Operative Perception and Environment Characterization

Lia Maglietta

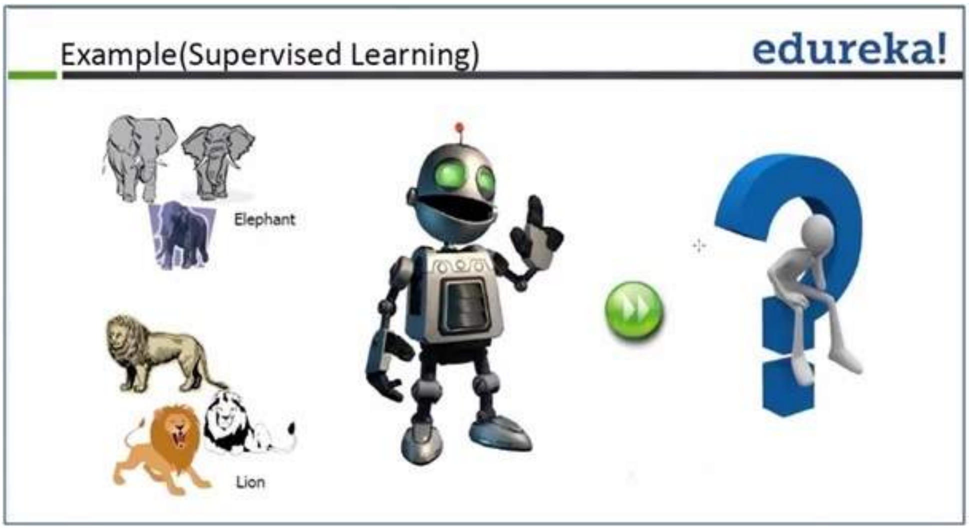
July 20th, 2016

Problem of supervised classification

MACHINE LEARNING uses algorithms that iteratively learn from data, allowing computers to find hidden insights without being explicitly programmed where to look.

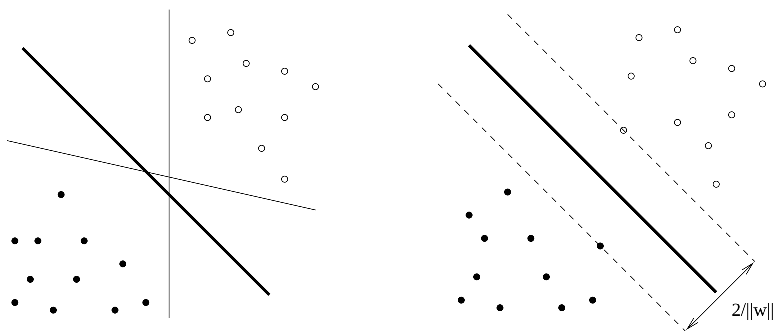
- Supervised learning algorithms are trained using labelled examples, build a model that can make predictions of the response value for new examples. Classification: for categorical response values, where the data can be separated into specific classes; Regression: for continuous-response values.
- Unsupervised learning is used against data that has no historical labels.

Problem of supervised classification



Support Vector Machine SVM

SVM seeks a hyperplane with a large margin, and minimizes the number of wrongly classified training data by using a regularization parameter C.



V. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, INC., 1998.

Support Vector Machine SVM

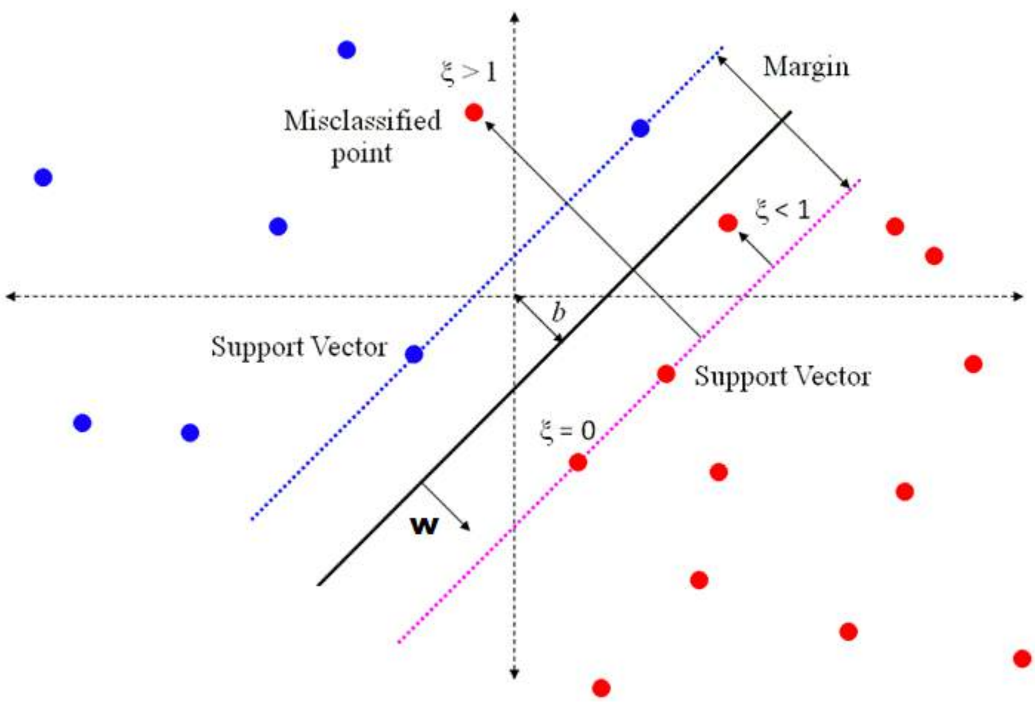
In the general hypothesis of non linearly separable classes, the optimal hyperplane $f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$, found by SVM is the following:

Problem P1

$$\begin{array}{ll} \min & \frac{1}{2}\mathbf{w} \cdot \mathbf{w} + C \sum_{i=1}^{m} \xi_i \\ \textit{subject to} & y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \;\; i = 1, \cdots, m \end{array}$$

where ξ_i are non-negative slack variables which take into account the misclassified training examples, $\frac{2}{||\mathbf{w}||}$ is the margin of the hyperplane, b is a scalar bias term.

Support Vector Machine SVM



Support Vector Machine SVM

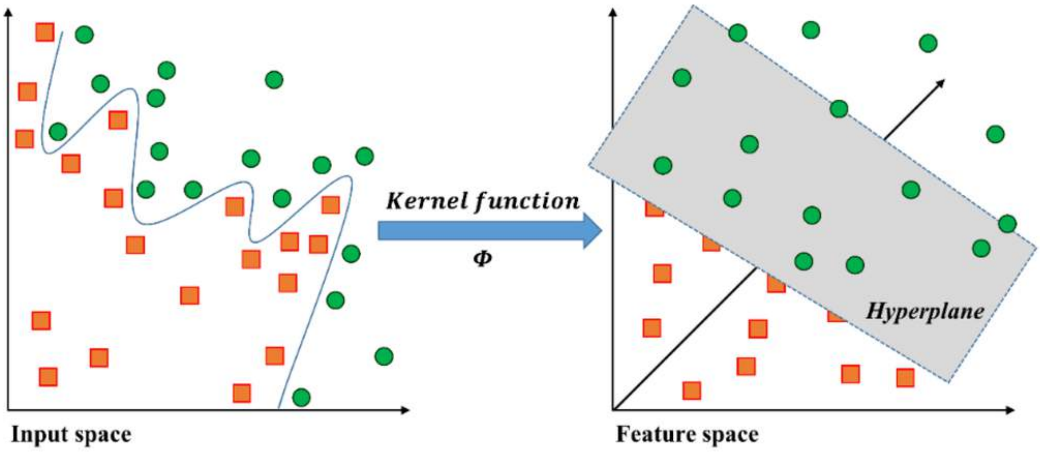
The classification of new data $\mathbf{x}$ involves the evaluation of the decision function:

$$y = \mathrm{sign}(f(\mathbf{x})) = \mathrm{sign}(\sum_{i=1}^{m} \lambda_i^* y_i \mathbf{x}_i \cdot \mathbf{x} + b^*) \quad (1)$$

The class $y \in \{-1, 1\}$ of $\mathbf{x}$ is expressed evaluating the dot product between the data and the support vectors of the training set S.

Support Vector Machine SVM

In case of non-linear SVM a kernel function is used.



Estimate of the generalization error

The ultimate goal of any classifier is *to generalize*.
The central problem is not classifying the training data in the training set S, because any sufficiently complex learning machine could separate S without errors.
The crucial problem is to design classifiers having low error rate on new data.

Cross Validation

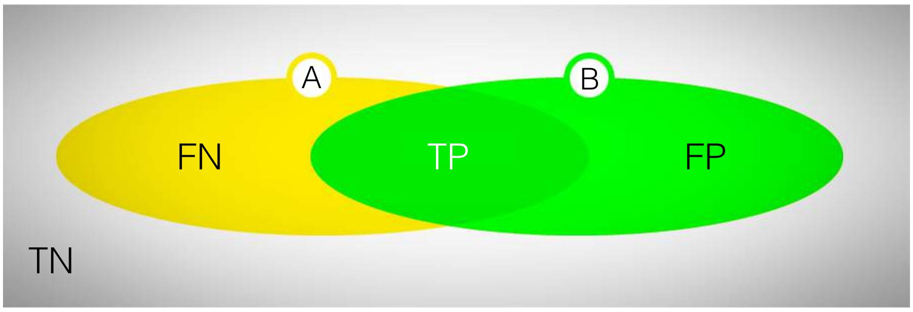
Leave-K-Out Cross Validation technique consists of building T_1 pairs of training and test sets composed of n and $\ell - n$ examples respectively by random sampling without replacement the data set S.
For each of these T_1 random splits, we evaluate the error rate e_{n_i} of the classifier trained on n examples, testing it on $\ell - n$ examples. So, the LKOCV error is given by:

$$error = \frac{1}{T_1}\sum_{i=1}^{T_1} e_{n_i}$$

Evaluation of Experimental Performance

From the confusion matrix, we define:

- True Positive (TP) = the actual positive data that are correctly classified
- False Positive (FP) = the negative data classified as positive
- True Negative (TN) = the actual negative data that are correctly classified
- False Negative (FN) = positive data classified as negative.



Evaluation of Experimental Performance

Statistical measures of the performance of a binary classifier can be easily derived as follows:

- error = (FP+FN)/(TP+FN+FP+TN)
- accuracy = (TP+TN)/(TP+FN+FP+TN)
- sensitivity = TP/(TP+FN)
- specificity = TN/(FP+TN)
- positive error = FN/(TP+FN)
- negative error = FP/(TN+FP)
- PPV = TP/(TP+FP)
- NPV = TN/(FN+TN)

Evaluation of Experimental Performance

Intuitive meaning:

- accuracy = The percentage of predictions that are correct
- sensitivity = The percentage of positive labeled instances that were predicted as positive
- specificity = The percentage of negative labeled instances that were predicted as negative
- positive error = The percentage of positive labeled instances that were predicted as negative
- negative error = The percentage of negative labeled instances that were predicted as positive
- PPV = The percentage of positive predictions that are correct
- NPV = The percentage of negative predictions that are correct.

Robotic Visual Inspection of Marine Vessels

Goal: development of a novel tool to classify sub-images as rust or non-rust using images acquired by Magnetic Autonomous Robotic Crawler (MARC) on-board camera

PICARD - Data set description

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. We assembled a data set containing 113 blocks, of which 61 were rust samples and 52 non-rust samples.

PICARD: Features extraction

Color moments (Mean, Variance, Skewness and Kurtosis) have been calculated in each channel of the HSV sub-images:

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

$$E(((X - E(X))^2) = \frac{1}{N}\sum_{n=1}^{N}(x_n - E(X))^2 \tag{3}$$

$$E(((X - E(X))^3) = \frac{1}{N}\sum_{n=1}^{N}(x_n - E(X))^3 \tag{4}$$

$$E(((X - E(X))^4) = \frac{1}{N}\sum_{n=1}^{N}(x_n - E(X))^4 \tag{5}$$

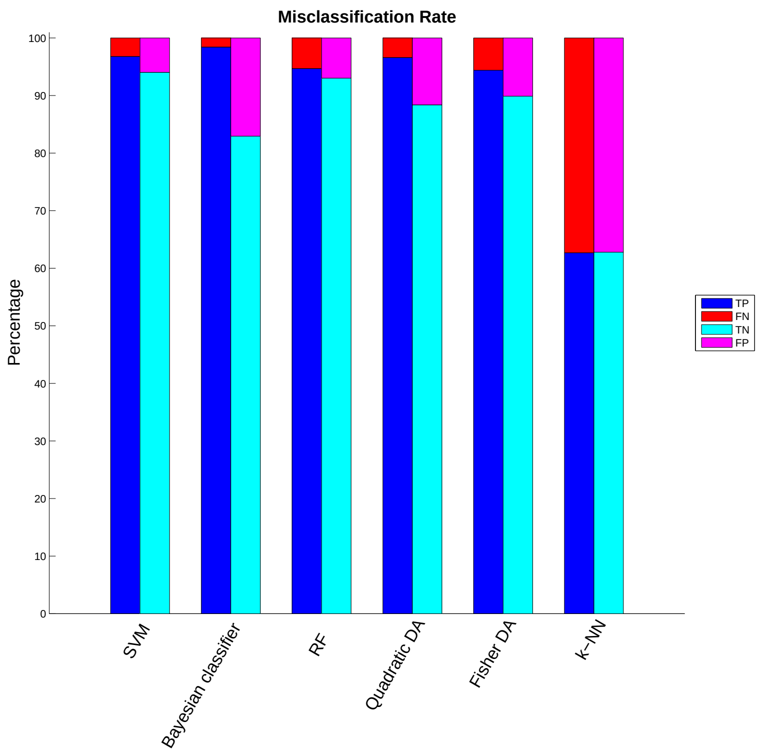
PICARD: Experimental results

SVM, Bayesian Classifiers (BC), Random Forest (RF), Quadratic Discriminant Analysis (QDA), Fisher Discriminant Analysis (FDA), K-Nearest Neighbor (K-NN) classifiers

Table: 1500 CV, the regularization parameter of SVM equals to 0.01 and number of trees of RF equals to 1000.

Indicator	SVM	BC	RF	QDA	FDA	k-NN
accuracy	0.954	0.907	0.939	0.925	0.921	0.627
sensitivity	0.968	0.984	0.947	0.966	0.944	0.627
specificity	0.940	0.829	0.930	0.884	0.899	0.628
positive error	0.032	0.016	0.053	0.034	0.056	0.373
negative error	0.059	0.171	0.070	0.116	0.101	0.372
PPV	0.942	0.852	0.932	0.893	0.903	0.628
NPV	0.967	0.981	0.946	0.963	0.941	0.627

PICARD: Experimental results



PICARD: Experimental results

Multiple Classifier System (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.

PICARD: Experimental results

Why do we need ensemble of classifiers? Condorcet's jury theorem: 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.

Two major limitations: the assumption that the votes are independent; there are only two possible outcomes.

PICARD: Experimental results

The success of the MCS depends on large extent on the proper selection of diverse classifiers for incorporation.
It is difficult to established if there is a measure that is best for proposes of developing committees that maximize accuracy.

Kuncheva and Whitaker, *Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy* Machine Learning 2003

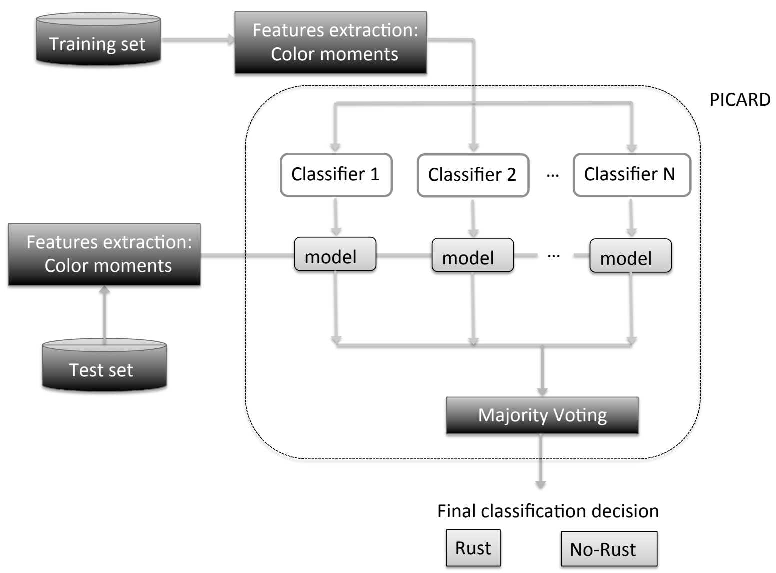
PICARD: Experimental results

Table: Summary of the diversity measures of SVM, BC and RF classifiers 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]	$-$	0.0045
Generalised diversity	[0, 1]	$+$	1
Coincident failure diversity	[0, 1]	$+$	1

PICARD: Experimental results

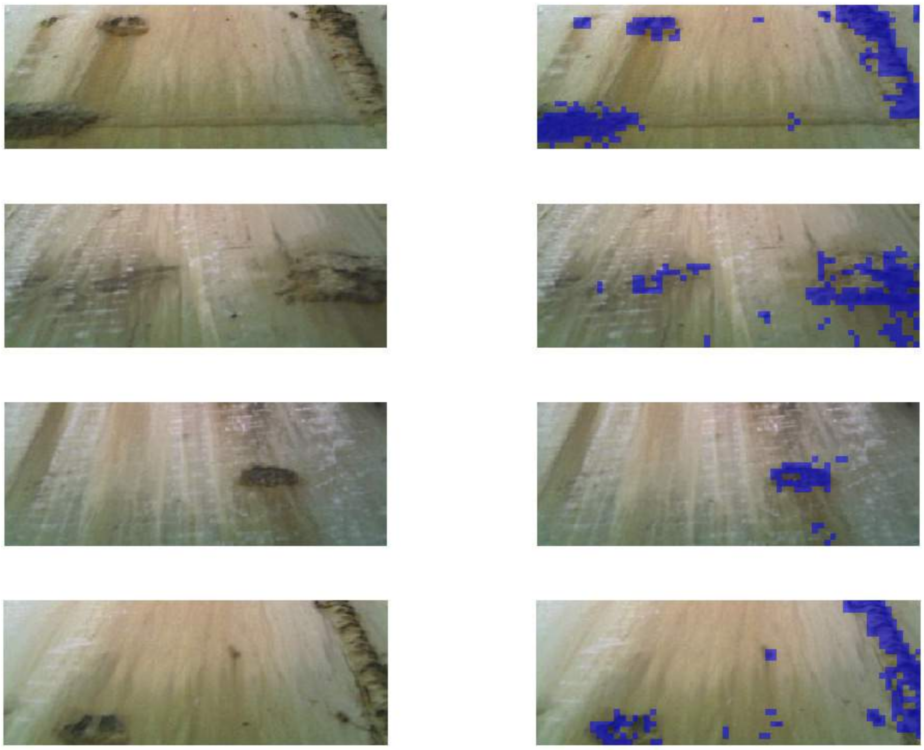
Parallel multIple Classifier system for Accurate Rust Detection (PICARD) is an ensemble of SVM, Bayesian classifier and RF, conceived for parallel and distributed computing to maintain low the computational cost.



PICARD: Experimental results

Indicator	**PICARD**	SVM	Bayesian classifier	RF
accuracy	**0.961**	0.954	0.907	0.939
sensitivity	**0.989**	0.968	0.984	0.947
specificity	**0.932**	0.940	0.829	0.930
positive error	**0.011**	0.032	0.016	0.053
negative error	**0.068**	0.059	0.171	0.070
PPV	**0.936**	0.942	0.852	0.932
NPV	**0.988**	0.967	0.981	0.946

PICARD: Experimental results



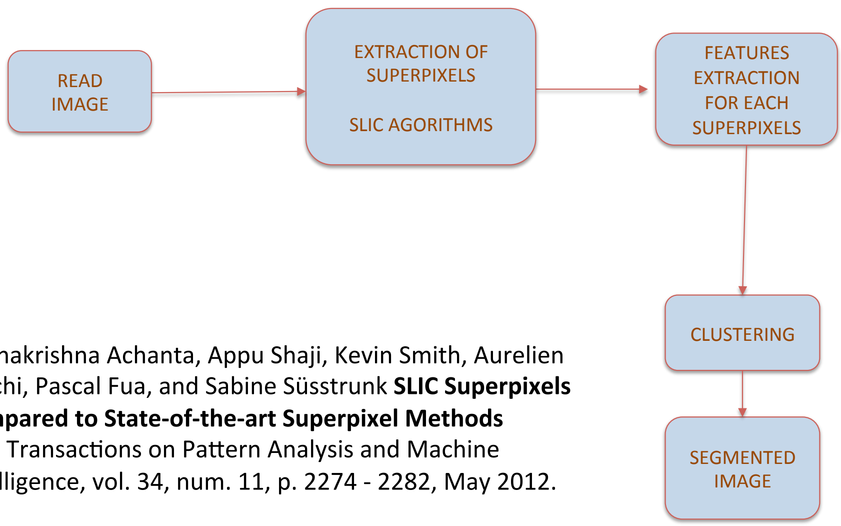
PICARD: Experimental results

PICARD: Experimental results

- R. Maglietta, A. Milella, M. Caccia and G. Bruzzone Parallel *Multiple Classifier System for Robotic Visual Inspection of Marine Vessels*
- A. Milella, R. Maglietta, M. Caccia and G. Bruzzone Parallel *Robotic visual inspection of ship hull surfaces using a magnetic crawler and a low-cost monocular camera*

Fast Segmentation System

The system is an integration of (Simple Linear Iterative Clustering) (SLIC) and Unsupervised learning algorithm.



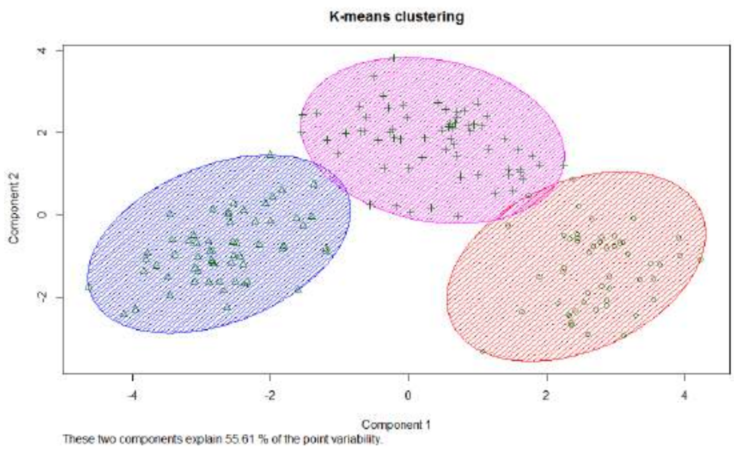
Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk **SLIC Superpixels Compared to State-of-the-art Superpixel Methods** IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, num. 11, p. 2274 - 2282, May 2012.

Fast Segmentation System

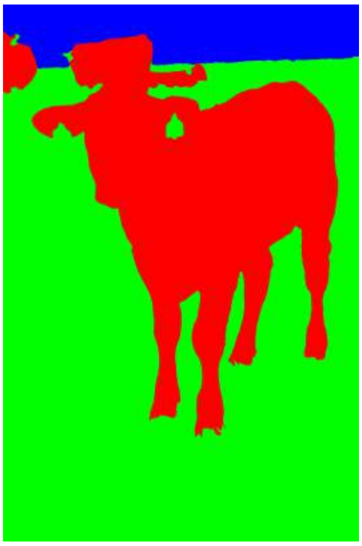
SLIC clusters pixels in the combined five-dimensional color and image plane space to efficiently generate compact, nearly uniform superpixels. The output is a desired number of regular, compact superpixels with a low computational.

Fast Segmentation System

Clustering: organization of unlabeled data into similarity groups called clusters based on specific metrics (i.e. euclidean distance). Data into cluster are "similar" between them and "dissimilar" to data items in other classes

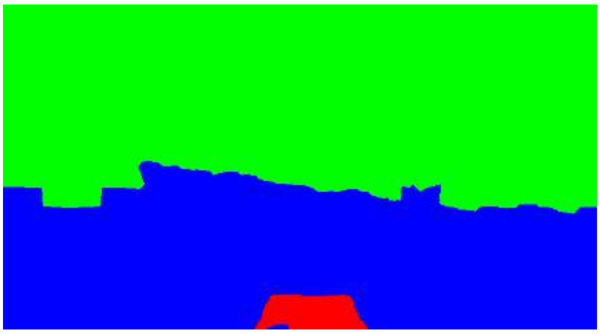


Fast Segmentation System



Fast Segmentation System

Automatic segmentation of images acquired in Polar Regions (Unmanned Vehicles for Autonomous Sensing and Sampling UVASS project).



Fast Segmentation System

The system is very practical, it has very low complexity and computational cost.
It should be very useful in several fields for fast segmentation of a large number of images.