Confidence Level Estimation in Multi-target Classification Problems

Shi Chang[†], Jason Isaacs[§], Bo Fu[†], Jaejeong Shin[†], Pingping Zhu[†], and Silvia Ferrari[†]

[†]Sibley School of Mechanical and Aerospace Engineering, Cornell University
[§]Naval Surface Warfare Center (NSWC) - Panama City Division

ABSTRACT

This paper presents an approach for estimating the confidence level in automatic multi-target classification performed by an imaging sensor on an unmanned vehicle. An automatic target recognition algorithm comprised of a deep convolutional neural network in series with a support vector machine classifier detects and classifies targets based on the image matrix. The joint posterior probability mass function of target class, features, and classification estimates is learned from labeled data, and recursively updated as additional images become available. Based on the learned joint probability mass function, the approach presented in this paper predicts the expected confidence level of future target classifications, prior to obtaining new images. The proposed approach is tested with a set of simulated sonar image data. The numerical results show that the estimated confidence level provides a close approximation to the actual confidence level value determined *a posteriori*, i.e. after the new image is obtained by the on-board sensor. Therefore, the expected confidence level function presented in this paper can be used to adaptively plan the path of the unmanned vehicle so as to optimize the expected confidence levels and ensure that all targets are classified with satisfactory confidence after the path is executed.

Keywords: Multi-target, classification, confidence level, estimation, information gain, path planning.

1. INTRODUCTION

Multi-target classification by imaging sensors, such as cameras, radar, or sonar, has a wide range of applications, including search and rescue, urban surveillance, and environmental monitoring.[1] In this paper, a simulated side-scan sonar installed on an unmanned underwater vehicle (UUV) obtains images of the sea floor that are then used to automatically detect and classify objects using a deep convolutional neural network (CNN) approach.[2] In this approach, the CNN convolutional features extracted from the sonar image matrix are provided as an input to a support vector machine (SVM) classifier that estimates the target class and an associated posterior probability. When target classification needs to be highly reliable, the problem is often divided into two stages: a survey stage during which objects are first detected by taking one or few images, and an identification stage during which additional images are obtained until the confidence in the classification decision is above a desired threshold.

When the target class is a hidden variable, the classification accuracy is always unknown, and the confidence level (CL) is unknown until the images are obtained by the vehicle. Thus, a challenging problem is how to plan the path of the vehicles efficiently, while also achieving a satisfactory classification performance. In conventional sensor measurements, expected information gain functions can be used to estimate the information value using, for example, expected entropy or Rényi divergence.[3–7] These information gain functions are challenging to derive for imaging sensors or for deep convolutional features. Furthermore, if the targets must all be classified within a desired confidence level, the information gain functions may not be applicable to finding a path that guarantees to cover all the targets subject to this requirement. This paper presents an approach for estimating the expected confidence level of a target *a priori*, assuming the classification will be performed based on the maximum *a posteriori* (MAP) decision rule. Since multiple images might be required to meet the desired CL threshold, an approach for recursive CL estimation is presented that takes into account the probabilities learned from data and all prior images. Numerical simulations show the effectiveness of the proposed approach on a simulated UUV-based imaging sonar.

Further author information: (Send correspondence to Silvia Ferrari)
Shi Chang: E-mail: sc2892@cornell.com
Silvia Ferrari: E-mail: ferrari@cornell.edu

2. PROBLEM FORMULATION

Consider the problem of classifying n targets in an underwater region-of-interest, $\mathscr{W} \subset \mathbb{R}^2$, were all targets are at fixed and unknown locations on the seafloor. While the onboard sonar is in operation, the UUV travels at constant speed and heading, in order to generate high-quality images of the seafloor. The sonar measurements are processed into an image matrix K, referred to as sonar image, that is automatically segmented into smaller images, denoted by I, using a matched filter.[2] The UUV goal is to classify targets from multiple sonar images and to adaptively compute the best configuration for obtaining additional sonar images. All target features are discretized and treated as discrete random variables. Upper case letters represent random variables and lower case letters represent the value (or realization) of the corresponding random variable. Each target has a set of physical characteristics, or features, chosen here as shape S, length L, width W, and height H. Let X denotes the set of actual target features, and $\hat{X}$ denote the set of features estimated from sonar images with range $\hat{\mathscr{X}}$. The aspect angle Θ and relative distance D between a target and the UUV are also discretized, and denoted by the discrete random variables D and Θ, respectively, such that $R = \{D, \Theta\}$ is the relative configuration between the UUV and each target. The problem considered in this paper is to classify each target into one of two classes (0 or 1), such that $Y \in \mathscr{Y} = \{0, 1\}$ is a binary classification variable that is hidden and random. Y is to be estimated from a set of sequential segmented sonar images $I^k = \{I_1, I_2, ..., I_k\}$ taken at configurations $r^k = \{r_1, r_2, ..., r_k\}$, where k is a time index that is not equally spaced and denotes time instants at which a sonar image is available.

At any time k, the UUV planning algorithm plans the next configuration r_{k+1} such that an additional image segment I_{k+1}, taken at r_{k+1}, will reduce uncertainty enough to bring the CL above a desired threshold. The later task is achieved by learning a function $f : \mathscr{R} \mapsto [0, 1]$, where $\mathscr{R}$ is the set of all the possible configuration r. $\mathscr{I}$ is the set of all the possible image segment I. Let $u \in \mathbb{R}$ be the utility of the next observation,

$$u = f(r_{k+1}; r^k, I^k), \tag{1}$$

where r^k and I^k are parameters. After obtaining the k^{th} measurement, r^k and I^k are known. Thus, u is only a function of next configuration r_{k+1} and can be maximized as follows:

$$\hat{r} = \underset{r_{k+1} \in \mathscr{R}}{\operatorname{argmax}} f(r_{k+1}; r^k, I^k). \tag{2}$$

The image processing and feature extraction approach previously developed by the authors is adopted and briefly reviewed here for completion.[2] The original sonar image is first down-sampled by a chosen factor to remove redundant information and reduce the computation required. The sonar image matrix is expressed on a grey scale by linear projection onto the interval $[0, 1]$,

$$K_g(i, j) = \frac{K(i, j)}{\max_{i,j}[K(i, j)]}, \tag{3}$$

and the histogram equalization technique[8] is applied to obtain a high contrast gray-scale image matrix, as shown by the example in Fig. 1. The pre-processed sonar image is segmented by a matched filter to obtain a sub-image, I, for each target detected during surveying.

The well-known pre-trained CNN ALexNet, overfitted to a sonar training database, is then used to extract the convolutional features from the segmented sonar image I. Using a pre-trained CNN in series with an SVM classifier is an effective approach when the training data set is too small to train a deep CNN from scratch, and also helps reduce the computation required.[2] The activation from the first fully connect layer is selected to form a (4096×1) convolutional feature vector, denoted by $\mathbf{z}$. Then the convolutional features are fed to a linear SVM trained for target classification on a labeled training set.[9] The SVM maps the convolutional features to the target classification label $\hat{y} \in \{0, 1\}$. This approach has shown to provide excellent classification accuracy and, therefore, is adopted in this paper to perform automatic target recognition on the simulated UUV.[2] It is then assumed, for simplicity, that the UUV has knowledge of the relative target position and aspect angle based on on-board instrumentation and prior sonar images.

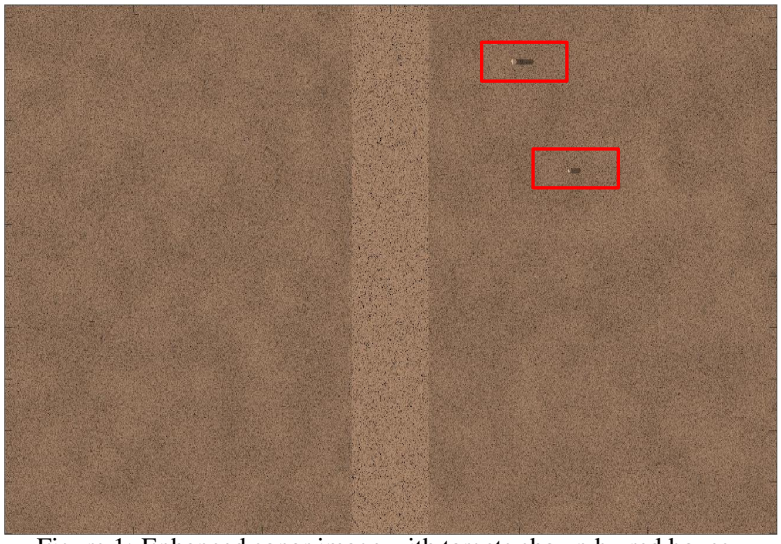


Figure 1: Enhanced sonar image with targets shown by red boxes.

3. RECURSIVE ESTIMATION AND PROBABILITY LEARNING

This section presents a recursive estimation approach for updating the belief on the target classification Y, based on the inferred target class and probability distribution obtained from previous images. At time k, a deep learning classification label $\hat{y}_k$ and the target feature estimates x_k are obtained for each target detected during surveying, using the segmented sonar image I_k taken at the UUV configuration r_k. Thus, the evidence set $M_k = \{e_1, e_2, ..., e_k\}$, where $e_k \triangleq \{\hat{y}_k, \hat{x}_k, r_k\}$, is used to develop a non-myopic strategy for information gathering. The conditional posterior distribution of the target class Y is obtained using Bayes' rule,

$$P(Y|M_k) = P(Y|e_k, M_{k-1}) = \frac{P(e_k|Y, M_{k-1})P(Y|M_{k-1})}{\sum_Y P(e_k|Y, M_{k-1})P(Y|M_{k-1})}. \tag{4}$$

In most sensor measurements, it can be assumed that e_k is conditionally independent on M_{k-1} given the target class, or

$$P(e_k|Y, M_{k-1}) = P(e_k|Y). \tag{5}$$

Thus, the posterior probability can be rewritten as,

$$P(Y|M_k) = \frac{P(e_k|Y)P(Y|M_{k-1})}{\sum_Y P(e_k|Y)P(Y|M_{k-1})}, \tag{6}$$

where $P(Y|M_{k-1})$ is known from time $(k-1)$. Using the chain rule of probability, the probability of evidence can be factored as,

$$P(e_k|Y) = P(\hat{y}_k, \hat{x}_k, r_k|Y) = P(\hat{y}_k|\hat{x}_k, r_k, Y)P(\hat{x}_k|r_k, Y)P(r_k|Y), \tag{7}$$

where $P(\hat{y}_k|\hat{x}_k, r_k, Y)$ and $P(\hat{x}_k|r_k, Y)$ conditional probability tables known from the sensor measurement model learned from the training set, and $P(r_k|Y)$ is assumed to be a uniform prior. Then, the the posterior distribution of the classification variable, Y, in (6), can be recursively updated as follows,

$$P(Y|e_1) = \frac{P(e_1|Y)P(Y)}{\sum_Y P(e_1|Y)P(Y)}, \tag{8}$$

where the prior $P(Y)$ is learned from the training set. By the maximum *a posteriori* (MAP) rule, the target class is decided based on the maximum posterior probability:

$$\hat{y} = \underset{y \in \mathscr{Y}}{\operatorname{argmax}}[P(Y|M_k)] = \underset{y \in \mathscr{Y}}{\operatorname{argmax}} \left[\frac{P(e_k|Y)P(Y|M_{k-1})}{\sum_Y P(e_k|Y)P(Y|M_{k-1})} \right]. \tag{9}$$

The actual classification confidence level (CL) associated with the decision (9) at time k is,

$$CL(e_k|M_{k-1}) = \max_{y\in\mathscr{Y}}[P(Y|M_k)] = \max_{y\in\mathscr{Y}}\left[\frac{P(e_k|Y)P(Y|M_{k-1})}{\sum_Y P(e_k|Y)P(Y|M_{k-1})}\right], \tag{10}$$

where a higher CL value suggests lower uncertainty in the chosen class.

4. EXPECTED CONFIDENCE LEVEL

In order to adaptively plan the next UUV configuration so as to meet the desired CL for all targets in the region of interest, $\mathscr{W}$, this paper presents an approach for predicting the confidence level of the next image, or $CL(e_{k+1}|M_k)$, where e_{k+1} is unknown. The expected confidence level is obtained by taking one step conditional expectation of the confidence level with respect the to next estimated target class variable, $\hat{Y}_{k+1}$, and target feature set, $\hat{X}_{k+1}$. At time k, the expected confidence level associated with a future configuration value r_{k+1} is

$$ECL(r_{k+1}|M_k) = \mathbb{E}_{\hat{Y}_{k+1},\hat{X}_{k+1}}[CL(\hat{Y}_{k+1},\hat{X}_{k+1},r_{k+1}|M_k)]. \tag{11}$$

From the definition of conditional expectation, the above equation can be rewritten as,

$$ECL(r_{k+1}|M_k) = \sum_{\hat{y}_{k+1}\in\mathscr{Y},\hat{x}_{k+1}\in\mathscr{X}} CL(\hat{y}_{k+1},\hat{x}_{k+1},r_{k+1}|M_k)P(\hat{y}_{k+1},\hat{x}_{k+1}|M_k,r_{k+1}), \tag{12}$$

where $CL(\hat{y}_{k+1},\hat{x}_{k+1},r_{k+1}|M_k)$ can be obtained recursively from Eq. (10) using each pair of possible realizations of $\hat{Y}_{k+1}$ and $\hat{X}_{k+1}$. The probability $P(\hat{Y}_{k+1},\hat{X}_{k+1}|M_k,r_{k+1})$ in Eq.(12) can be computed by marginalizing the joint probability $P(Y,\hat{Y}_{k+1},\hat{X}_{k+1}|M_k,r_{k+1})$ over the unknown target classification Y

$$\begin{aligned} P(\hat{Y}_{k+1},\hat{X}_{k+1}|M_k,r_{k+1}) &= \sum_Y P(Y,\hat{Y}_{k+1},\hat{X}_{k+1}|M_k,r_{k+1}) \\ &= \sum_Y P(\hat{Y}_{k+1},\hat{X}_{k+1}|Y,M_k,r_{k+1})P(Y|M_k,r_{k+1}). \end{aligned} \tag{13}$$

Under the sensor conditional independence assumptions, $P(\hat{Y}_{k+1},\hat{X}_{k+1}|Y,M_k,r_{k+1}) = P(\hat{Y}_{k+1},\hat{X}_{k+1}|Y,r_{k+1})$, and, since the target class is independent of the sensor range, $P(Y|M_k,r_{k+1}) = P(Y|M_k)$, and Eq. (13) can be written as,

$$\begin{aligned} P(\hat{Y}_{k+1},\hat{X}_{k+1}|M_k,r_{k+1}) &= \sum_Y P(\hat{Y}_{k+1},\hat{X}_{k+1}|Y,r_{k+1})P(Y|M_k) \\ &= \sum_Y P(\hat{Y}_{k+1}|\hat{X}_{k+1},Y,r_{k+1})P(\hat{X}_{k+1}|Y,r_{k+1})P(Y|M_k), \end{aligned} \tag{14}$$

where $P(\hat{Y}_{k+1}|\hat{X}_{k+1},Y,r_{k+1})$ and $P(\hat{X}_{k+1}|Y,r_{k+1})$ can be obtained from sensor measurement model (Section 3). Thus, the ECL can be calculated recursively at every time step from Eqs. (11), (12) and (14), and a path can be planned such that a desired ECL is obtained for all n targets in $\mathscr{W}$.

5. SIMULATION RESULTS

The proposed approach for the recursive computation of the expected confidence level is tested on a simulation of $n = 520$ targets distributed in a region of interest $\mathscr{W} = [-L,\ L]\times[-L,\ L]$, where $L = 1200$ meters. Training and testing data sets are obtained for training the CNN and SVM classifier,[2] as well a Bayesian network measurement model of the UUV-based sonar.[10] The testing data set has $n = 215$ targets, each with a hidden target class $Y \in \{0,\ 1\}$ and a diverse distribution of target features. During the survey stage, a first set of sonar images are obtained by the UUV using a lawnmower path in the North-South designed to cover $\mathscr{W}$.

From the survey images, the target features and classification are estimated by applying the CNN and SVM classifier to the segmented sonar images. Subsequently, the ECL is estimated as a function of the UUV configuration, r_{k+1}, and estimated target class and features, as shown in Sections 3-4. By this approach, the ECL is estimated before obtaining additional sonar images, and the best UUV path can be planned accordingly. When the identification stage begins, and the UUV obtains new sonar images, they are similarly processed to obtain new class and feature estimates, as well as an actual

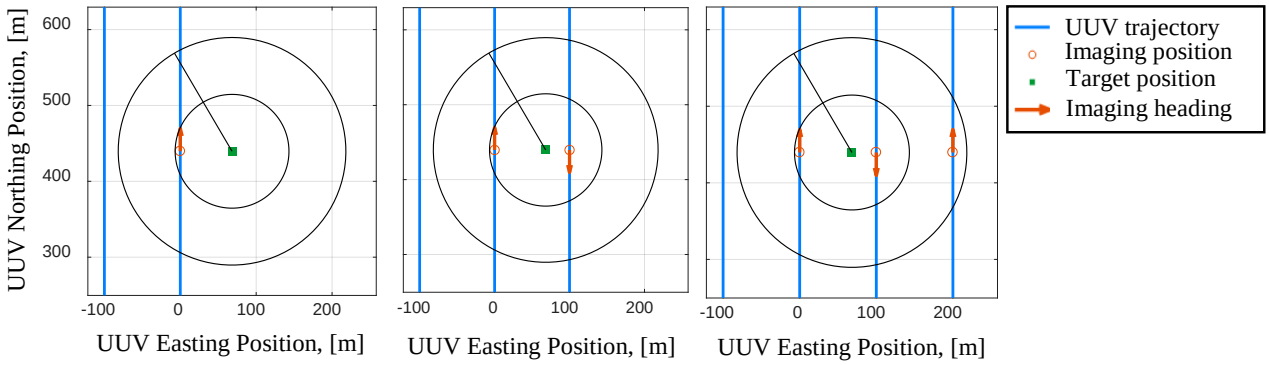


Figure 2: UUV trajectory and UUV (imaging) position and heading at the times sonar images are obtained from the target.

CL for the classification decision, according to (10). Then, if the actual CL is not above the desired threshold, the ECL is also updated recursively according to (14), in order to plan additional measurements online. The error in the CL estimate is given by,

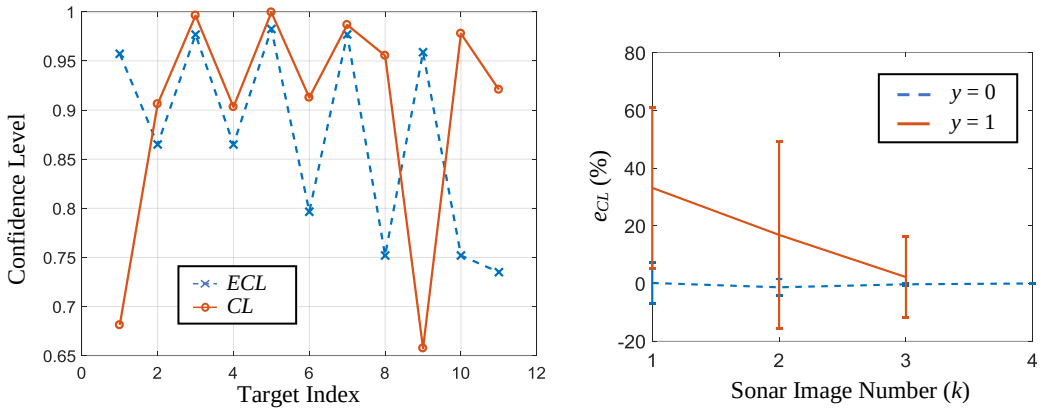
$$e_{CL} = \frac{|ECL - CL|}{CL} \cdot (100\%) \tag{15}$$

and is computed here for validation, after every new image is obtained. Figure 2 and Table 1 show the estimated and actual classification of a target observed three consecutive times by the UUV-based sonar. It can be seen that the the accuracy of both estimates improves as additional images are obtained at different relative range and aspect angles.

Table 1: Confidence levels and classification results obtained after each of three sonar images is obtained, sequentially, from the same target (index 3), as shown in Fig. 2.

Sonar Image Number (k)	1	2	3
Expected CL ($ECL(r_k \mid M_{k-1})$)	0.9580	0.7514	0.9768
Estimated target classification ($\hat{y}_k$)	0	1	1
Actual CL ($P(\hat{y}_k \mid M_k)$)	0.5312	0.9739	0.9962
Actual target classification (y)	1	1	1

Figure 3(a) and 4 show a comparison of actual and estimated CL for all of the targets in the testing data set, organized by actual classification, $y = 1$ and $y = 0$, respectively. Although the ECL is inaccurate when the prior probability and corresponding MAP estimate are incorrect, it can be seen that in many cases, the ECL follows the same trend as the actual CL. From (15), the maximum CL error is 45.82% when $y = 1$, and 9.90% when $y = 0$. The majority of the target with $y = 0$ display a CL error smaller than 4%. Therefore, the expected confidence level is a good estimation of the actual confidence level. Half of the targets with $y = 1$ have a CL error less than 5%. For those few targets that display a high CL error (Table 1), due to the poor quality of the survey image(s), the ECL accuracy rapidly improves as more images are obtained during the identification stage. This is especially true when the UUV configuration is planned so as to maximize the ECL, as will be shown in a separate paper. The mean and standard deviation of the CL error, obtained from the entire data base, are plotted in Fig. 3(b) as a function of number of images per target. The results in Fig. 3(b) show that, when $y = 1$ more images are required to accurately estimate the CL then when $y = 0$, but in both cases the ECL improves over time. It was also found that the ECL accuracy improves with the quality of the sensor measurement model, which is typically learned from data and/or first principles.[10] As a result, the effectiveness of this approach highly depends on the training and testing data sets. Since the availability of labeled data is limited in many imaging applications, future work will also investigate how sampling and re-sampling algorithms can be utilized to improve ECL and classification accuracy.



(a) (b)

Figure 3: Expected and actual confidence level for targets of class $y = 1$(a), and CL error statistics for multiple images (b).

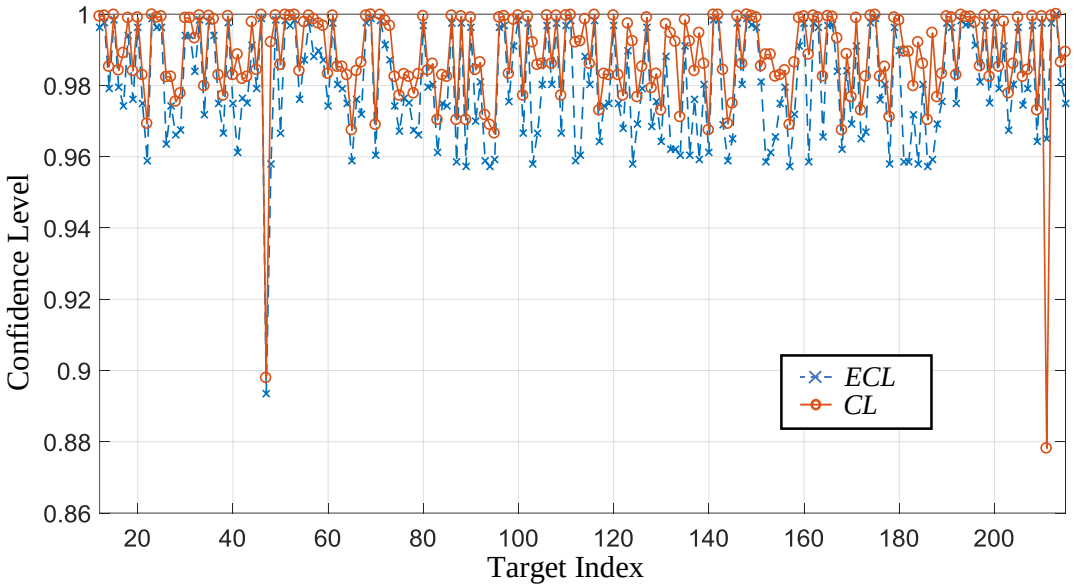


Figure 4: Expected and actual confidence level for targets of class $y = 0$.

6. CONCLUSION

This paper presents an approach for computing the expected confidence level in target classification problems for imaging sensors, such as sonar. A recursive relationship is derived to update the expected confidence level efficiently after every new image of the target is obtained at a new range and orientation. Using a Bayesian measurement model and a deep learning approach for automatic target recognition and feature extraction, the expected confidence level can be obtained from prior information, comprised of processed images, by taking the conditional expectation with respect to target class and feature estimates. Simulation results show that the expected confidence level can be accurately estimated for most targets. When prior images do not allow for accurate estimates, the recursive relationship can be used to improve the estimate efficiently over time, as new target images become available. Therefore, the proposed approach can be used to develop path planning algorithms for sensor applications in which targets must be classified with high confidence, but the image-based target recognition and classification problem is challenging and can only be solved by obtaining multiple images at different positions and orientations.

ACKNOWLEDGMENTS

This work was funded by ONR grant N00014-15-1-2595.

REFERENCES

[1] Zarco-Tejada, P., Diaz-Varela, R., Angileri, V., and Loudjani, P., "Tree height quantification using very high resolution imagery acquired from an unmanned aerial vehicle (uav) and automatic 3d photo-reconstruction methods," *European Journal of Agronomy* **55**, 89 – 99 (2014).

[2] Zhu, P., Isaacs, J., Fu, B., and Ferrari, S., "Deep learning feature extraction for target recognition and classification in underwater sonar images," in [*Decision and Control (CDC), 2017 IEEE 56th Annual Conference on*], 2724–2731, IEEE (2017).

[3] Zhang, G., Ferrari, S., and Cai, C., "A comparison of information functions and search strategies for sensor planning in target classification," *IEEE Transactions on Systems, Man, and Cybernetics - Part B* **42**(1), 2–16 (2012).

[4] Wei, H., Lu, W., Zhu, P., Ferrari, S., Liu, M., Klein, R. H., Omidshafiei, S., and How, J. P., "Information value in nonparametric dirichlet-process gaussian-process (dpgp) mixture models," *Automatica* **74**, 360–368 (2016).

[5] Wei, H., Lu, W., Zhu, P., Huang, G., Leonard, J., and Ferrari, S., "Optimized visibility motion planning for target tracking and localization," in [*Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*], 76–82 (2014).

[6] Cai, C. and Ferrari, S., "On the development of an intelligent computer player for CLUE®: a case study on preposterior decision analysis," in [*American Control Conference*], 4350– 4355 (2006).

[7] Cai, C. and Ferrari, S., "Information-driven sensor path planning by approximate cell decomposition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **39**, 672–689 (June 2009).

[8] Gonzalez, R. and Woods, R., [*Digital Image Processing*], Prentice Hall, Upper Saddle River, New Jersey (2008).

[9] Murphy, K. P., [*Machine learning: a probabilistic perspective*], MIT press (2012).

[10] Cai, C., Ferrari, S., and Ming, Q., "Bayesian network modeling of acoustic sensor measurements," in [*Proc. IEEE Sensors*], 345–348 (2007).