Signal, Image and Video Processing manuscript No.
(will be inserted by the editor)

A Visual-based System for Robotic Inspection of Marine Vessels

Rosalia Maglietta · Annalisa Milella · Massimo Caccia · Gabriele Bruzzone

Received: date / Accepted: date

Abstract This paper presents a novel intelligent system for the automatic visual inspection of vessels consisting of three processing levels: (a) data collection: acquisition of images using a magnetic climber robot equipped with a low-cost monocular camera for hull inspection; (b) feature extraction: all the images were characterized by 12 features made by color moments in each channel of HSV space; (c) classification: a novel tool, based on an ensemble of classifiers, was used to classify sub-images as rust or non-rust. This paper provides a helpful roadmap to guide future research on detection of rusting of metals using image processing.

Keywords Image processing · segmentation · classification

1 INTRODUCTION

Vessels represents one of the most common ways of transport around the world. Their maintenance entails visual inspection of the hull, since the structural failure due to the deterioration of metallic parts of vessels, mostly observable in the form of rust, is the major cause of shipwrecks and catastrophic events for the environment. Regular surveys of vessels are mandatory to make sure that these are maintained according to the technical standards defined by the so-called *classification societies*, non-governmental organizations that establishes technical rules for the construction and maintenance of ships and offshore structures. Nowadays human surveyors visually estimate structural

R. Maglietta, A. Milella and M. Caccia
Institute of Intelligent Systems for Automation, National Research Council
via Amendola 122/D-O, 70126 Bari, Italy
E-mail: {maglietta,milella,caccia}@ba.issia.cnr.it

G. Bruzzone
Institute of Intelligent Systems for Automation, National Research Council
via de Marini 6, 16149 Genova, Italy E-mail: bruzzone@ba.issia.cnr.it

damage, pitting and corrosion of vessels getting within arm's reach of the part under observation using temporary staging, lift and movable platform installed around the vessel. Clearly this procedure carries a high risk for the surveyor and the result depends on observer experience. Moreover, due to the size of vessels the inspection process becomes tedious and long task, and the total cost of a full vessel inspection could become very high depending on the time along which the ship is inoperable. For these reasons, developing novel tools to enhance the ship surveying process is strategic.

The development of a fleet of robots to make ship inspections safer and more cost-efficient was the objective of the MINOAS project (Marine Inspection rObotic Assistant System) [1]. In this context, the Magnetic Autonomous Robotic Crawler (MARC, see Figure 1) was proposed as part of an automated or semi-automated inspection system [2]. MARC is provided with magnetic tracks, which make it able to crawl along vertical slopes, such as the vessel hull, carrying a number of sensors to perform inspection tasks. In [3] MARC was demonstrated to be successful in performing thickness measurements using an electrical robotic arm and an ultra-sound probe.

In [4] an image-processing model for rust detection using digital images of metal has been presented. The acquisition required cameras and sensors to take a representative picture of the surface of metallic pieces. A simple Bayesian classifier using mean descriptor was used to classify oxide-containing from oxide-free regions. Real images were used in order to measuring the performances of the rust detector model in real industrial situations with limited acquisition conditions; simulated images with Perling Noise were used for exposing the detector to extreme conditions of corrosion.

In this work MARC has been used for visual close-up surveys of vessel hulls using an on-board low-cost monocular camera and our main objective was to develop a novel intelligent system that enables new levels of performance in rust detection using images acquired by the on-board camera. Pictures from real materials under extreme corrosion conditions have been used in this study (see Figure 2). Firstly, the effectiveness of existing learning algorithms, popular in main application domains, was analyzed and evaluated for the rust detection, compared with the Bayesian classifiers used in [4]. In particular, Support Vector Machine (SVM) [5], [6], Random Forest (RF) [7], [8], Fisher and Quadratic Discriminant Analysis (DA) [9], [10] and K-Nearest Neighbor (K-NN) classifiers [11], [12] were used to classify images according to color moments, i.e. mean, variance, skewness and kurtosis for each HSV channel. We estimated statistical measures (accuracy, sensitivity, specificity, positive error, negative error, Positive Predictive Value (PPV), Negative Predictive Value (NPV)) of the performance of all the classifiers through the cross validation strategy. The best result, in terms of accuracy, was obtained by SVM followed by RF, while Bayesian classifier was the most sensitive algorithm. For all the algorithms, similar positive and negative errors of classification were obtained using the four color moments to represent images, instead of the color mean as proposed in [4].

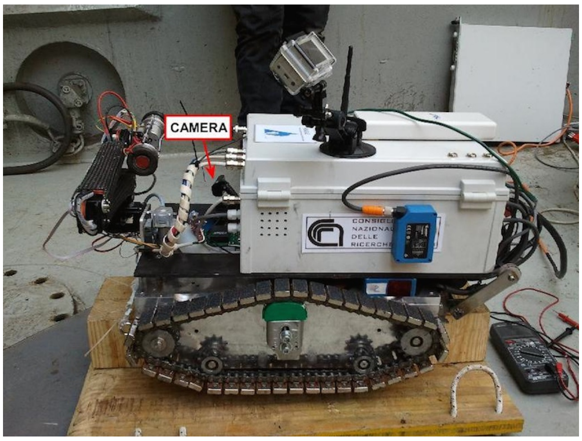

Fig. 1 The Magnetic Autonomous Robotic Crawler (MARC).

One of the most attractive areas of research in machine learning has been to study methods for constructing classifier ensemble which is a set of base classifiers whose individual predictions are combined in some way to classify new examples. The main discovery is that ensembles are often much more accurate than the individual classifiers that make them up [13]. Along these lines we propose a novel and highly accurate tool, Parallel multIple Classifier system for Accurate Rust Detection (PICARD), designed to detect materials subject to corrosion and rust, using images acquired by the MARC on-board camera. The PICARD is an ensemble of SVM, Bayesian classifier and RF, conceived for parallel and distributed computing to maintain low the computational cost. The individual classifiers were designed independently, trained on the same data, and their outputs were combined using major voting strategy. Ensembles of classifiers are nowadays widely used and outperform single classifiers in several applications, i.e. land cover change detection, hyperspectral data classification, IRIS data classification, hand-written digit recognition, medical decision support, and the fraud detection [14], [15], [16], [17]. The success of the ensembles depends on large extent on the proper selection of diverse classifiers for incorporation [18]. However, the diversity for classification task, largely discussed by Kuncheva and Withaker in [19], is not yet quite well defined and a better understanding of diversity could be expected to lead to higher ensemble accuracy [18], [20], [21]. In the current literature there are many studies on MCS and diversity of the classifiers, and, to the best of our knowledge, none of them was applied to the rust detection. In this study, SVM, Bayesian classifier and RF showed optimal combinations for the non-symmetric diversity, confirming that these measures, grouped together, exhibit a slightly higher correlation with the ensemble accuracy as previously hypothesized in [19]. The performance of PICARD, with a prediction accuracy of 96.1%, was superior to those of the standard classifiers for rust detection.

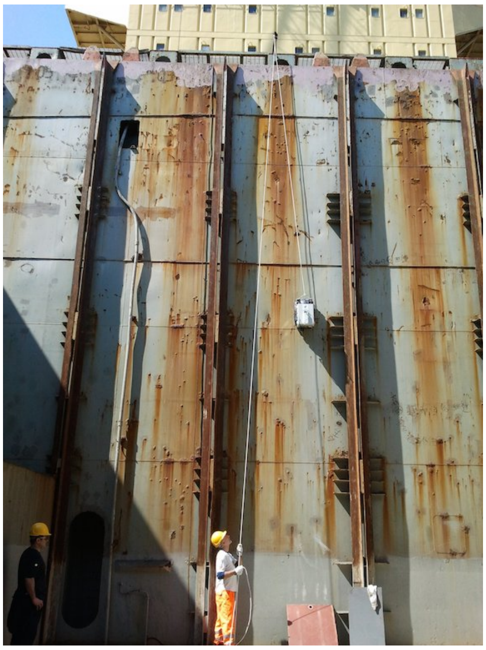

Fig. 2 Experimental setup: MARC climbing a ship hold wall during a typical inspection operation. Note that the rope visible in the pictures is just a safety measure, i.e. it would not support the vehicle under normal working condition.

2 Materials and Methods

2.1 The Robotic Inspection System MARC

The proposed inspection system features a Magnetic Autonomous Robotic Crawler (MARC) [2] (see Figure 1). MARC is provided with magnetic tracks, which make it able to crawl along vertical slopes, while carrying various sensors for autonomous navigation and data gathering, including a monocular camera. The latter consists of a low-cost Mediacom USB digital webcamera mounted on a tilt support at a height of approximately 20-25 cm from the inspected surface, with a tilt angle ranging between 50-70 deg. Figure 2 shows MARC during a typical inspection operation in a shipyard in Varna, Bulgaria. The rope visible in the pictures is just a safety measure, i.e., it would not support the vehicle under normal working conditions. The experiments consisted in MARC's following a vertical stiffener frame, while climbing a bulkhead. Images were acquired by the on-board camera at a resolution of 320 x 240 pixels and a frame rate of 10 Hz and were processed offline.

2.2 Data set description and Features

From the images acquired by MARC on-board camera, a subset of 23 salient frames were selected as the most visibly informative in terms of rust content. In this subset, 16×16 pixels sub-images or blocks, assigned to rust and non-rust classes, were extracted. The block size resulted as trade-off between the consideration that larger size be not suitable to distinguish wether or not the surface is rusted and smaller size increases the computational cost. We assembled a data set containing 113 blocks, of which 61 were rust samples and 52 non-rust samples.
Color moments (Mean, Variance, Skewness and Kurtosis) have been calculated in each channel of the HSV sub-images by the following equations:

$$E(X) = \frac{1}{N}\sum_{n=1}^{N} x_n \tag{1}$$

$$E(((X - E(X))^2) = \frac{1}{N}\sum_{n=1}^{N}(x_n - E(X))^2 \tag{2}$$

$$E(((X - E(X))^3) = \frac{1}{N}\sum_{n=1}^{N}(x_n - E(X))^3 \tag{3}$$

$$E(((X - E(X))^4) = \frac{1}{N}\sum_{n=1}^{N}(x_n - E(X))^4 \tag{4}$$

where x_n is the frequency for each pixel value and $n = 1, 2, ..., N$ refers to imaging quantization. The mean (eq. 1) indicates where the individual color generally lies in the HSV color space. The second moment (eq. 2) incorporates the information on the spread or scale of the color distribution. Non corroded surfaces are often homogeneous and they imply low variance. The third moment (eq. 3) measures the asymmetry of the data around the sample mean and indicates when the HSV values lie toward maximum or minimum in the scale. The fourth moment (eq. 4) measures the flatness or peakedness of the color distribution [22].

2.3 Proposed classification approach PICARD

In this section the proposed classification scheme is discussed. Generating an ensemble of classifiers, called Multiple Classifier System (MCS), is one of the most promising directions in pattern recognition which gained a lot of interest in the recent years. MCS is a set of pattern classifiers whose individual decisions are integrated, according to a certain combination approaches, to classify new examples. MCS are viewed as one effective way to improve classification performances [18]. Why do we need ensemble of classifiers? The answer lies in the Condorcet's jury theorem which refers to a jury of voters who need to make a decision regarding a binary outcome. If each voter has a probability p of being correct and the probability of a majority of voters being correct is L then:

- $p > 0.5$ implies $L > p$
- Also L approaches 1, for all $p > 0.5$ as the number of voters approaches infinity.

This theorem has two major limitations: the assumption that the votes are independent; and that there are only two possible outcomes. Nevertheless, if these two preconditions are met, then adding more voters whose judgments are better than a random vote increases the probability of the majority decision to be correct.

The success of the MCS depends on large extent on the proper selection of diverse classifiers for incorporation [18], in the sense that the classifiers in the ensemble should make different errors on unseen data. The diversity for classification task is not quite well defined and a better understanding of diversity could be expected to lead to higher ensemble accuracy [18], [20], [21]. In the classification problems, there have been several attempts to define diversity measures which can be categorized into two groups: pairwise and non-pairwise measures. The first are computed for each pair of classifiers in the ensemble and then averaged. The non-pairwise measures either use the concept of entropy or correlation of individual outputs with the averaged output of the ensemble or are based of the "difficulty" of the data points. Assuming the classifier output equals to 1 (correct decision) if the example is correctly recognized by the learner, and 0 (incorrect decision), otherwise, the measures of diversity should be symmetrical with respect to swapping 0 and 1 [19], [23]. In [19] Kuncheva and Withaker showed that, although it was difficult to established if there is a measure that is best for proposes of developing committees that maximize accuracy, the non-symmetrical measures exhibited a slightly higher correlation with the team accuracy and tended to be grouped together. As suggested, in this work we evaluated three non-pairwise and non-symmetrical diversity measures for constructing ensemble: Measure of difficulty, Generalised diversity, Coincident failure diversity. The problem of measuring the team diversity and how to use it for effectively building better classifier ensemble is out of the scope of this work. Table 1 shows a summary of the three diversity measures, including their types and literature source. For complete definition of these diversity measures see [19].

Moreover, there are different methods to combine the outputs of the voters and make the final decision in an ensemble of classifiers. The proposed classification tool, PICARD (see figure 3), used the parallel combination where multiple classifiers are independently designed without any mutual interaction and their outputs are combined according to majority vote strategy to make the final decision. Majority vote is the most popular ensemble approach where each classifier votes for specific class and has equal importance in the ensemble, then the class collecting the majority votes is the one predicted by the ensemble. The parallel approach is a way to reduce computational cost of the processing. The

Table 1 Summary of the non-pairwise and non-symmetrical diversity measures of classifier ensemble. The measures assume values in the range [0,1]. The + means that diversity is greater when the measure is larger, and the - means that diversity is greater when the measure is smaller.

Name	range of values	+/−	Reference
Measure of difficulty	[0, 1]	−	[24]
Generalised diversity	[0, 1]	+	[25]
Coincident failure diversity	[0, 1]	+	[25]

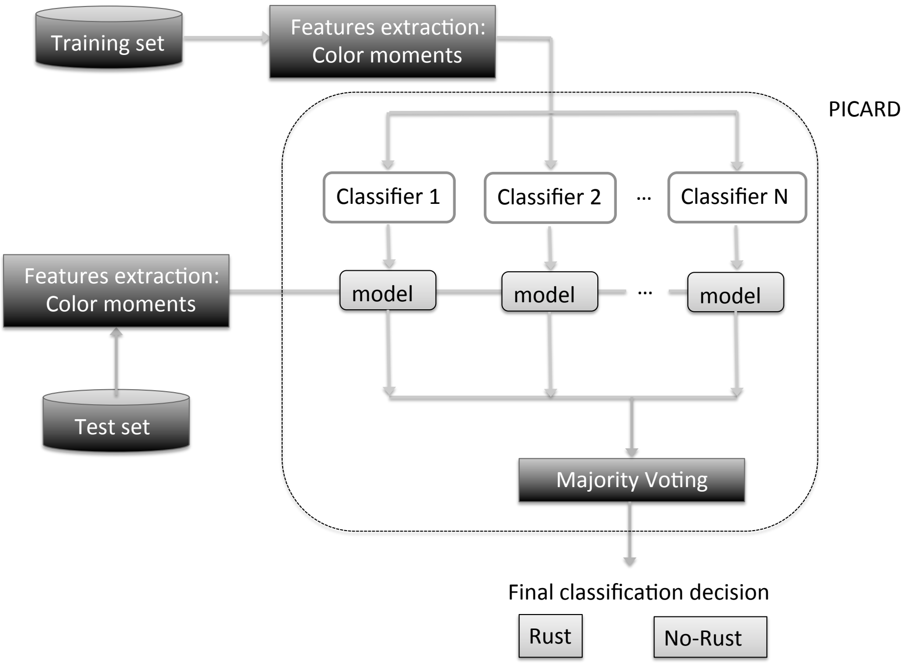

Fig. 3 One round of the cross validation technique employed to evaluate the performances of the PICARD algorithm for rust detection

performances of the PICARD were estimated using cross validation strategy, each round of which involved partitioning a sample of data into complementary subsets, training and test sets, building the classifier on the first set, and validating the model on the second set. To reduce variability multiple rounds of cross-validation were performed using different partitions, and the validation results were averaged over rounds.

2.4 Evaluation of Experimental Performance

Statistical measures of the performance of a binary classifier, used in this work, are accuracy, sensitivity, specificity, positive error, negative error, Positive Predictive Value (PPV), Negative Predictive Value (NPV). These metrics can be easily derived by the confusion matrix as follows:

accuracy = (TP+TN)/(TP+FN+FP+TN)	The percentage of predictions that are correct.
sensitivity = TP/(TP+FN)	The percentage of positive labeled instances that were predicted as positive.
specificity = TN/(FP+TN)	The percentage of negative labeled instances that were predicted as negative.
positive error = FN/(TP+FN)	The percentage of positive labeled instances that were predicted as negative.
negative error = FP/(TN+FP)	The percentage of negative labeled instances that were predicted as positive.
PPV = TP/(TP+FP)	The percentage of positive predictions that are correct.
NPV = TN/(FN+TN)	The percentage of negative predictions that are correct.

where TP is the number of True Positive, i.e. the actual positive data that are correctly classified, FP is the number of False Positive, i.e. negative data classified as positive, TN is the number of True Negative, i.e. the actual negative data that are correctly classified, and FN is the number of False Negative, i.e. positive data classified as negative. The intuitive meaning of each measure is also reported.

3 Results and Discussion

All of the experiments have been effected on a Workstation HP Z820 equipped with 2 CPU Intel Xeon and eight cores E5-2650, RAM 64Gb-2x1000Gb. All data were analyzed in MatLab (MathWorks, Natick, MA).
A cross validation technique was used in order to estimate how accurately a predictive model will perform in practice on new examples. As suggested in [4],[8],[26], 2/3 of data were used for the training set and 1/3 for the test set each set containing the same ratio of rusted and non-rusted samples, and the evaluation metrics were estimated performing 1500 rounds of cross validation. Six algorithms, widely employed in many areas of pattern classification, were used for rust detection: SVM, Bayesian classifier, RF, Fisher and Quadratic DA, K-NN classifiers. Internal cross validation on training set was used to implement the parameters tuning for SVM and RF in a wide range of parameter values, with the goal of optimizing measures of the algorithm's performance on an independent data set. The regularization parameter of the SVM classifier was tuned in the range [0.01, 5000] with linear kernel function, and the

optimal value was equal to 0.1; parameter tuning of RF was performed using a number of trees in the range [1, 10000] and the optimal number resulted to be 1000. Figure 4 shows the misclassification rate of the six classifiers and the summary of the experimental results on the rust data set is shown in table 2. The best performance, in term of accuracy, was that of SVM classifier (accuracy = 0.954), followed by RF classifier (accuracy= 0.939). Among the six classifiers, Bayesian classifier has the highest percentage of positive labeled instances (rust samples) that were predicted as positive (rust) (sensitivity = 0.984), however it has a percentage of negative labeled instances (no-rust samples) that were predicted as negative (no-rust) (specificity = 0.829) much lower than SVM and RF.

Table 2 Statistical measures of the performances of PICARD, SVM, Bayesian classifier, RF, Quadratic DA, Fisher DA and k-NN classifier. The analysis was performed on the rust data set using 1500 CV and the regularization parameter of SVM equals to 0.01 and number of trees of RF equals to 1000.

Indicator	**PICARD**	SVM	Bayesian classifier	RF	Quadratic DA	Fisher DA	k-NN
accuracy	**0.961**	0.954	0.907	0.939	0.925	0.921	0.627
sensitivity	**0.989**	0.968	0.984	0.947	0.966	0.944	0.627
specificity	**0.932**	0.940	0.829	0.930	0.884	0.899	0.628
positive error	**0.011**	0.032	0.016	0.053	0.034	0.056	0.373
negative error	**0.068**	0.059	0.171	0.070	0.116	0.101	0.372
PPV	**0.936**	0.942	0.852	0.932	0.893	0.903	0.628
NPV	**0.988**	0.967	0.981	0.946	0.963	0.941	0.627

In [4] the authors proposed a system, based on Bayesian classifier, to identify rusted and non-rusted areas, using mean descriptor from the HSV image, with the aim of reducing the computational cost and maximizing the usability of the proposed methodology in fast placed industrial settings. To be consistent with real conditions of image acquisitions, the behavior of the proposed system has been studied under the influence of the variance of Additive White Gaussian Noise (AWGN) (summary of the results shown in table 3). The authors concluded that the best performing system was the one without noise added. In fact, as the sensitivity increases with the added noise, so does the negative error, therefore the specificity decreases. This behavior yielded an increase of the classification errors since the detector classified all the regions into the rust class, including the non-rust areas. We evaluated the performance of SVM and Bayesian classifier using the only mean descriptor. The SVM parameter was newly tuned in the range [0.01, 5000], obtaining the optimal value C = 0.9. The performance indicators of SVM and Bayesian classifiers are shown in table 3. We obtained sensitivities (0.969 for SVM and 0.973 for Bayesian classifier) similar to those of table 2 (0.968 for SVM and 0.984 for Bayesian classifier), while much lower specificities (0.776 for both SVM and Bayesian classifier) were obtained compared with those of table 2 (0.940 for SVM and 0.829 for Bayesian classifier). Experimental results suggested that using the four color

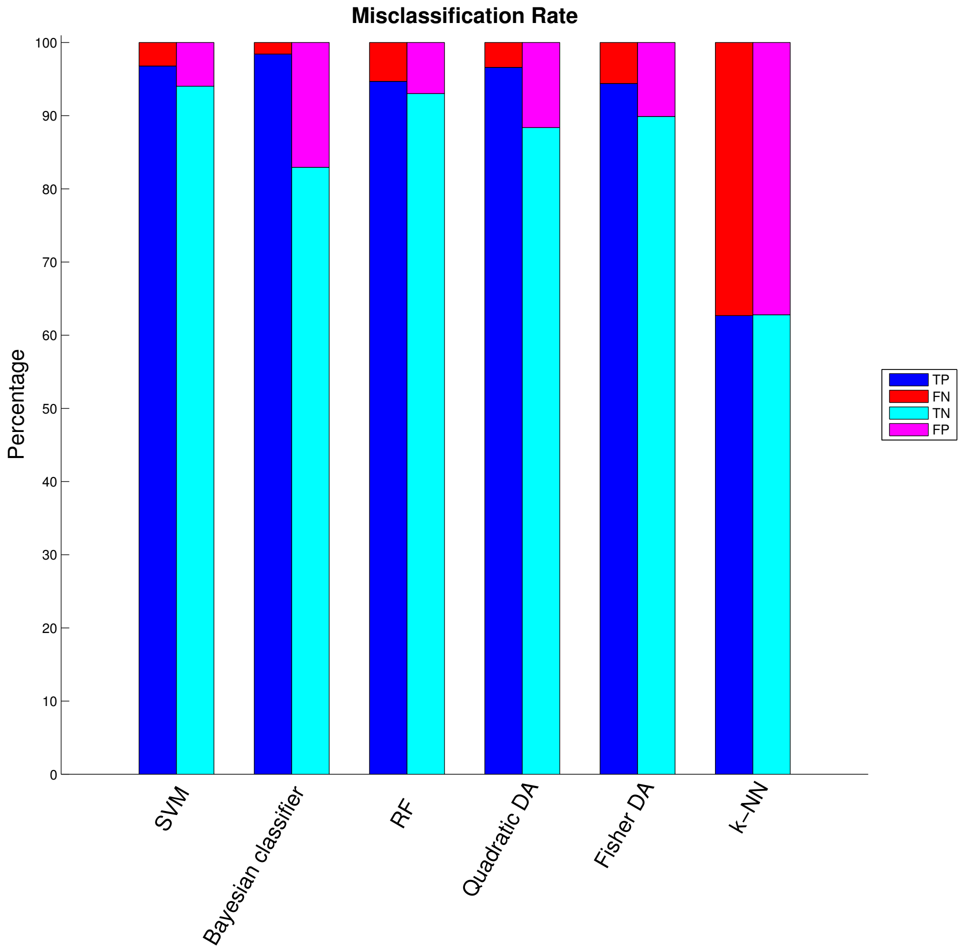

Fig. 4 Misclassification rate of SVM, Bayesian classifiers, RF, Quadratic DA, Fisher DA, K-NN classifiers.

descriptors (mean, variance, skewness and kurtosis) significantly improve classification accuracy and specificity. Note that in our experiments extremely rusted surfaces are available, so we can validate algorithms performance using images acquired by the MARC on-board camera, without introducing any kind of simulation and added noise.

Table 3 Performance indicators for SVM and Bayesian classifiers trained using the only mean descriptor on the rust data set. Last four columns report the performance indicators of the system proposed in [4] in different cases of AWGN.

Indicator	SVM	Bayesian classifiers	Noise-free in [4]	r=0.001 in [4]	r=0.002 in [4]	r=0.005 in [4]
accuracy	0.872	0.875	-	-	-	-
sensitivity	0.969	0.973	0.979	0.969	0.978	1
specificity	0.776	0.776	0.982	0.971	0.914	0.152
positive error	0.031	0.020	0.021	0.030	0.021	0
negative error	0.224	0.224	0.017	0.028	0.085	0.848
PPV	0.812	0.813	0.982	0.971	0.919	0.541
NPV	0.962	0.967	0.978	0.969	0.976	1

The empirical evidences suggested that a careful combination of classifiers should potentially compensate for the individual classification error and thus achieve better robustness and performance. We used a novel tool, PICARD, that is a parallel combination of SVM, Bayesian classifiers and RF, using the majority voting strategy for the fusion of the classifiers predictions. The selection of these three classifiers from the six base voters were done using three diversity measures (see table 1 and 4): Measure of difficulty, Generalised diversity, Coincident failure diversity. These measures should not be a replacement for the estimate of the team accuracy but should be stem from the intuitive concept of diversity. Starting from the six classifiers, all teams of three classifiers were generated, forming 20 classifier ensembles, and the diversity measures were computed for these ensembles. The team composed by SVM, Bayesian classifier and RF showed optimal values for the three diversity measures. In fact the Measure of difficulty, varying in the interval $[0.0045, 0.0726]$, was equal to 0.0045 for the SVM, Bayesian classifier and RF ensemble, very near to the optimal value of 0. Both the Generalised diversity and Coincident failure diversity assumed values from 0.5 to 1, and for the selected team they were equal to 1. The higher are the their values, the greater is the diversity.
The summary of the experimental results of PICARD are shown in the first column of the table 2. PICARD achieved an accuracy of 0.961 better than that of individual SVM, Bayesian classifier and RF. Moreover, PICARD showed a positive error of 0.011 lower than the minimum positive error of 0.016 obtained by Bayesian classifier, and much lower than that achieved by both SVM (positive error = 0.032) and RF (positive error = 0.053). The PICARD negative error of 0.068 was much lower than the one of the Bayesian classifier (negative error = 0.171) and very similar to the SVM (negative error = 0.059) and RF (negative error = 0.070) negative errors.

Table 4 Summary of the non-pairwise and non-symmetrical measures of diversity for SVM, Bayesian classifier and RF.

Name	value
Measure of difficulty	0.0045
Generalised diversity	1
Coincident failure diversity	1

Finally the proposed PICARD was applied to detect rust areas of new images acquire by MARC on-board camera. Working with large images, normal image processing techniques can sometimes break down. To overcome this limitation, the new image has been processed incrementally, that is reading and classifying one region of the image at time, where the region was defined as a block of size 5×5, and then assembling the results into an output BW image (mask). The classification of each block was performed by the PICARD, using the 1500 machines previously trained in the cross validation procedure, using the major voting strategy for combining the predictions. Rust detection successfully obtained on four novel images by PICARD is shown in 5.

4 Conclusions

In this paper a vision-based robotic system has been used for inspection of ship hulls and a novel intelligent system for rust detection, PICARD, has been developed and tested on images acquired by the on-board camera. Using diversity measures as guidance for the selection of multiple classifiers combination, the optimum was an ensemble of SVM, Bayesian classifier and RF. The experimental results showed that a decision support system based on PICARD achieved very high accuracy and overall satisfactory performance if compared with those of existing and popular learning algorithms. The developed tool is general and can be applied for the quality assessment of metallic pieces and iron machines.

Based on our analysis, we identified that the incorporation of diversity among the classifiers can help to obtain better classification by the team. A thorough analysis of all the diversity measures and of the relationship between diversity and ensemble accuracy are out of the scope of this work. However we experimentally highlighted the effectiveness of the non-pairwise and non-symmetrical measures of diversity for constructing ensembles with good generalization performance, also validating our system on novel full images acquired by the MARC on-board camera.

A limitation of our approach is the number of classifiers in the ensemble, in fact increasing this number usually increases the computational coast and decreases their comprehensibility, and in general it depends on the computer's architecture available for parallel learning.

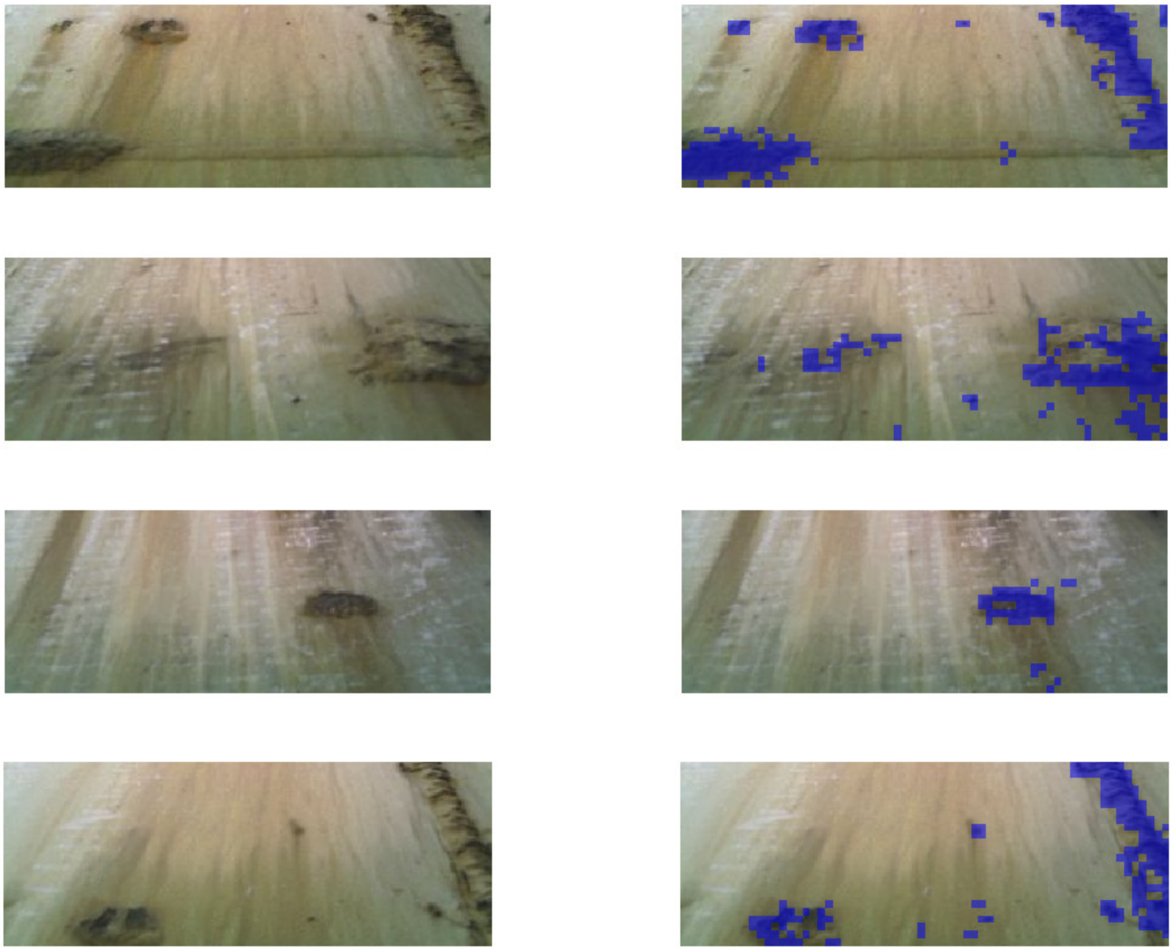

Fig. 5 Rust areas automatically detected by the PICARD on new test images.

Future objective of this work will be to develop classification framework, based on PICARD, for online detection of corrosion areas during completely automatic marine vessel inspection by robotic systems.

Acknowledgements This work has been funded by the EC FP7 SST.2008.5.2.1 Innovative Product Concepts project: Marine INspection rObotic Assistant System, grant n. SCP8-GA- 2009-233715-MINOAS, and by the BANDIERA (Best Action for National Development of International Expert Researchers Activities) project RITMARE. The authors thank Giorgio Bruzzone, Edoardo Spirandelli and Mauro Giacopelli for their extraordinary contribution to the design, development and test of the MARC platform, and Arturo Argentieri for technical assistance. Special thanks to the MINOAS consortium and, in particular, to Albena Todorova and all the personnel of Dolphin shipyards who supported with their high professionalism and kindness the field validation of the system.

References

1. M. Caccia, R. Robino, W. Bateman, M. Eich, A. Ortiz, L. Drikos, A. Todorova, I. Gaviotis, F. Spadoni, V. Apostolopoulou, in *Proc. of IAV 2010, 7th IFAC Symposium on Intelligent Autonomous Vehicles* (2010)
2. M. Bibuli, G. Bruzzone, G. Bruzzone, M. Caccia, M. Giacopelli, A. Petitti, E. Spirandelli, in *Conference on Computer Applications and Information Technology in Maritime Industries (COMPIT), Liegi (Belgio)* (2012), pp. 62–75
3. M. Eich, F. Bonnin-Pascual, E. Garcia-Fidalgo, A. Ortiz, G. Bruzzone, Y. Koveos, F. Kirchner, Journal of Field Robotics **31(2)**, 319 (2014)
4. M. Acosta, J. Daz, N. Castro, Corrosion Science. **88**, 141 (2014)
5. V. Vapnik, IEEE Trans on Neural Networks **10**(5), 988 (1999)
6. N. Ancona, R. Maglietta, E. Stella, Pattern Recognition **39**(9), 1588 (2006)
7. L. Breiman, Machine Learning **45**, 5 (2001)
8. R. Maglietta, N. Amoroso, M. Boccardi, S. Bruno, A. Chincarini, G. Frisoni, P. Inglese, A. Redolfi, S. Tangaro, A. Tateo, R. Bellotti, Pattern Analysis and Application pp. 1–13 (2015)
9. Y. Guo, T. Hastie, R. Tibshirani, Biostatistics **8**(1), 86 (2007)
10. T. Hastie, R.Tibshirani, J.Friedman, *The elements of statistical Learning - Data Mining, Inference and Prediction* (Springer, 2008)
11. T. Cover, P. Hart, IEEE Trans. Inform. Theory **13**, 21 (1967)
12. K. Hattori, M. Takahashi, Pattern Recognition **32**(3), 425 (1999)
13. T. Dietterich, AI Magazine **18**(4), 97 (1997)
14. H. Nemmour, Y. Chibani, ISPRS J. Photogramm. Remote Sens. (2006)
15. X. Ceamanos, B. Waske, J. Benediktsson, J. Chanussot, M. Fauvel, J. Sveinsson, Int. J. Image Data Fusion **1**, 293 (2010)
16. H. Kim, S. Pang, H. Je, D. Kim, S. Bang, Pattern Recognition **36**(12), 2757 (2003)
17. K. Jackowski, B. Krawczyk, M. Woniak, Intelligent Data Engineering and Automated Learning IDEAL **7435**, 850 (2012)
18. P. Du, J. Xia, W. Zang, K. Tan, Y. Liu, S. Liu, Sensor **12(4)**, 4764 (2012)
19. L. Kuncheva, C. Whitaker, Machine Learning **51**, 181 (2003)
20. R. Banfield, L. Hall, K. Bowyer, W. Kegelmeyer, Information Fusion **6**(1), 49 (2005)
21. G. Brown, J. Wyatt, R. Harris, X. Yao, Journal of Information Fusion **6(1)** (2006)
22. F. Medeiros, G. Ramalho, M. Bento, L. Medeiros, EURASIP J. Adv. Signal Process **2010** (2010). DOI 10.1155/2010/817473
23. D. Ruta, B. Gabrys, *Multiple Classifier Systems: Second International Workshop, MCS 2001 Cambridge, UK, July 2–4, 2001 Proceedings* (Springer Berlin Heidelberg, 2001), chap. Application of the Evolutionary Algorithms for Classifier Selection in Multiple Classifier Systems with Majority Voting, pp. 399–408
24. L. Hansen, P. Salamon, IEEE Transactions on Pattern Analysis and Machine Intelligence **12**(10), 993 (1990)
25. D. Patridge, W. Krzanowski, Information and Software Technology **39**, 707 (1997)
26. N. Ancona, R. Maglietta, A. piepoli, A. D'Addabbo, R. Cotugno, M. Savino, S. Liuni, M. Carella, G. Pesole, F. Perri, BMC Bioinformatics **6** (2005)