Active Learning for Road Lane Landmark Inventory in Highly Uncontrolled Image Capture Conditions

JM Lopez-Guede, Asier Izquierdo, Julian Estevez, Manuel Graña

Computational Intelligence Group, UPV/EHU

Abstract

Road landmark inventory is becoming an important industry for the maintenance of transport infrastructures. Several commercial sensors are available which include optical cameras allowing to build panoramic images around the vehicle used for road inspection. We focus on a specific region of such panoramas which contains most of the relevant information. The landmark detection is posed as a two class classification problem that may be solved by some standard approaches, such as random forest (RF) and ensembles of extreme learning machines (V-ELM). Besides model parameter selection, a central problem is the construction of the labeled training dataset due to human labor cost and the highly uncontrolled conditions of image capture. We propose an open ended Active Learning approach with a human operator in the loop who can start the Active Learning process when detection quality is degraded by the change in imaging conditions in order to achieve adaptation to them. We report encouraging results over a collection of sample images selected from an industrial operation. As an additional contribution, we have assessed the ability of Active Learning to overcome the issues raised by highly class imbalanced datasets.

1. Introduction

Road landmark inventory is a flourishing industry around the world, as the traffic becomes more dense and the drivers must rely on a well maintained infrastructure. Specifically, horizontal signals and lane landmarks, such as lines, arrows or other drawings on the asphalt, are of great public concern. In this introductory section we present the problem definition and motivation, some introductory review of Active Learning, the description of the proposed approach, and finally the paper contributions and content.

1.1. Problem definition and motivation

The task ahead is driven by the industrial exploitation of a car mounted sensor nicknamed "ladybug" (figure 1). More precisely, the sensor is the IP-S3 HD1[1] product of Topcon, Japan. It is composed of a positioning system (wheel encoder, GPS receiver, Inertial Measurement Unit), five cameras pointing at regular arc intervals of the circumference, and a sixth one pointing up. with

---

[1] https://www.topcon.co.jp/en/positioning/products/product/3dscanner/IP-S3_HD1_E.html

Figure 1: The ladybug sensor mounted in a car

one pointing directly at the rear of the car, and a LiDAR sensor. The views of the cameras are composed into panoramas such as the one shown in figure 2, where the central view corresponds to the rear view of the car, so that the front view is splitted left and right of the panorama. At the bottom of the panorama it is possible to appreciate the scaffolding supporting the sensors. Notice the changes in hue and saturation in the image due to the different sensitivities of the cameras. The LiDAR data and the image capture has a frequency of 1Hz. The task ahead is to create an inventory of the road signals and landmarks using both LiDAR and image data. All images are tagged spatially with coordinates provided by onboard GPS. In this paper we deal with the extraction of the horizontal signals drawn on the road: lane limiting lines, arrows, and other horizontal signaling. The image capture conditions, specially the illumination, is wildly changing from one traveling capture to another, or during the same trip due to the changes in the position of the sun, the time of the year, the weather, etc. Besides, the road maintenance is often in bad condition, so the lines may be fading or interrupted. Finally, for supervised classification approaches the construction of the labeled dataset would be costly so we are extremely interested in exploiting cheaper alternatives, such as the Active Learning strategy[34].

Figure 2: An example panorama from the ladybug sensor, obtained by the composition of the views of the five cameras.

2

1.2. Active Learning

Classification approaches need careful selection and labeling of training data samples from the available data. In response to this issue, Active Learning [15] tries to achieve the most accurate classification using the smallest possible training set, minimizing the user interaction needed to label the training samples. Active Learning starts with a minimal training sample, adding new labeled samples in an iterative process. Aiming to provide the greatest increase in classifier accuracy [30], the additional samples are selected according to some classification uncertainty measure which does not require knowledge of the actual data label. Besides its benefits in economy of computation and data labeling, Active Learning is also useful when the underlying data statistics are non stationary, so that the classifier built at one time instant may not be optimal later on. Active Learning has been successfully applied to classification of remote sensing images [29, 34, 33], and image retrieval based on semisupervised Support Vector Machines [20]. Active Learning inspiration for the selection of a minimal collection of training images is proposed in [25] for the development of combined generative and discriminative models in the segmentation of CT scans. An active feedback approach is used in [32] to improve the classification based annotation of radiographs.

1.3. Proposed approach

In this paper we formulate the road landmark segmentation problem [26] as a pixel classification into road landmark and background classes. The experimental setup for validation is illustrated in Fig. 3. We collect the input panorama images provided by the so-called "ladybug" sensor to be described later. As shown in figure 2, the central part of the panorama contains most of the image useful information given by the rear view of the vehicle, so we crop this central part for processing, discarding the remaining parts of the panorama. We compute the Gabor features over the images collecting all the image features in a unique pool for the training of the classifiers and their validation. We apply an Active Learning strategy in order select the optimal training dataset. The classifier trained with the optimal training dataset is validated over the entire images, producing the performance report for the specific classifier. We repeat the validation for the diverse classifiers and classifier parameters explored. The Active Learning oracle providing sample labels in the reported experiments is the ground truth provided by manual segmentation. In the practical application the oracle will be the user through some graphical interface for the selection of most informative pixels to be added to the training dataset.

For pixel classification we explore the results of Random Forest (RF) [9, 11] and Ensembles of Extreme Learning Machines (V-ELM) classifiers based on texture features computed at pixel level. Specifically we apply a bank of Gabor filters, so that the feature vector of each pixel is composed of the Gabor coefficients plus some spatial localization information. We report performance results over a collection of road images in order to assess the most adequate classifier and parameter settings.

1.4. Paper contributions and content

Some specific contributions of the approach proposed in this paper relative to the state of the art of road image segmentation algorithms are: (1) Active
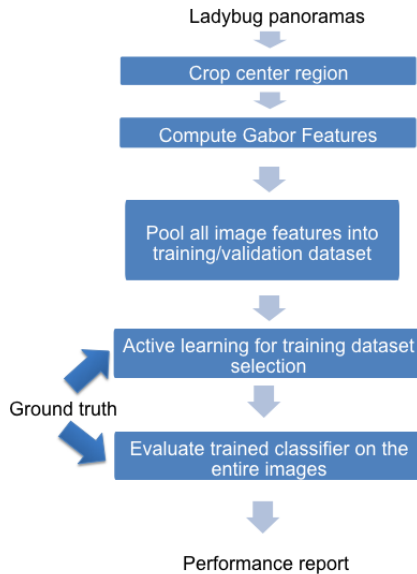
3

Figure 3: Pipeline of the experimental setup for the Active Learning segmentation process

Learning reduces the human intervention to the minimum in the process of training data selection and labeling, (2) we compare two extremely efficient and fast classifier approaches, namely RF and V-ELM, which allow quick adaptation to incremental training datasets, (3) the approach does not required *a priori* information or geometric models, (4) feature extraction is based on an specific systematic approach, i.e. Gabor filters (5) if we need to transfer the trained classifier to new data stream we only need to pick new training samples according to the Active Leaning approach, i.e. the process is an open ended learning process with a human in the loop. (6) In our experimental exploration we have found that Active Learning may provide an alternative avenue to tackle the issues raised by heavily class imbalanced datasets.

The structure of the paper is as follows: Section 2 describes the Active Learning framework, the machine learning approaches, and the image feature generation method. Section 3 describes the experimental set-up. Section 4 provides the experimental results. Finally, Section 5 provides our conclusions and some hints for future work.

2. Methods

2.1. Random Forest Classifiers

Random Forest (RF) algorithm is a classifier [11] that encompasses bagging [10] and random decision forests [1, 19], whose performance has been demonstrated in a variety of applications [9, 27]. RF became popular due to its simplicity of training and tuning while offering a competitive performance to other machine learning approaches, such as support vector machines. Consider a RF

4

as a collection of decision tree predictors, built so that they are as much decorrelated as possible, denoted:

$$\{h(\mathbf{x}; \psi_t); t = 1, ..., T\},$$

where $\mathbf{x}$ is a d-dimensional random sample of random vector X, ψ_t are independent identically distributed random vectors modeling the stochastic nature of the tree building process. Each tree $h(\mathbf{x}; \psi_t)$ casts a unit vote in order to decide the class assignment of $\mathbf{x}$. RF captures complex interaction structures in data, and are proposed [11] to be resistant to both over-fitting of data when individual trees are very deep and no pruned, and under-fitting when individual trees are too shallow.

Given a dataset of N samples, a bootstrapped training dataset is used to grow a tree $h(\mathbf{x}; \psi_t)$ on a randomly selected subset of data dimensions $\hat{d}$ such that $\hat{d} \ll d$. Decision tree growing recursively picks the best data split of each node based on these information measure of each dimension. In RF pruning is not required [11]. The RF training process picks randomly the dimension and the dataset bootstrapping according to independent identically distributed random vectors ψ_t. This randomness is the source of RF individual tree diversity ensuring the decorrelation of their outputs.

Classification of a new input $\mathbf{x}$ is done by majority voting over the responses of the trees in the RF $C_u(x)$. The critical parameters of the RF classifier for the experiments reported below are the number of trees in the forest, the dimension of the random subspace, and the maximum tree depth. We report experiments assessing the effect of these parameters in our specific study.

2.2. Ensembles of Extreme Learning Machines

Extreme Learning Machine (ELM) [24, 21] was proposed for fast and efficient training of Single-Hidden Layer Feedforward Neural Networks (SLFN). It is composed of the random generation of the input-to-hidden layers weights followed by least squares estimation of the hidden-to-output layer weights. Random weight generation allows to solve the non-linear training problem as a linear problem, gaining in speed at the cost of some instability that has been shown empirically to be affordable. In general terms, a supervised classifier is a map from input feature space into a target value space, for the SFLN this map has the following form:

$$f_L(\mathbf{x}) = \sum_{i=1}^{L} \boldsymbol{\beta_i} \cdot h_i(\mathbf{x}) = \mathbf{h}(\mathbf{x}) \boldsymbol{\beta}, \tag{1}$$

where $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_L]^T$ is the matrix composed of the weights of the connections hidden-to-output units, and the transformation $\mathbf{h}(\mathbf{x})$ is the ELM nonlinear transformation from input to hidden space. The hidden units of ELM may be any piecewise continuous function that satisfies simple universal approximation conditions [22, 23]. Given a sample of feature vectors and corresponding labels $\{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, the estimation of the labels given the input features can be written in matrix notation as

$$\mathbf{H}\boldsymbol{\beta} = \hat{\mathbf{Y}}, \tag{2}$$

where $\mathbf{H}$ is the matrix of responses of the hidden units to the input features in the sample, and $\hat{\mathbf{Y}}$ is the achieved approximate value of the labels of the samples.

The process of training an SLFN following ELM approach has two stages: (1) feature mapping implemented by random sampling of the hidden units activation in response to the input features in order to build the feature kernel matrix $\mathbf{H}$, and (2) least squares resolution of the following minimization problem:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{L \times m}} \|\mathbf{H}\boldsymbol{\beta} - \mathbf{Y}\|^2 , \tag{3}$$

whose optimal solution is given by $\hat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{L}$, where $\mathbf{H}^{\dagger}$ is the Moore-Penrose generalized inverse of $\mathbf{H}$. Note that the definition of ELM allows naturally multivariate output SFLN.

ELMs can be combined very efficiently in ensembles which can be homogeneous or heterogeneous [5]. Applications on the prediction of readmissions in hospital environments [3], remote sensing image processing [4, 6, 7, 8] have already been reported. The simplest ensemble of ELM is the Voting ELM (V-ELM) [13], consisting of a collection of ELMs each trained on a bootstrap of the training dataset. Every SFLN has the same number of hidden neurons, and the decision on the class is performed by majority voting over the responses of the SFLN in the ensemble. Examples of applications of V-ELM are hyperspectral image classification [?], remote sensing data classification [18]and natural gas reservoir characterization [2], and wastewater quality index modeling [35]. It has been proved that accuracy converges to perfect classification as the size of the ensemble grows under the mild assumption that individual classifiers have performance better than random choice [13]). The V-ELM has spurred some variations in the literature, such as the use of soft-class dependent voting schemes [12] trying to increase model reliability. We apply the basic scheme in our work.

2.3. Active Learning

The motivation of the Active Learning approach [14, 15, 33] is to facilitate the task of data labeling for supervised classification. The two objectives pursued by this approach are to reduce as much as possible the manual effort of data sample labeling, and the selection of those data samples which are more informative towards the building a robust and efficient classifier. Data labeling is a costly process involving manual labor and some interactive data visualization tool that facilitates the labeling process. In some cases, data labeling involves quite complex field work, so minimizing the number of required labeled samples and ensuring that each labeled sample contributes significantly to the quality of the classifier is of paramount importance. Moreover, manual data labeling may introduce erroneous classifications, which further interfere and degrade the training process. Such kind of errors are less likely in small datasets. Additionally, in the case of highly class imbalanced datasets, the guided selection of balanced training datasets may enhance the results as we have found in our experiments reported later.

Let us denote $X = \{\mathbf{x}_i, y_i\}_{i=1}^l$ and $U = \{\mathbf{x}_i\}_{i=l+1}^{l+u} \in \mathbb{R}^d$ the training set containing labeled samples, and the unlabeled samples in the *pool of candidates*, respectively. We have that $u \gg l$, $\mathbf{x}_i \in \mathbb{R}^d$, and $y_i \in \{1, \dots, N\}$. Active Learning is an iterative algorithm. At iteration t, the algorithm selects q candidates from U^t to be added to the current training set X^t, aiming at maximal gain in performance of the classifier trained with the incremented training dataset,

while reducing the classification model uncertainty. An oracle provides the labels $\{y_m\}_{m=1}^q$ to the selected samples $S^t = \{\mathbf{x}_m\}_{m=1}^q \subset U$. The oracle can be a human carrying out an interactive segmentation, or available ground truth if we are carrying out computational experiences, such as we do in this paper. The current training set is increased with the candidates selected from the pool($X^{t+1} = X^t \cup S^t$) , while being removed from the pool of candidates ($U^{t+1} = U^t \backslash S^t$)The process stops when some criterion is met, such as the achieving accuracy over a preset threshold θ_{max}. Algorithm 1 summarizes the Active Learning process.

Algorithm 1 Active learning general algorithm

Imputs
–Initial training set $X^t = \{\mathbf{x}_i, \ y_i\}_{i=1}^l (X \in \mathcal{X}, \ t = 1)$.
–Pool of candidates $U^t = \{\mathbf{x}_i\}_{i=l+1}^{l+u}(U \in \mathcal{X}, \ t = 1)$.
–Number of voxels q to add at each iteration (defining the batch of selected voxels S).

1: **repeat**
2:     Train a classifier with current training set X^t
3:     **for** *each* candidate in U^t **do**
4:         Evaluate a user-defined heuristic
5:     **end for**
6:     Rank the candidates in U^t according to the score of the heuristic
7:     Select the q most interesting voxels $S^t = \{\mathbf{x}_k\}_{k=1}^q$
8:     The system assigns a label to the selected voxels $S^t = \{\mathbf{x}_k, \ y_k\}_{k=1}$
9:     Add the batch to the training set $X^{t+1} = X^t \cup S^t$
10:     Remove the batch from the pool of candidates $U^{t+1} = U^t \backslash S^t$
11:     $t = t + 1$
12: **until** $t > T$

2.3.1. Classification uncertainty in ensemble classifiers

RF and V-ELM classifiers implement committee approaches to the decision of the class corresponding to an unlabeled sample. This distributed cooperative decision can be used for classification uncertainty estimation [33]. Given a committee of k base classifiers, then we have k labels for each candidate unlabeled sample $\mathbf{x}_i \in U$. The data sample class label is provided by the majority voting. Heuristically, we equate the measure of the classification uncertainty of $\mathbf{x}_i$ with the standard deviation $\sigma(\mathbf{x}_i)$ of the class labels generated by the committee. Let us order the pool of candidates. i.e.: $U^* = \{\mathbf{x}_{j_i}\}_{i=l+1}^{l+u}$, according to $\sigma(\mathbf{x}_{j_i}) > \sigma(\mathbf{x}_{j_{i+1}})$. The *standard deviation query-by-bagging* heuristic selection of samples to be added to the train set is stated as the following selection:

$$S^t = \{\mathbf{x}_{j_m}\}_{m=1}^q \tag{4}$$

The standard deviation of ensemble label predictions is a powerful heuristic measure of classification uncertainty. A candidate sample that is classified equally by all ensemble members has a zero prediction standard deviation, hence its inclusion in the training set does not add any information. On the other hand, if a candidate has uniformly distributed responses from the ensemble, its standard

deviation is maximal, hence it contributes maximal information when included in the training dataset.

2.3.2. Active Learning for Image Segmentation

We want to classify image pixels into two classes, the target and the background [27]. Target in our case are the pixels corresponding to the lane marks and other landmarks in the road. In a nutshell, an Active Learning system returns to the user an image whose intensity value corresponds to the degree of uncertainty in the classification of the pixel. Upon this image, the user, in its role as the oracle will pick some of the pixels with greatest intensity labeling them for insertion in the training dataset. Then, a new instance of the classifier is trained [30]. The features of each pixel are the result of the application of a bank of Gabor filters, the pixel intensity and its coordinates. Though the feature vector dimensionality is relatively high, we do not carry out any feature selection procedure because we prefer to leave open the possibility that a certain orientation or scale may be meaningful in future images. We assume that the stream of images may produce images that are quite different from the ones used for training. Hence, the final implementation will allow to restart the Active Learning process when the human operator detects deviations from optimal segmentation. It is an open-end learning process with a human in the loop. For the computational exploration reported in this paper, we do not resort to a human oracle. Instead, we use a hand delineated ground truth of a collection of images. Hence, the labeling process consists in the selection of the pixels with maximal uncertainty, applying random selection to solve ties.

2.3.3. Active Learning and class imbalance

Class imbalance is a widely know issue in machine learning [17] that is gaining increased attention due to its very strong bias effect in classifier training by almost any modeling approach. In the case of digital mapping, the effect of class imbalance has been also recognized [31]. The data dealt with in this paper is highly class imbalanced, hence we must take into account this fact in the use of performance metrics, and in the way we select the samples that are added to the training set. We have carried out comparative experiments ensuring that the added samples are class balanced and not ensuring this class balance. Usually, using subsampled balanced datasets for training does not achieve competitive test results. However, we hypothesized that the optimal selection of the samples followed by the Active Learning may compensate for the subsampling. We carried the comparative over the RF experiments to test this hypothesis.

2.4. Gabor texture features

In order to have a systematic characterization of the surroundings of each pixel we use a bank of Gabor filters. The magnitude of the responses of the Gabor filters are used a the feature vector for classification. In other words, we use the local texture descriptor of the image as features [16, 28] for classification. Formally, a Gabor filter is defined by the product of a sinusoidal wave, a plane wave in 2D, and a Gaussian function. The Gaussian component modulates the scale of the filter, while the wave component acts as a selector of the orientation and spatial frequency of the detected objects. In many implementations, Gaussian scale and wavelength are linked, so only the wavelength is specified. The
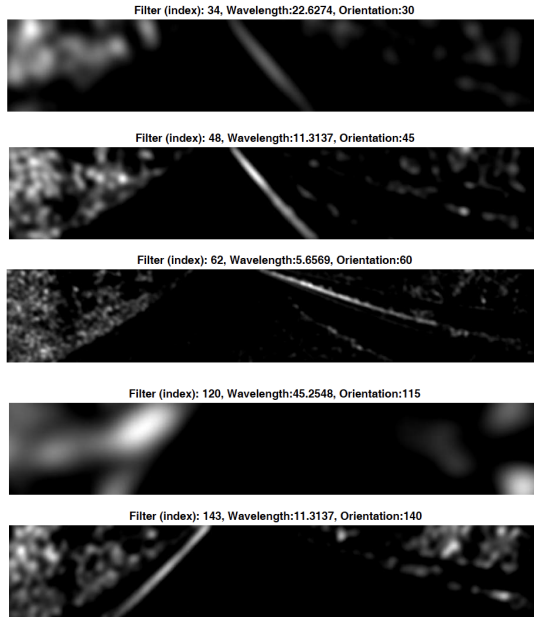
Figure 4: Examples of filtered images with Gabor filters of various orientations and scales.

Gabor filter provides a complex valued response, so its magnitude and phase can be used as features. In this paper we use only the phase. Formally, a the impulse response of single filter is defined as follows:

$$g\left(x,y\right) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right)\exp\left[-\frac{1}{2}\left(\frac{x'^2}{\sigma_x^2} + \frac{y'^2}{\sigma_y^2}\right)\right]\exp\left[2\pi i\left(Ux + Vy\right)\right],$$

where we rotate the Euclidean coordinates by θ such that $x' = x\cos\left(\theta\right) + y\sin\left(\theta\right)$, and $y' = x\sin\left(\theta\right) + y\cos\left(\theta\right)$. Parameters σ_x, σ_y define the spatial support and bandwidth of the filter. The complex exponential factor is 2D sinusoid wave of frequency $F = \sqrt{U^2 + V^2}$ and orientation $\gamma = \tan^{-1}\left(V/U\right)$. Figure 4 shows some examples of the magnitude response of different filters over the same road image. Diverse orientations highlight different road lane marks, and diverse wavelengths produce crisp or blurred images, where details are highlighted or removed.

3. Experimental setup

3.1. Dataset

From the actual collection of landscapes gathered from several trips of the data recording car we have selected 15 images in order to carry out the experimental work. Instead of working on the full panoramas illustrated in Figure 2, we have automatically cropped a central region of the panorama that contains most of the rear view of the ladybug. Figure 5 shows some of the experimental images (left column) and their manually delineated ground truth. The actual imbalance ratio of the data is 1:39, the target minority class accounts 2.45%
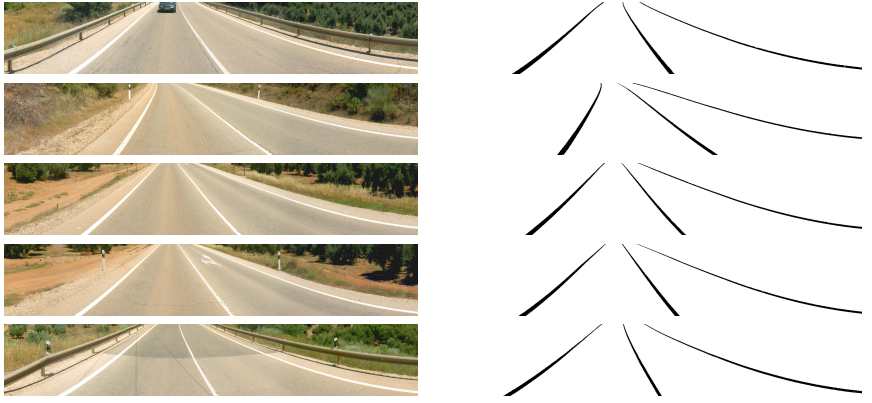
Figure 5: Some of the experimental images (left column) and their corresponding manually delineated ground truth (right column) white=background

of the dataset, the remaining 97.55% corresponds to the background, including the road and the environment.

3.2. Model parameter exploration.

We have carried out an exhaustive exploration of the model parameter settings in order to assess robustness of the approach and to find out which parameter is more influential. As a conclusion we can provide some guidance on the best parameter setting and the most promising model. Regarding RF computational experiments, we have explored the influence of the number of trees (NT), the number of variables taken into account at node split (NVS), and the size of the sample (NS) that is added to the current training dataset. We have also make the comparative between ensuring that the dataset subsample added to the current training dataset is class balanced and not ensuring class balance. This experience is not repeated on the ELM experiments. Regarding ELM computational experiments, we have explored the influence of the ensemble size (NELM), the number of hidden units (NHELM), and the size of the sample (NS) added to the current training dataset. We have used the standard sigmoid activation function. In all reported experiments we have performed a fixed number of iterations (T=10) of the Active Learning algorithm.

3.3. Validation.

We report the sensitivity (SEN), specificity (SP), accuracy (AC), and true positive ratio (TPR) of the pixel-wise classification of the entire images using the classifiers built upon the selected training datasets at the end of the Active Learning process. The most valuable metric is the TPR because of the strong class imbalance of the dataset.

The pool of pixels used for the selection of the training dataset is composed of pixels of all labeled images, so the selection tries to have representatives from all images, in order to avoid overtraining on one image. Hence, at each Active Learning iteration we compute the classification uncertainty over all images. However we do not ensure that the selection is fair, in the sense of picking the same amount of pixels from each image to be added to the training dataset.
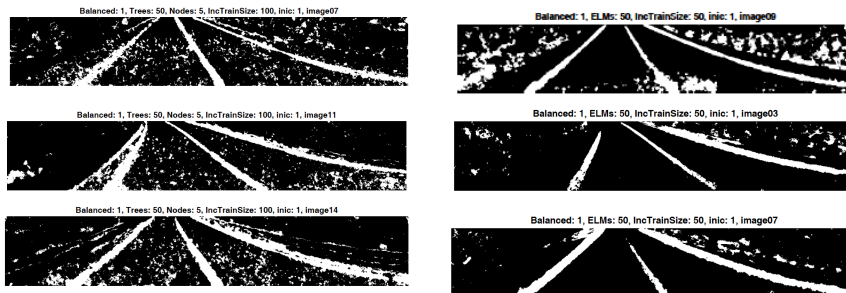
Figure 6: Some visual results of the trained RF (left column) and V-ELM (right column) ensemble classifiers using balanced training sample increments.

Regarding the issue of the separation of training and test data for validation, it is ensured as far as we are reporting the performance measures over the pixels not in the training set. Active Learning is *per se* safe in this regard, because never uses the labeling information of data outside the training dataset [34].

4. Experimental Results

For a qualitative appraisal of the results we show in Figure 6 some images of the detections achieved by RF (left) and V-ELM (right) classifiers trained on the final training datasets of Active Learning processes. Notice that we are not reporting results after post-processing and noise removal of the classification results. Removing noisy detections may be easily achieved applying morphological operators and some standard computer vision tools. We think that these processes correspond to the industrial implementation of the approach, and including them here would somehow obscure the intended contribution of the paper. Regarding the quantitative analysis of the results, we have organized it into several research questions.

Impact of the sample selection strategy. We compare the performance achieved by the RF classifiers when the training datasets are increased by candidate samples with or without ensuring class balancing, as shown in Table 1 and Table 2, respectively. A one-sided t-test comparing the SP and TPR of the balanced versus imbalanced increments confirms that there is a significant increment in performance ($p < 1e-10$ for TPR, $p < 1e-8$ for SP). Wilcoxon rank sum test is even more significant ($p < 1e-11$ for TPR, $p < 1e-9$ for SP). This result is meaningful in the sense that it provides an alternative approach to the resolution of the bias induced by strong class imbalance. In this situation, if we examine the AC performance we are mislead, because ignoring the minority class lead to very high AC values, but very poor TPR values as can be seen in Table 2 where imbalanced training dataset growth lead to high AC and SEN but very poor SP and TPR. In fact the mean TPR in Table 2 is 0.051 while in Table 1 it is 0.552. The result is so definitive that we have not carried out the equivalent experiment on the V-ELM classifiers, which have been trained only with balanced incremental training datasets.

Table 1: Active Learning using RF classifiers ensuring that set of samples added at each iteration is class balanced. Performarce results measured by sensitivity (SEN), specificity (SP), accuracy (AC), True Positive Ratio (TPR) of RF varying the number of trees (NT), the number of variables considered for the split at each node (NVS), the number of samples added in each iteration of the active learning algorithm (NS). The added set of samples is class balanced.

NS	NVS	NT	SEN	SP	AC	TPR
100	5	50	0.858	0.283	0.843	0.898
100	5	100	0.816	0.344	0.804	0.865
100	5	150	0.732	0.483	0.726	0.890
100	5	200	0.995	0.010	0.970	0.090
100	10	50	0.851	0.254	0.836	0.873
100	10	100	0.853	0.304	0.839	0.857
100	10	150	0.850	0.264	0.835	0.812
100	10	200	0.998	0.002	0.973	0.034
100	15	50	0.716	0.488	0.710	0.859
100	15	100	0.759	0.458	0.752	0.890
100	15	150	0.733	0.501	0.727	0.851
100	15	200	0.984	0.034	0.960	0.158
100	20	50	0.864	0.310	0.850	0.812
100	20	100	0.882	0.227	0.865	0.804
100	20	150	0.812	0.383	0.801	0.800
100	20	200	0.891	0.172	0.873	0.363
50	5	50	0.972	0.096	0.950	0.389
50	5	100	0.943	0.156	0.923	0.638
50	5	150	0.999	0.010	0.974	0.043
50	5	200	1.000	0.013	0.975	0.013
50	10	50	0.953	0.128	0.932	0.581
50	10	100	0.924	0.183	0.905	0.622
50	10	150	0.978	0.058	0.954	0.333
50	10	200	0.991	0.011	0.966	0.083
50	15	50	0.874	0.256	0.858	0.599
50	15	100	0.940	0.137	0.919	0.486
50	15	150	0.927	0.181	0.908	0.617
50	15	200	0.910	0.151	0.890	0.274
50	20	50	0.851	0.296	0.837	0.718
50	20	100	0.917	0.168	0.898	0.596
50	20	150	0.945	0.114	0.924	0.533
50	20	200	0.917	0.192	0.899	0.311

Table 2: Active Learning using RF classifiers without ensuring that set of samples added at each iteration is class balanced. Performarce results measured by sensitivity (SEN), specificity (SP), accuracy (AC), True Positive Ratio (TPR) of RF varying the number of trees (NT), the number of variables considered for the split at each node (NVS), the number of samples added in each iteration of the active learning algorithm (NS).

NS	NVS	NT	SEN	SP	AC	TPR
100	5	50	0.997	0.012	0.972	0.078
100	5	100	0.999	0.008	0.973	0.032
100	5	150	0.996	0.014	0.971	0.108
100	5	200	0.999	0.020	0.973	0.057
100	10	50	0.996	0.000	0.97	0.09
100	10	100	0.998	0.000	0.972	0.043
100	10	150	0.992	0.011	0.967	0.172
100	10	200	0.998	0.017	0.973	0.064
100	15	50	0.998	0.005	0.972	0.059
100	15	100	1.000	0.000	0.974	0.015
100	15	150	0.998	0.015	0.973	0.067
100	15	200	0.996	0.001	0.97	0.079
100	20	50	0.998	0.012	0.973	0.043
100	20	100	0.999	0.011	0.974	0.03
100	20	150	0.999	0.000	0.973	0.039
100	20	200	0.999	0.018	0.974	0.043
50	5	50	1.000	0.004	0.974	0.015
50	5	100	1.000	0.004	0.974	0.009
50	5	150	0.995	0.032	0.97	0.132
50	5	200	0.999	0.000	0.973	0.024
50	10	50	0.999	0.005	0.973	0.037
50	10	100	0.995	0.008	0.97	0.127
50	10	150	0.999	0.001	0.973	0.04
50	10	200	0.998	0.002	0.973	0.04
50	15	50	0.997	0.007	0.972	0.055
50	15	100	1.000	0.000	0.974	0.005
50	15	150	0.998	0.004	0.973	0.046
50	15	200	1.000	0.000	0.974	0.013
50	20	50	1.000	0.000	0.974	0.009
50	20	100	1.000	0.003	0.974	0.014
50	20	150	0.999	0.000	0.973	0.034
50	20	200	0.999	0.006	0.973	0.027

Table 3: Active Learning using V-ELM ensemble classifiers ensuring that set of samples added at each iteration is class balanced. Performarce results measured by sensitivity (SEN), specificity (SP), accuracy (AC), True Positive Ratio (TPR) of RF varying the number of ELM components (NELM), the number of hidden units at each ELM (NHELM), the number of samples added in each iteration of the active learning algorithm (NS).

NHELM	NELM	NS	SEN	SP	AC	TPR
25	50	50	0.45	0.9524	0.9401	0.45
25	50	100	0.8059	0.8534	0.8522	0.8059
25	100	50	0.828	0.8217	0.822	0.828
25	100	100	0.5241	0.9289	0.919	0.5241
25	150	50	0.8277	0.7921	0.8277	0.8277
50	50	50	0.6395	0.9021	0.8956	0.6395
50	50	100	0.8442	0.8398	0.84	0.8442
50	100	50	0.8289	0.8151	0.8154	0.8289
50	100	100	0.8469	0.9399	0.9376	0.8469
50	150	50	0.8001	0.8425	0.8415	0.8001
50	150	100	0.8972	0.9227	0.9221	0.8972
50	200	50	0.8167	0.8234	0.8232	0.8167
50	200	100	0.9097	0.8931	0.8935	0.9097
100	50	50	0.8792	0.8025	0.8044	0.8792
100	50	100	0.8713	0.9101	0.9091	0.8713
100	100	50	0.9237	0.845	0.8469	0.9237
100	100	100	0.8984	0.921	0.9204	0.8984
100	150	50	0.8091	0.8866	0.8846	0.8091
100	150	100	0.8345	0.8091	0.8097	0.8345
100	200	50	0.9099	0.8764	0.8771	0.9099
100	200	100	0.8444	0.8186	0.8192	0.8444
150	50	50	0.6467	0.9247	0.918	0.6467
150	50	100	0.8223	0.8974	0.8955	0.8223
150	100	50	0.9038	0.721	0.7255	0.9038
150	100	100	0.9284	0.8785	0.8798	0.9284
150	150	50	0.8418	0.7856	0.7871	0.8418
150	150	100	0.8272	0.8859	0.8844	0.8272

Effect of parameter settings. The size of the sample (NS) that is added to the training dataset at each iteration is a global parameter of the Active Learning process. In the RF experiments, we find moderate but significant effect of the sample size 100 versus 50 in Welch one-sided t-test ($p < 0.004$ for TPR, $p < 0.005$ for SP) and in Wilcoxon rank sum test ($p < 0.02$ for TPR, $p < 0.007$ for SP) on the data of Table 1. In the V-ELM experiments reported in Table 3 we find even lower significance in Welch one-sided t-test ($p < 0.5$ for TPR, $p < 0.05$ for SP) and in Wilcoxon rank sum test ($p < 0.5$ for TPR, $p < 0.09$ for SP). However, we think that the greater sample increment is preferable, for instance, for V-ELM the mean value of TPR for NS=100 is 0.82, while for NS=50 it is 0.77. Similar values for RF are 0.67 and 0.42.

Regarding the effect of the RF parameters, we find that the number of trees in RF has a stronger effect than NVS in an analysis of variance ($p < 1e-5$ for TPR, $p < 0.005$ for SP). It is very interesting to observe a clear overfitting effect, i.e. often there is a decrease in performance when NT goes from 150 to 200 all other parameters unchanged. Looking at the mean values of TPR for NT values 50, 100, 150, 200 we obtain 0.71, 0.72, 0.60, and 0.16, respectively. Hence there is catastrophic overfitting effect when we try to use more than 150 trees.

Regarding the effect of the parameters of V-ELM, namely the number of hidden units and the number of ELMs in the ensemble, their effect is minor. An analysis of variance on the data of Table 3 reports a minor effect of NHELM on TPR and SEN ($p < 0.6$). If we look in detail at the performance for varying NHELM values 25, 50, 100, and 150 we obtain average TPR values 0.68, 0.81, 0.87, and 0.82, respectively. Therefore, we may have a big performance increase going from 25 to 100 hidden units, but afterwards there is a degradation of the classifier performance. Although the number of ELMs in the ensemble seems to be rather irrelevant, it must be noted that the average TPR for NELM values 50, 100, 150, and 200 is 0.74, 0.83, 0.83, and 0.81, respectively. Therefore there is significant improvement increasing the number of ELMs until saturation around 100 ELMs which leads to performance degradation at 200 ELMs but not as dramatic as in the case of RF number of trees.

Effect of the classifier ensemble. When comparing the V-ELM against the RF, we find a highly significant difference in Welch two sample t-test ($p < 1e-4$ for TPR and $p < 1e-15$ for SP) and in Wilcoxon rank sum test ($p < 1e-3$ for TPR, $p < 1e-6$ for SP). The mean vales of the TPR and SP for V-ELM over Table 3 are 0.80 and 0.86, respectively, while for the RF in Table 1 are 0.55 and 0.20, respectively. The best TPR and SP values achieved by V-ELM classifiers are 0.92 and 0.95, while RF achieves 0.89 and 0.50, respectively. Adding up the insensitivity to the parameter settings and the improved results, V-ELM appears to be the most promising approach for the full development of the large scale road images segmentation.

5. Conclusion and future works

In this paper we present an Active Learning approach to deal with the labeling of road landmarks in images obtained by an onboard sensor that includes LiDAR as well as positioning sensors for the purpose of detailed road signaling inventory. The underlying problem is a two class classification problem with

strong class imbalance, and potentially large volume of images taken under very diverse light and atmospheric conditions, as well as road conditions. The proposed solution is an open ended process segmentation with a human in the loop that may start the adaptation to new images at any moment. Due to the cost of image labeling, the adaptation follows an Active Learning approach, where the training set is built incrementally with the most informative image samples. We have explored the performance of two ensemble classifiers, namely the random forest (RF) and the ensemble of extreme learning machines (V-ELM). Our computational experiments have found that the V-ELM improves RF in terms of true positive ratio (TPR), a performance measure more appropriate than Accuracy for strong class imbalanced datasets. Additionally, we have found a novel way to deal with class imbalance through Active Learning selection of optimal balanced training datasets. We think that the approach deserves further exhaustive study, as it has not been previously proposed in the literature. Besides, we have found in our computational exploration that both RF and V-ELM show some overfitting issues as the model complexity grows. This exploratory work may lead to more detailed experimentation in order to propose optimal model parameterizations for the task at hand. Future works would be addressed to the exploitation of the fused image and LiDAR information in order to enhance the road landmark recognition.

References

[1] Amit, Y., Geman, D., 1997. Shape quantization and recognition with randomized trees. Neural computation 9 (7), 1545–1588.

[2] Anifowose, F. A., Labadin, J., Abdulraheem, A., 2015. Ensemble model of non-linear feature selection-based extreme learning machine for improved natural gas reservoir characterization. Journal of Natural Gas Science and Engineering 26 (0), 1561–1572.

[3] Artetxe, A., Ayerdi, B., Graña, M., Rios, S., 2017. Using anticipative hybrid extreme rotation forest to predict emergency service readmission risk. Journal of Computational Science 20, 154 – 161.

[4] Ayerdi, B., Graña Romay, M., May 2016. Hyperspectral image analysis by spectralÐspatial processing and anticipative hybrid extreme rotation forest classification. IEEE Transactions on Geoscience and Remote Sensing 54 (5), 2627–2639.

[5] Ayerdi, B., Graña, M., 2014. Hybrid extreme rotation forest. Neural Networks 52, 33 – 42.

[6] Ayerdi, B., Graña, M., 2016. Hyperspectral image nonlinear unmixing and reconstruction by elm regression ensemble. Neurocomputing 174, 299 – 309.

[7] Ayerdi, B., Maiora, J., d'Anjou, A., na, M. G., 01 2014. Applications of hybrid extreme rotation forests for image segmentation. International Journal of Hybrid Intelligent Systems 11, 13–24.

[8] Ayerdi, B., Marques, I., Graña, M., 2015. Spatially regularized semisupervised ensembles of extreme learning machines for hyperspectral image segmentation. Neurocomputing 149, 373 – 386.

[9] Barandiaran, I., Paloc, C., Graña, M., 2010. Real-time optical markerless tracking for augmented reality applications. Journal of Real-Time Image Processing 5, 129–138.

[10] Breiman, L., 1996. Bagging predictors. Machine learning 24 (2), 123–140.

[11] Breiman, L., 2001. Random forests. Machine learning 45 (1), 5–32.

[12] Cao, J., Kwong, S., Wang, R., Li, X., Li, K., Kong, X., 2015. Class-specific soft voting based multiple extreme learning machines ensemble. Neurocomputing 149, Part A (0), 275 – 284.

[13] Cao, J., Lin, Z., Huang, G.-B., Liu, N., 2012. Voting based extreme learning machine. Information Sciences 185 (1), 66 – 77.

[14] Chyzhyk, D., Ayerdi, B., Maiora, J., 2013. Active learning with bootstrapped dendritic classifier applied to medical image segmentation. Pattern Recognition Letters 34 (14), 1602 – 1608.

[15] Cohn, D., Atlas, L., Ladner, R., 1994. Improving generalization with active learning. Machine Learning 15, 201–221.

[16] Fogel, I., Sagi, D., Jun 1989. Gabor filters as texture discriminator. Biological Cybernetics 61 (2), 103–113.

[17] Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., Bing, G., 2017. Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications 73, 220 – 239.

[18] Han, M., Liu, B., 2015. Ensemble of extreme learning machine for remote sensing image classification. Neurocomputing 149, Part A (0), 65 – 70.

[19] Ho, T., 1998. The random subspace method for constructing decision forests. Pattern Analysis and Machine Intelligence, IEEE Transactions on 20 (8), 832–844.

[20] Hoi, S. C. H., Jin, R., Zhu, J., Lyu, M. R., 2009. Semisupervised SVM batch mode active learning with applications to image retrieval. ACM Transactions on Information Systems 27 (3), 1–29.

[21] Huang, G., Huang, G.-B., Song, S., You, K., ????. Trends in extreme learning machines: A review. Neural Networks, 32 – 48.

[22] Huang, G.-B., Chen, L., 2007. Convex incremental extreme learning machine. Neurocomputing 70 (16–18), 3056 – 3062.

[23] Huang, G.-B., Chen, L., 2008. Enhanced random search based incremental extreme learning machine. Neurocomputing 71 (16–18), 3460 – 3468.

[24] Huang, G.-B., Zhu, Q.-Y., Siew, C.-K., ????. Extreme learning machine: Theory and applications. Neurocomputing (1–3), 489 – 501.

[25] Iglesias, J., Konukoglu, E., Montillo, A., Tu, Z., Criminisi, A., 2011. Combining generative and discriminative models for semantic segmentation of ct scans via active learning. In: Information Processing in Medical Imaging. Springer, pp. 25–36.

[26] Izquierdo, A., Lopez-Guede, J. M., Graña, M., 2019. Road lane landmark extraction: A state-of-the-art review. In: Pérez García, H., Sánchez González, L., Castejón Limas, M., Quintián Pardo, H., Corchado Rodríguez, E. (Eds.), Hybrid Artificial Intelligent Systems. Springer International Publishing, Cham, pp. 625–635.

[27] Maiora, J., Ayerdi, B., Graña, M., 2014. Random forest active learning for aaa thrombus segmentation in computed tomography angiography images. Neurocomputing 126, 71 – 77.

[28] Maldonado, J. O., Graña, M., 2009. Recycled paper visual indexing for quality control. Expert Systems with Applications 36 (5), 8807 – 8815.

[29] Mitra, P., Shankar, B. U., Pal, S. K., 2004. Segmentation of multispectral remote sensing images using active support vector machines. Pattern Recognition Letters 25 (9), 1067 – 1074.

[30] Settles, B., 2010. Active learning literature survey. SciencesNew York 15 (2).

[31] Sharififar, A., Sarmadian, F., Malone, B. P., Minasny, B., 2019. Addressing the issue of digital mapping of soil classes with imbalanced class observations. Geoderma 350, 84 – 92.

[32] Tao, Y., Peng, Z., Jian, B., Xuan, J., Krishnan, A., Sean Zhou, X., 2010. Robust learning-based annotation of medical radiographs. In: Medical Content-Based Retrieval for Clinical Decision Support. Vol. 5853 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg, pp. 77–88.

[33] Tuia, D., Pasolli, E., Emery, W., 2011. Using active learning to adapt remote sensing image classifiers. Remote Sensing of Environment 115 (9), 2232–2242.

[34] Tuia, D., Volpi, M., Copa, L., Kanevski, M., Munoz-Mari, J., june 2011. A survey of active learning algorithms for supervised remote sensing image classification. Selected Topics in Signal Processing, IEEE Journal of 5 (3), 606 –617.

[35] Zhao, L., Yuan, D., Chai, T., Tang, J., 2011. {KPCA} and {ELM} ensemble modeling of wastewater effluent quality indices. Procedia Engineering 15 (0), 5558 – 5562, {CEIS} 2011.