On Human Perception and Automatic Target Recognition: Strategies for Human-Computer Cooperation

David P. Williams, Michel Couillard, and Samantha Dugelay
NATO Science and Technology Organization
Centre for Maritime Research & Experimentation (CMRE)
La Spezia, Italy
{williams, couillard, dugelay}@cmre.nato.int

***Abstract*—This work addresses the task of underwater object recognition in sonar imagery when both human operators and automated algorithms are available. We discuss the issues that have impeded previous attempts at automation, raise key insights related to human perception, present strategies to exploit the skills of humans and computers synergistically, and demonstrate the utility of the proposed approaches on a real object-recognition task employing actual humans acting as operators. Importantly, the strategies outlined here can be immediately adopted in existing (unautomated) target recognition systems with minimal cost, effort, and risk, while still achieving potentially significant performance gains. Moreover, this progress lays the foundation for the acceptance of still-further automated systems in the future. Experimental results are provided from a real mine-search exercise at sea, with recognition performance as a function of human operator effort given for various human-computer divisions of labor.**

I. Introduction

The detection and classification of objects in remote-sensing imagery is an important task that is common to many sensor modalities across diverse domains. Although automation has seen acceptance in certain civilian applications, high-risk (*i.e.,* life-or-death) object-recognition tasks involving military assets are still performed almost exclusively by human operators, often *in situ*. In order to remove humans from the threat and to cope with ever-increasing data-collection rates, automation will be imperative for future object-recognition systems. However, it has been acknowledged [1] that this transition away from human operators will be a gradual process that proceeds incrementally as trust in the automated systems grows. This caveat suggests that the path to future automation first requires development of methods by which humans and computers (*e.g.,* automatic target recognition (ATR) algorithms) can work in concert to achieve improved performance.

In this work, we discuss the issues that have impeded previous attempts at automation, raise key insights related to human perception, present strategies to exploit the skills of humans and computers synergistically, and demonstrate the utility of the proposed approaches on a real object-recognition task employing actual humans acting as operators. Although this work addresses these themes in the context of underwater target recognition in sonar imagery, the findings should be applicable to a wide range of object-recognition tasks in various types of remote-sensing data that seek to employ humans and computers in a cooperative manner.

The remainder of this paper is organized as follows. Flaws in earlier automation attempts are discussed in Sec. II, with this providing the motivation for the cooperative human-computer strategies proposed in Sec. III. Sec. IV describes a realistic at-sea experiment involving human operators and the detection and classification of underwater targets in sonar imagery; results for various human-computer combinations are also shown. Concluding remarks are given in Sec. V.

II. Flaws in Earlier Automation Attempts

The automation of target detection and classification tasks currently performed by human operators is a worthy goal. It is known that the successful completion – by humans – of these largely repetitive tasks is compromised by factors such as fatigue, boredom, distraction, and disinterest [2], [3]. However, earlier attempts intended to automate underwater mine recognition were resounding failures that wound up inhibiting the acceptance of automation [1], [4]. These failures can be attributed to an immaturity in the ATR algorithms at the time, which suffered from excessive false alarm rates and an inability to adapt to different environmental conditions or sites. But more fundamentally, these earlier automation attempts strove for arguably misguided goals and ostensibly overlooked matters related to human perception.

If an algorithm is designed with the intent to mimic the processes and decisions that a human performs, the best-case performance of the automation is essentially limited to only matching – and not surpassing – human performance (albeit peak performance when operating in optimal conditions, sans fatigue, *etc.*). The benefits engendered by the automation then pertain strictly to saving resources (namely, time and human effort) and enabling repeatable performance. Only by exploiting elements that fall beyond the limits of human perception can automation lead to performance that exceeds that of humans. For example, rather than developing features that attempt to capture cues that humans rely on for detection, other phenomena difficult or impossible to see with the human eye – but detectable by a computer – can be exploited. The complementary viewpoints provided by the human and computer then make the cooperative fusion of their efforts a logical approach. Therefore, when it is known that

human operators and automated algorithms will be employed cooperatively, research should be directed with this in mind. Specific example recommendations to this end, in the context of the underwater target recognition problem, are provided in Sec. III-D.

In earlier human-computer cooperation attempts, the outputs of detection algorithms were used to cue human operators to alarms, with this effected by overlaying boxes on the imagery on the operator's display screen [5]. However, this approach has fundamental flaws because it constructs an artificial saliency map [6] that overrides that of the remote-sensing imagery itself. By flagging specific alarms, one directs the attention of the operator to those locations, and as a result, the operator becomes *more likely* to miss targets that are *not* cued [7], [8]. Therefore, if the detection algorithm cannot be ensured to achieve a perfect probability of detection, such cuing will degrade performance.

Another issue with this intrusive cuing approach is that the burden on the human operator actually *increases* when the false alarm rate is high. Rather than saving the operator time, the well-intentioned automation in fact makes the task more difficult. It is no surprise then that the first action many operators took when given systems with automated cuing in field tests was to shut off the cuing program completely [4], [8], [9]. Alternative approaches to using automation in the detection stage are given in Sec. III-B.

III. Strategies for Human-Computer Cooperation in Target Recognition

We propose various ways of leveraging human operators and automated algorithms to improve object-recognition performance – importantly, without significantly increasing operator burden – depending on mission requirements and resources (*e.g.,* time or manpower available). The strategies are described in the context of a task relying on sonar imagery of a large area of seabed over which underwater targets (*e.g.,* mines) must be detected and classified. The following recommendations proceed in terms of increasing levels of operator effort required.

A. Human as Aid to Classifier

A human operator can aid automated algorithms via active learning [10], in which the operator provides (potentially noisy [11]) labels for a subset of the most informative alarms flagged by a detection algorithm. This human feedback injects expertise that better tailors the classifier to the site under consideration. The number of alarms to query the operator for will necessarily depend on the amount of operator effort available.

B. Detector as Aid to Human

An automatic detection algorithm can be used to provide a rapid assessment of the feasibility of performing minehunting in the area. If the overall alarm rate of the detector is excessively high, this can be interpreted as a sign that the area is unhuntable (in the time allotted) and that it is wiser to seek an alternate route (mission requirements permitting) through a different area. Using the automated detector to provide a fast meta-analysis of the mission area in this manner enables potential time savings (*e.g.,* preventing the scenario in which an operator searches an area only to come to the conclusion after much effort that the area is unhuntable) and can also provide the operator with rough estimates regarding the time that will be needed to complete the task.

If a human operator is instead supplied – on his first inspection of the data – with the actual alarms of the detection algorithm, his own decisions might get biased. Moreover, such an arrangement could set the precedent in which the operator eventually relies too heavily on the algorithm, after which the operator's own skills deteriorate [8], [12].

An operator should be presented with the alarms flagged by the automated detection algorithm – time-permitting – only after he has inspected the data himself. This presentation, however, should be in the form of both a "mugshot" of each alarm *as well as* the larger image scene from which it derives, because contextual information is valuable ("the perceptual saliency of stimuli critically depends on surrounding context" [6]). For example, a mugshot of a rock may appear very mine-like when it is viewed in isolation, but the same rock may be easily rejected as a false alarm if it is seen as part of a large boulder field.

C. Classifier as Aid to Human

If multiple human operators are available, it is useful for each to make detection and classification decisions in isolation, as this engenders view diversity among the operators. (This is analogous to seeking a second opinion on a medical diagnosis.) Subsequently, the individual predictions can be fused, with ensemble methods [13] providing principled justification for such an approach. If the operators instead work in a single team, the consensus that is arrived at in the decision-making process will tend to eliminate any view diversity and potential for improvement via fusion. For the same reasons, the inclusion of predictions from an automated classification algorithm – treated, in essence, as an additional operator – is also advised as another form of human-computer cooperation that provides potential performance benefit without increasing the burden on the human operator.

D. Directions for ATR Research

We argue that the directions for future ATR research should differ depending on whether or not human-computer cooperation is intended. The fusion of human operators and ATR algorithms can exceed the performance of humans alone if the latter are developed appropriately. Specifically, complementary information and viewpoints should be sought from humans and computer algorithms.

The ATR algorithms to be exploited in human-computer efforts should attempt to capture elements that a human has difficulty perceiving. A human operator performing underwater object detection will naturally be attracted to salient parts of a scene, like bright highlights and crisp shadows that a target produces under ideal conditions. The potential benefit of an automated detection algorithm lies instead in cases where a human may struggle. Examples of these include an object lacking a shadow due to poor image quality or multipath effects, an object in sand ripple fields where highlights of the object and ripples blend together, and an object at short range

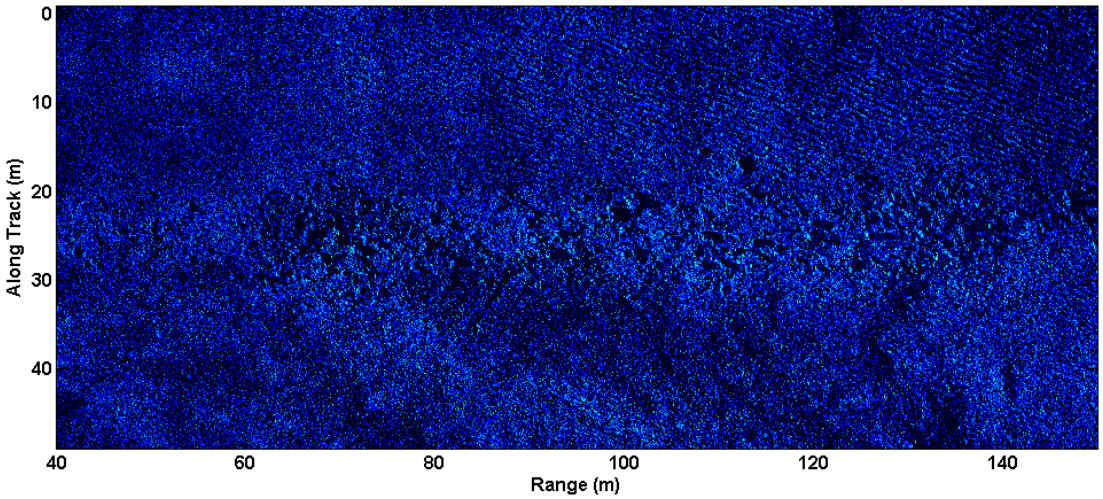

Fig. 1. One example SAS image from the experiment.

(from the sensor) where the shadow is shorter (and hence less noticeable) due to geometry. A detection algorithm that can detect targets in these cases – in the "blind spots" of humans where failure is likely – is truly worthwhile and cooperative.

The development of features for use in an automated classification algorithm should similarly not try to simply mimic clues on which a human focuses. For example, from interferometric synthetic aperture sonar systems, phase information can provide object height estimates that are not perceivable to a human operator examining standard sonar imagery. By developing these sorts of features beyond the limits of human perception, the view diversity between human and computer predictions will be strengthened. In turn, potential performance gains will be possible via fusion because of the complementary information available for exploitation.

Lastly, a human asked to describe an underwater environment would likely think in terms of sediment type (*e.g.,* sand, mud, silt) rather than on more subtle characteristics that directly impact feature calculations. Therefore, an automated classification algorithm with the ability to use *in situ* through-the-sensor data to characterize the local environment and adapt accordingly can provide benefit where a human is perceptually limited. In particular, such a classifier might be able to discern which historical data from other locations is most relevant for the test site.

IV. At-Sea Experiment

A. Background

In April 2013, CMRE participated from the R/V Alliance in a minehunting exercise conducted at sea by the Spanish navy off the coast of Cartagena, Spain. The objective of one task was to detect and classify mines contained in a designated one square nautical mile area of seabed, within a specific time allotment. To perform this task, CMRE first deployed a SAS-equipped autonomous underwater vehicle (AUV) called MUSCLE to collect sonar imagery of the mission area. The center frequency of the SAS is 300 kHz and the bandwidth is 60 kHz.

After the entire area was surveyed, the AUV was recovered and the raw sonar data were downloaded and processed into SAS imagery with an across-track resolution of 1.5 cm and an along-track resolution of 2.5 cm; each image spanned 110 m in the across-track direction and 50 m in the along-track direction. As each image was created, it was assigned a unique image number and presented to multiple human operators who were tasked with detecting mines present in the image. Both the number and type of mines contained in the mission area were unknown to the operators. In total, 635 SAS images were presented to the operators; one example SAS image from the experiment is shown in Fig. 1.

Conducting a real, time-limited detection and classification experiment with human operators on actual data under realistic conditions at sea is very rare and difficult to achieve. For example, the laboratory simulation in [14] was not time-constrained, informed participants of the target types in advance, and compared only *classification* performance between humans and an automated algorithm, skipping the important detection stage of the task. To our knowledge, our study is the first in the open literature to assess the performance of the full suite of human-computer combinations, from detection through classification.

B. Human Operators

The experiment was conducted with four operators.[1] The operators were scientists with varying levels of experience dealing with mines in sonar imagery. The operators worked in isolation and did not discuss the data with each other during the experiment. Each operator examined the images and recorded the location of each object he felt was a target, and assigned a (subjective) confidence value from 1 to 5 (5 being most confident of its identity as a target) to the alarm. On average, the operators had approximately 30 seconds to examine each image.

For the subsequent performance analysis, each operator confidence score – essentially the operator's classification

[1] One "operator" was actually a team of two humans, so in total five humans participated.

prediction for an alarm – was scaled (by dividing by 5) to represent the probability of being a mine, as this comports with standard scoring rubrics used by the defense community [3].

For cases presented later simulating active learning, it was assumed that the operator labeling of one detection-algorithm-flagged alarm took 10 seconds. This (binary) labeling was effected by translating into a label of "target" only those alarms for which the operator confidence was 4 or 5 (otherwise a label of "non-target" was assigned).

C. Automated Computer Algorithms

For the cases employing automated algorithms, we employed specially tailored detection and classification algorithms that we have recently developed. The detection algorithm [15] is an unsupervised method that handles various elements outlined in Sec. III-D – including poor image quality, objects in sand ripples, and shadow-length range-dependence – that tend to be challenging for humans.

The classification algorithm [16] learns an ensemble of classifiers in which the local environmental characteristics, measured through-the-sensor, of each alarm determine the degree to which each base classifier is trusted. The base classifiers in this framework are relevance vector machines [17] and were trained using historical data from eight geographical sites with diverse environmental conditions. Importantly, the level of similarity between the test site and the historical data need not be known *a priori*. Space constraints prevent a more thorough description of the automated algorithms here.

D. Assessment

The locations and identities of the targets present were disclosed after the experiment, from which it was determined that the images contained 16 target views. The number of alarms flagged by Operators 1 through 4 were 80, 104, 74, and 51, respectively, with a collective total of 183 unique alarms. The distribution of the confidence values that each operator assigned to alarms is shown in Fig. 2. Operators 1, 2, and 3 each failed to detect 1 target, while Operator 4 missed 3 targets. (These results illustrate the variability that can be obtained with human operators.) The automated detection algorithm flagged 282 alarms and failed to detect 3 targets. Pooling all alarms from among the operators and the automated detection algorithm, there were 353 unique alarms, including all 16 targets.

The area under an ROC curve (AUC) [18] provides a convenient summary measure of performance (with scalar values in $[0, 1]$) where higher values indicate better performance. To facilitate the direct comparison of the performance using various human-computer combinations for detection and classification considered here (with different sets of alarms), all AUC values were computed based on the full universe of 353 unique alarms. We omit the full ROC curves (from which the AUC values were computed) of the various cases due to space constraints.

E. Results of Human-Computer Combinations

Fig. 3 shows the AUC values for various human-computer combinations of detection and classification as a function of

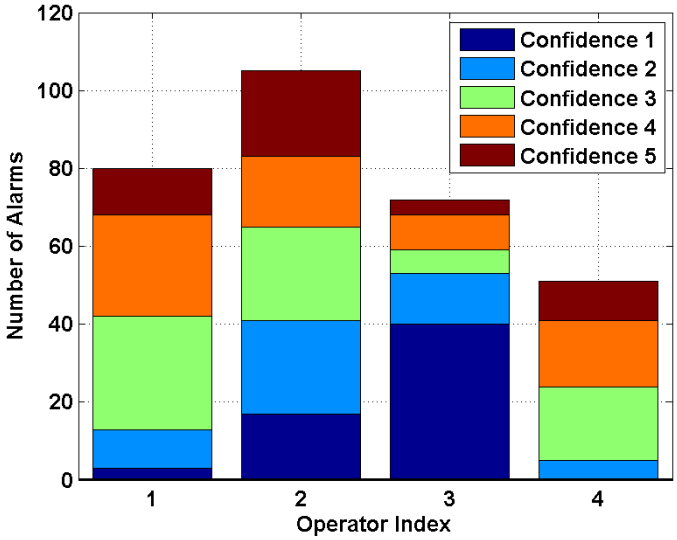

Fig. 2. Distribution of the confidence values that each operator assigned to alarms.

human operator effort, measured in time (on a logarithmic scale). As one moves to the right on the x-axis, automation (and trust in it) decreases. The lettered cases below reference the legend entries.

Case A corresponds to the scenario in which both detection and classification are performed by the automated algorithms, with no human operator involvement. Cases B – E correspond to the scenario in which the automated detection algorithm generates the list of alarms, an operator then provides labels (target or non-target) for 10 alarms, and then the automated classification algorithm learns a classifier using both the historical data and the newly labeled data jointly, and then makes final predictions on all alarms. For these cases, the operator was queried for the 10 alarms with the highest detection score, with the operator confidence scores translated into binary labels as described earlier. This active learning is a noisy-labeling process because the operators do not know the true labels with certainty; for the experiments here, only Operator 2 was equivalent to an "oracle" assigning perfect labels. Nevertheless, by incorporating minimal feedback from an operator in this active learning process, performance improved.

In cases F – I, the automated detection algorithm generates the list of alarms, but each operator assigns predictions (*i.e.,* the operator confidence scores) to all alarms. Although operator effort increases for these cases, performance does not necessarily improve. Case J is similar except the final predictions are obtained as the mean confidence score from all four operators. (Detector alarms that an operator did not himself flag were assigned a confidence score of 0, which is consistent with the scoring convention.) Because this case involves the effort of four operators, the total human effort is more; however, performance also improves beyond that of any single operator, as the fusion procedure effectively leverages the viewpoint diversity of the operators.

Cases K – N correspond to the scenarios in which one operator performs both detection and classification with no computer aid. These cases require substantial operator effort,

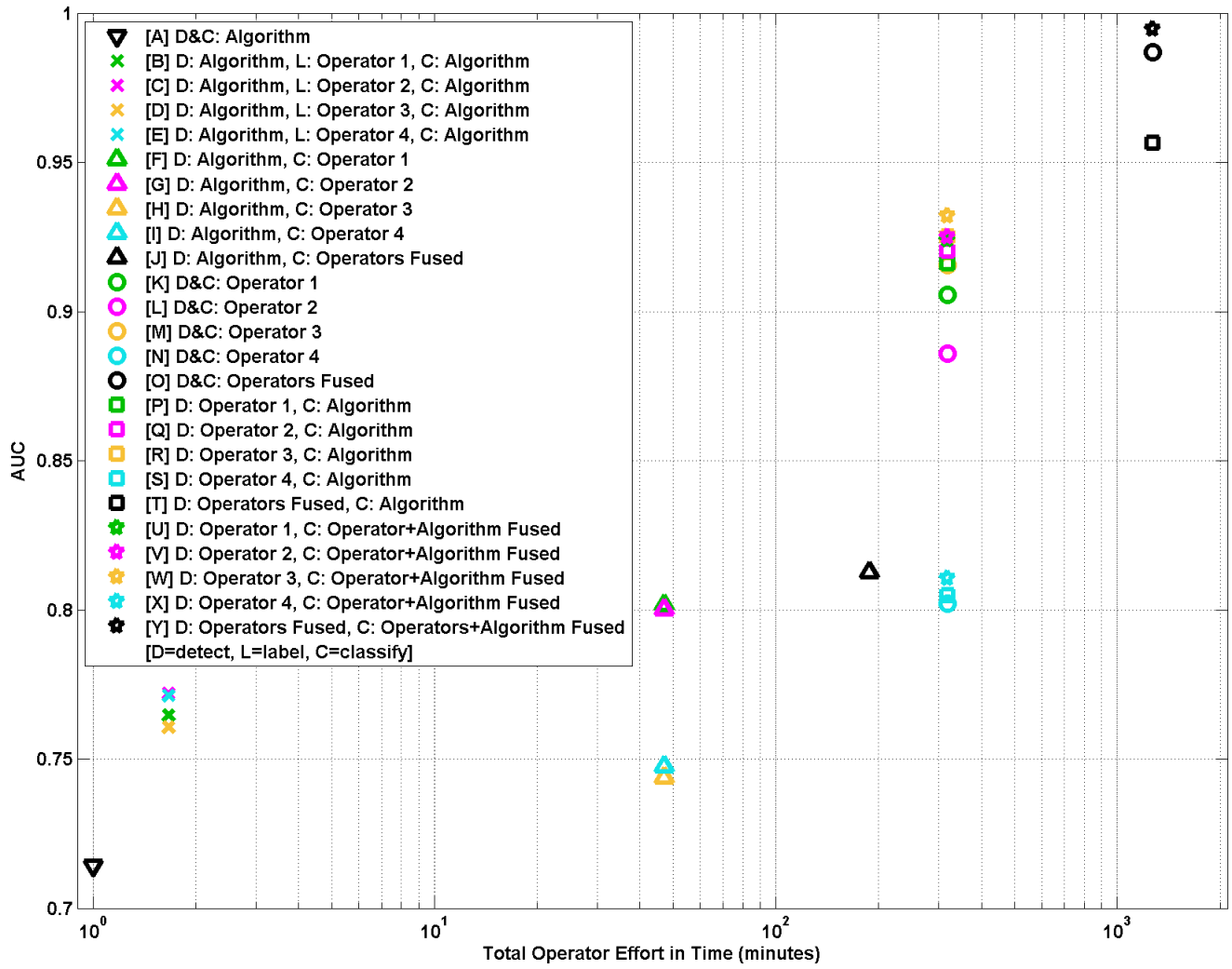

Fig. 3. Classification performance as a function of human operator effort for various human-computer combinations of detection and classification.

but the performance is typically significantly improved as well. However, performance can improve still further without *any* additional human effort by employing the automated classification algorithm. This is demonstrated in Cases P – S, where the alarms are generated by a human operator but the final classification is performed by the automated classifier. Still further improvement can be achieved, again with no additional operator effort, by fusing the predictions of the operator and the classifier. This is demonstrated in Cases U–X where the alarms are again generated by a human operator but final predictions are taken to be the mean of the operator's (scaled) scores and the classifier's predictions. The performance improvement in these cases can be attributed to the viewpoint diversity that the human operators and automated algorithms provide. These results illustrate the immediate performance gains – without additional operator burden – that can be achieved by incorporating human-computer cooperation.

The value of fusing the efforts of multiple human operators can be seen in Case O, which pools the alarms generated by the four human operators and makes predictions based on the mean confidence score. Case T corresponds to the same scenario except that the final predictions are based on the automated classification algorithm rather than the operators' scores. Case Y corresponds to a similar scenario, except the final predictions are taken to be the mean of the operators' (scaled) scores and the classifier's predictions. These three cases require the most operator effort because all four operators are used. However, the considerable viewpoint diversity elicited from multiple human operators and the automated algorithm result in the best performance. In these cases, the inclusion of the automated classifier – which can be viewed as another operator – is even more valuable than an additional *human* operator because the automated classifier has the potential to provide complementary information that is impossible for a human to perceive.

Although the differences in AUC values are small among certain cases, it should be realized that the reduction of even a single false alarm is valuable. Typically, each alarm declared a target would necessitate the time-consuming and dangerous task of optical inspection – either with human divers or camera-equipped remotely-operated vehicles. Because this process can take upwards of a half hour per object, even slight improvements in the AUC translate to significant time-savings in real operations.

V. Conclusion

Specific strategies for cooperatively employing human operators and automated computer algorithms have been provided for underwater target recognition applications. It has been experimentally demonstrated that fusing the skills of a human (or multiple humans) and computers can significantly improve performance beyond that which is achievable with only one type of operator thanks to the view diversity that is engendered. This progress lays the foundation for the acceptance of still-further automated systems in the future.

Acknowledgment

The authors would like to thank John Fawcett (DRDC, Canada) and Romain Huiban (DGA, France) for their participation as the operator team.

References

[1] J. Stack, "Automation for underwater mine recognition: Current trends and future strategy," in *SPIE Defense, Security, and Sensing*, vol. 8017, 2011.

[2] D. Kobus and L. Lewandowski, "Critical factors in sonar operation: A survey of experienced operators," Naval Health Research Center, Tech. Rep. NHRC-91-19, 1991.

[3] G. Kuperman, "Human system interface (HSI) issues in assisted target recognition (ASTR)," in *Proc. IEEE 1997 National Aerospace and Electronics Conference*, vol. 1, 1997, pp. 37–48.

[4] N. Allen and R. Kessel, "The roles of human operator and machine in decision aid strategies for target detection," in *RTO HFM Symposium on the Role of Humans in Intelligent and Automated Systems*, October 2002.

[5] R. Kessel, "On-screen alarms in computer-aided detection systems: Combining signal processing, human factors, and system design," Canadian National Defence Research and Development Branch, Tech. Rep. DREA TM 2001-184, 2001.

[6] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[7] J. See, I. Davis, and G. Kuperman, "Aided and unaided operator performance with synthetic aperture radar imagery," in *Proc. IEEE 1998 National Aerospace and Electronics Conference*, 1998, pp. 420–427.

[8] R. Parasuraman, T. Sheridan, and C. Wickens, "A model for types and levels of human interaction with automation," *IEEE Trans. Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 30, no. 3, pp. 286–297, 2000.

[9] J. Irvine, "User issues associated with sensor fusion and ATR," in *Proc. International Conference on Information Fusion*, 2003.

[10] B. Settles, "Active learning literature survey," University of Wisconsin–Madison, Computer Sciences Technical Report 1648, 2009.

[11] V. Sheng, F. Provost, and P. Ipeirotis, "Get another label? Improving data quality and data mining using multiple, noisy labelers," in *Proc. 14th ACM SIGKDD International Conference On Knowledge Discovery And Data Mining*, 2008, pp. 614–622.

[12] K. Cosenzo and M. Barnes, "Human-robot interaction research for current and future military applications: From the laboratory to the field," in *SPIE Defense, Security, and Sensing*, vol. 769204, 2010.

[13] T. Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems*. Springer, 2000, pp. 1–15.

[14] R. Kessel and V. Myers, "Discriminating man-made and natural objects in sidescan sonar imagery: Human versus computer recognition performance," in *SPIE Defense and Security*, 2005.

[15] D. Williams, "On adaptive underwater object detection," in *Proc. International Conference on Intelligent Robots and Systems*, 2011, pp. 4741–4748.

[16] D. Williams and E. Fakiris, "A new environmentally adaptive classification algorithm for underwater mines in SAS imagery," in *Proc. International Conference and Exhibition Underwater Acoustics*, 2013, pp. 703–711.

[17] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[18] J. Hanley and B. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, pp. 29–36, 1982.