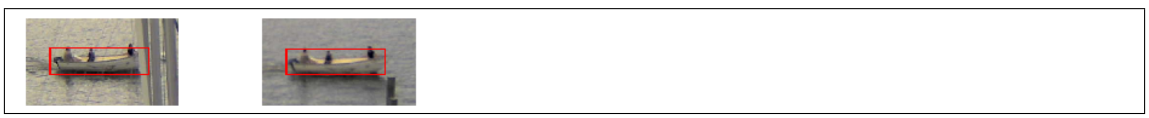


Figure 31: Personalsmallmotormedium

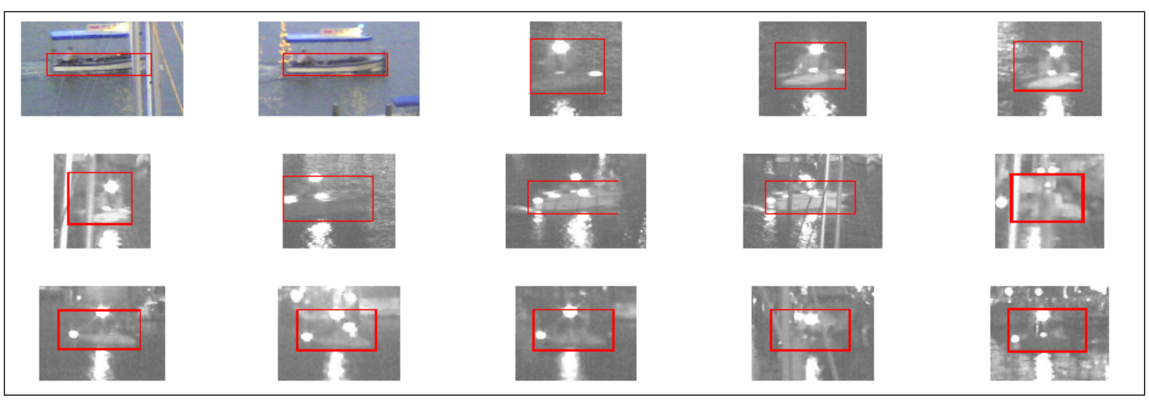


Figure 32: Publicmediummotorfast

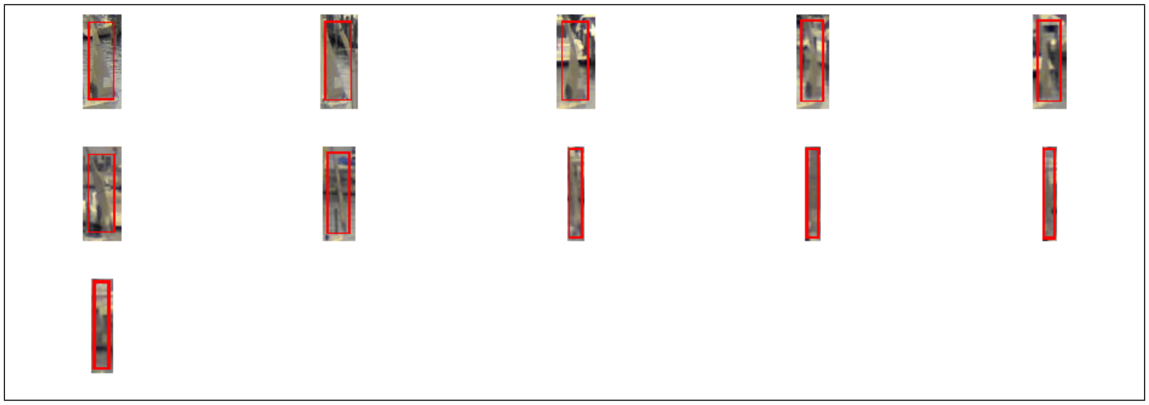


Figure 33: SailNANANA

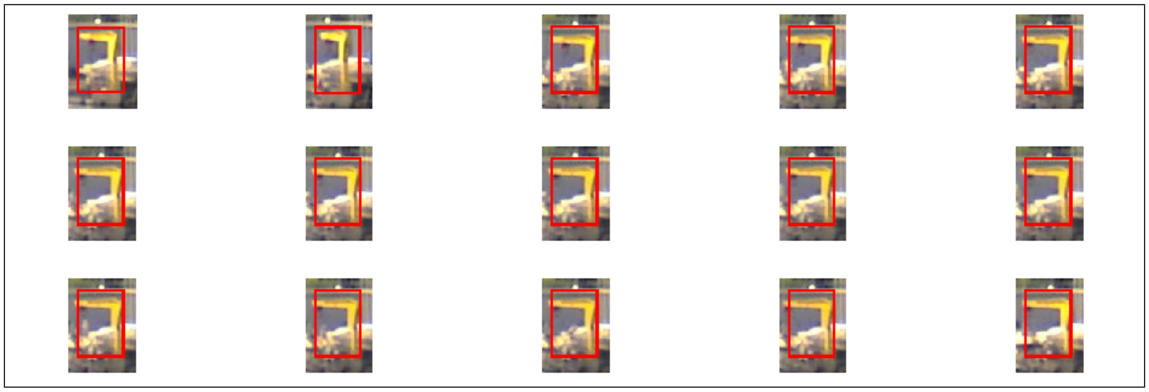


Figure 34: CraneNANANA

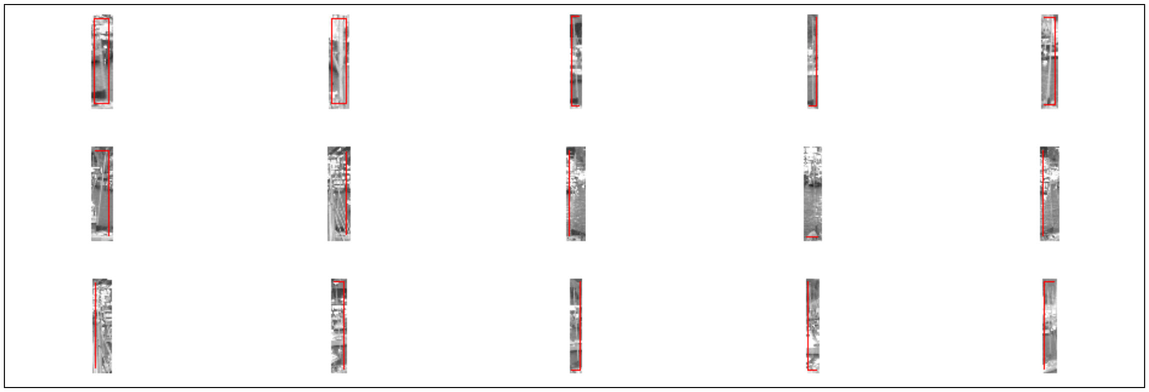


Figure 35: MastNANANA