Adapting to Intra-Class Variations using Incremental Retraining with Exploratory Sampling

Young-Woo Seo, Chris Urmson, David Wettergreen, Rahul Sukthankar

CMU-RI-TR-10-36

October 2010

Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

Abstract

Variations in appearance can detrimentally impact the accuracy of object detectors leading to an unacceptably high rate of missed detections. We propose an incremental retraining method that combines a self-training strategy with an uncertainty-based model for active learning. This enables us to augment an existing training set with selectively-labeled instances from a larger pool of examples that exhibit significant intra-class variation while minimizing the user's labeling effort. Experimental results on an aerial imagery task demonstrate that the proposed method significantly improves over conventional passive learning techniques. Although the experiments presented in this paper are in the domain area of visual object recognition, our method is completely general and is applicable to a broad category of problems in machine learning.

Contents

1 Introduction **1**

2 Related Work **2**

3 Incremental Retraining with Exploratory Sampling **4**

4 Experimental Results **8**

5 Conclusions and Future Work **10**

6 Acknowledgements **10**

1 Introduction

Overhead imagery, such as that obtained using satellite and aerial sensors is a valuable source of information in providing structural overviews since it helps users to effectively grasp geographic characteristics of the region. Although large quantities of high-resolution images have become widely available[1], manually annotating them is infeasible. There has been considerable research on automatically analyzing, mining and updating information extracted from overhead-view imagery but a key difficulty is that even relatively simple features, such as parking spots, vary significantly in appearance and are therefore challenging to detect in such images.

Recovery of road structures such as road segments, intersections, and parking lots in overhead-view imagery is an important task for generating route maps for autonomous vehicles and maintaining cartographic databases. Road structures may be recovered by first detecting partial structures and then by optimally interpreting detection results, with respect to geometric and image characteristics of the results. The interpreted results are then used to delineate the geometry of road structures. The underlying idea of such part-based object detection approach is conceptually appealing because structures or complete objects can be modeled by part-objects in a deformable configuration [8]. For example, detecting all of the visible parking spots (or spaces) narrows down to image regions that need a further analysis for recovering parking lot geometry [19]

Learning a part detector requires a large quantity of labeled examples for the detector to learn the model of a target object. Manual labeling has traditionally served as the primary means of preparing a training data for such detection tasks. Since it is expensive and tedious, there is significant interest in reducing the human effort in generating labeled data [15, 21, 20, 3, 16].

We have developed aerial image analysis algorithms that produce a map of drivable regions in a parking lot aerial image. To minimize the effort of manual labeling in this task, we developed a self-labeling method that automatically collects the training data. Our self-labeling method extracts and analyzes low-level image patterns such as lines to automatically obtain a set of parking spot sub-images [19]. Using these self-labeled examples, we learnt the local appearance model of parking spots and showed promising results on detecting parking spots on parking lot overhead imagery.

However, because our self-labeler relies on easily detectable low-level image patterns, it is unable to acquire the training parking spot sub-images from aerial imagery with challenging appearances such as shadows, faded road-markings, and occlusions. This affects the performance of our hypothesis generation method that collects candidate parking spot sub-images based on the spatial layouts and geometric characteristics of self-labeled examples. Because our detector only examines these candidate sub-images, the true parking spots undiscovered by the hypothesis generation would not be recovered, resulting in an unacceptable false negative rate. This is caused primarily by intra-class appear-

[1] e.g., USGS provides foot-resolution maps of the U.S. and its territories.

ance variation that objects belonging to the same category look different based on their physical properties and illumination conditions. This problem is common in many object detection and data-mining tasks where the performance of a detector is limited by the quality of the available training data.

Our goal is to develop a self-learning part-object detector for road structure recovery that automatically collects training examples by analyzing low-level image patterns, learns the model of the target object, and automatically updates the learned model when a detection accuracy drops.

This paper focuses on the second aspect of this goal: how to incrementally update the learned model so as to cover the intra-class variations while minimizing the user's manual labeling effort.

In this paper we present a incremental retraining approach motivated by uncertainty sampling [6] to improve the performance of our detector with a small quantity of manually-labeled data. The initial detector is trained using self-labeled examples and classifies the newly-available labeled data based on the learned model. Classified examples with low confidences are assumed to lie near the decision boundary of a binary classification. Thus, retraining the detector using these examples enables the detector to explore uncertain regions and to better generalize its current model of objects with previously-unseen examples.

2 Related Work

Learning a reliable visual object detector requires a substantial amount of labeled examples. Manual labeling has been employed as the primary means of obtaining high-quality examples of the target object. Because manual labeling is expensive, there has been a great deal of research conducted in developing methods that minimize human involvement in a learning task while ensuring the generalization of a learner.

Active learning is a popular approach for reducing manual labeling. Instead of passively receiving training data, a learner actively chooses examples from a pool of unlabeled instances based on a sampling criterion and requests a human labeler to provide the labels for these selected examples [6, 13]. The sampling criterion is the key component that enables any active learning method to pick more informative examples and a learner to generalize its model with less labeled data.

In the general setting of active learning, a learner estimates the initial model of the target object from the initial labeled examples. The learner assigns unlabeled examples with confidences that explain how certain the learner is about its classification. The closer the location of examples to the current decision boundary the lesser certain their memberships are. Thus examples located near the decision boundary are clearly important for improving the learner's performance. Tong et al. utilize the spatial proximity of unlabeled examples to the decision boundary as a sampling criterion [23]. Gurevich et al. use this criterion to select negative examples for learning a visual object recognizer [11]. Near-miss examples are examples that are slightly different from

positive ones and hence located around the decision boundary. They are very informative to learn a visual detector that tightly focuses on the target object. Wu et al. combine this criterion with two other measures such as diversity and representativeness for selecting images in an image retrieval application [26]. The diversity criterion is closely related to our coverage measure in that favors requesting labels for instances that are less redundant with the labeled data [2].

Self-labeling is an alternative way of minimizing the use of manually labeled data. Most of self-labeling methods implement heuristics derived from domain knowledge to automatically collect labeled instances using an initially trained classifier.

In a multimodal input data setting, one can utilize the most precise data source to label other data sources that are complementary, but unlabeled. For example, because laser range measurements are quite reliable within range finders' operating ranges, they are used to predict the roughness [21] and the traversability [20] of terrain and detect road boundaries [14]. Stavens and Thrun estimate the associations between inertial data and laser readings on the same terrain and use the learned rules to predict possible high shock areas in upcoming terrains [21]. Similarly, Sofman et al. use local range estimates as self-labeled examples to learn relations between the characteristics of local terrain and corresponding regions in aerial images [20]. The learned relations are used to map aerial images to long range estimates of traversability over regions that a robot is exploring. Lieb et al. devised a self-supervised approach to road following that analyzes image characteristics of previously traversed roads and extracts templates for detecting boundaries of upcoming roads [14].

In a unimodal input data setting, one can utilize low-level data patterns for self-labeling. Nair and Clark exploit foreground movements on static background to automatically collect training examples for the task of detecting people from the video of an office corridor scene [15]. Rosenberg et al. introduce a weakly labeled data that loosely confines the possible image regions containing the target objects and are selectively utilized to improve the detector's performance [17]. Our self-labeler extracts lines forming parking lot road-markings and analyzes spatial layouts between them, resulting in a collection of parking spot image patches which can be used as training examples [19]. We additionally use these initial parking spots to guide a random selection of negative examples.

Our approach is closely related to a semi-supervised learning technique called self-training where the learner obtains the initial model from the labeled data, classifies the unlabeled data, and updates the model using augmented training data with the labeled data and the predicted unlabeled data [17]. For example, Rosenberg et al. iteratively retrain a human face detector with the initial labeled data and the weakly-labeled data. The weakly-labeled data is the unlabeled data with high-confidences that are considered as the training data. Their approach is different from ours in that we add the k most uncertain unlabeled examples to increase the training data's coverage on variation of visual appearances.

In the GIS community, there is extensive work on extracting road structures in aerial imagery: estimating geometric characteristics and connectivity of intersections by analyzing rasterized images [5], updating [10, 1, 9] and conflating [4]

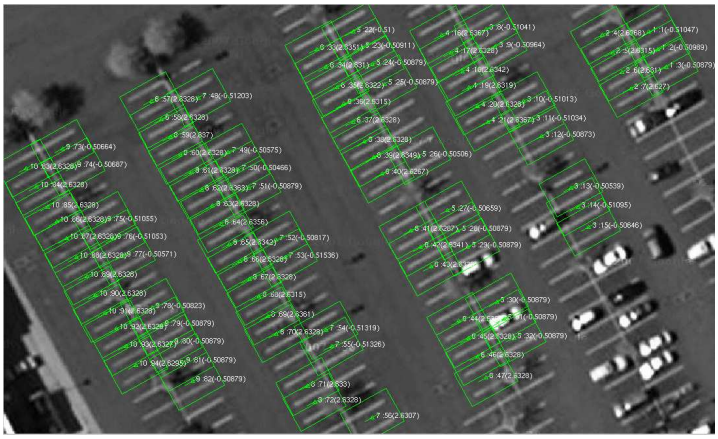


Figure 1: A parking lot image used for self-labeling. Each of the (green) rectangles is a self-labeled training parking spot example.

roadmap databases by analyzing ortho-imagery.

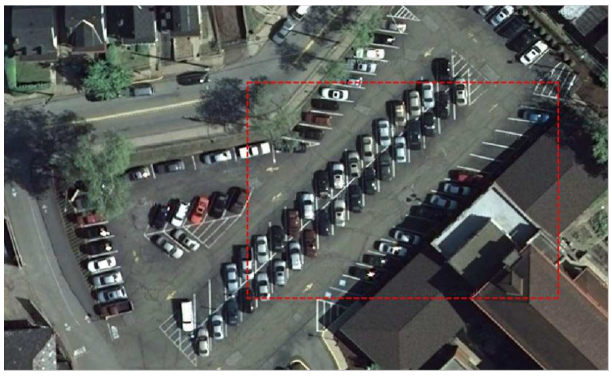
Although there is significant research in extracting road network structures from overhead aerial imagery, to the best of our knowledge, ours is the first work that mines the geometric structures of parking lots from large collections of images, using self-supervised machine learning techniques.

3 Incremental Retraining with Exploratory Sampling

Our task is to learn a binary classifier for parking spots. For the preparation of a training data, we developed a self-labeling method that extracts lines from a parking lot ortho-image[2] and analyzes their spatial layouts to automatically collect parking spot sub-images [19]. These self-collected images are used to train a parking spot detector. Figure 1 shows examples of such self-labeled parking spots. However, because our self-labeler relies on easily-detectable line patterns (the outcome of our self-labeling hinges on the quality of a given gray image), it is unable to obtain the training parking spot sub-images from aerial imagery with challenging appearances. Figure 2 shows the examples of aerial images used in this work where varying illumination conditions and occlusions degrade the performance of our self-labeler. This variation in object appearances breaks our parking spot detection procedure based on self-labeling and make our detector unable to correctly classify parking spot sub-images with challenging appearances, resulting in an unacceptable false negative rate.

[2] An ortho-image is an aerial image where terrain relief and camera tilt are removed through a rectification process.

(a) A parking lot image with faded road-markings.



(b) Parking spots with challenging appearances: occlusions, shadow, and fish-bone shapes.

Figure 2: Examples of the parking lot ortho-imagery with challenging appearances.

To handle this problem, we propose an active-learning based retraining. We assume that a human labeler detects a performance drop and provides our detector with manually labeled parking spot sub-images with challenging appearances. Because manual labeling is expensive, our retraining algorithm aims at minimizing the use of manually labeled data while maintaining a low rate of label noise.

Our method is an extension of a pool-based active learning [13] that enables discriminative supervised machine learning techniques to improve their performance with small amounts of manually-labeled data. Unlike a passive (or batch) learning method, where the learner passively requires a completely-labeled dataset to learn the target function, our method gradually labels and uses the available data through a retraining process. For an individual retraining processes, the current binary classifier assigns its confidences of class

Algorithm 1 Incremental Retraining with Exploratory Sampling.

Input: - $S = \{s_1, ..., s_{|S|}\}$, a set of self-labeled examples; $U = \{u_1, ..., u_{|D|}\}$, a set of unlabeled data $s_i, u_j \in \mathbb{R}^m$; h_0, an initial classifier; h_k, the current classifier; $bs = \{bs_1, ..., bs_K\}$, a list of bucket size; Δ, the size of an increment

Output: h^*, an optimal classifier

Initialization: $h_k \leftarrow h_0$,
for $k = 1$ to K **do**
  $Classify(h_k, U)$, assign classification confidences to the unlabeled data
  $Score(U)$, compute the selection scores of U and $Sort(U)$ in descending order of the scores
  $B_k = Sampling(U, bs_k)$, $|B_k| = bs_k$
  Present examples in B_k to the oracle for labeling
  - $B_k = B_s \cup B_r$, $B_s = \phi$, exploratory sample set, $B_r = B_k$, the remaining data
  while $B_r \neq \phi$ **do**
    $B_s = Sampling(B_k, \delta)$, $B_r = B_k - B_s$
    $h_k^{'} = Train(h_k, D)$, train h_k using $D = S \cup B_s$
    $Classify(h_k^{'}, B_r)$
  end while
end for
Return $h^* \leftarrow h_k$

assignments to each of the sampled data that is compiled by a random sampling. Examples with low confidence values are selected to retrain the current classifier. From the perspective of the current classifier's model, those examples with low confidences are the ones at the decision boundary. Retraining the detector with these examples enables the detector to explore uncertain regions of the decision space.

Algorithm 1 shows the pseudo-code of our method. The initial classifier, h_0, is trained by only using self-labeled examples, S. The list of the buckets defines the number of examples used for sampling in each of the retraining steps. The bucket size increases as the retraining proceeds. The predefined list of buckets ensures a careful and incremental use of the limited labeled data in terms of the ratio of the number of labeled data to the performance change. In other words, this list allows us to observe the correlation between the number of newly added examples and the performance increase if any.

Initially, *Classify*(h_k, U) executes a binary classification where the initial classifier, h_k, assigns the unlabeled data, U, with confidences of its classification decisions based on the current model that is obtained from the self-labeled examples, S. In the $Sampling(U, bs_k)$ step, we sample bs_k instances of unlabeled data based on two different selection criteria: the informativeness, $\delta_{\mathrm{I}}(u_j)$, and the abnormality, $\delta_{\mathrm{C}}(u_j)$, of an unlabeled example, u_j. The informativeness measures how much information an example contains based on its assigned classification confidence. The current classifier assigns an unlabeled example

with a low confidence about its decision when the example is located closely at the decision boundary and hence its class membership is uncertain. Such examples are likely to be informative for the classifier to update its model.

$$\delta_{\mathrm{I}}(u_j) = 1 - conf(u_j),$$

where $conf(u_j) \in [0, 1]$ is a classification confidence assigned by the current classifier, h_k.

Another sampling criterion is the abnormality of an unlabeled data, which measures how far an unlabeled example is located from the centroid of positive examples in an local appearance space. The local appearance space is a high-dimensional space where objects with similar appearances are proximate. We construct this appearance space by first computing the eigenvectors of an affinity matrix among all of the parking spots and then projecting m-dimensional parking spot image onto the k most significant eigenvectors [25]. The affinity matrix, $\mathbf{W}$, is computed by

$$\mathbf{W}_{i,j} = \exp\left\{-\frac{d(\mathbf{g}_i - \mathbf{g}_j)}{2\sigma^2}\right\},$$

where $d(\mathbf{g}_i - \mathbf{g}_j)$ is a function that measures Euclidean distance between two parking spot image patches, $\mathbf{g}_i$ and $\mathbf{g}_j$; and σ is the width of a kernel that controls the range of neighbors. The farther an unlabeled example is located from the centroid of positive examples the more unusual its appearance.

$$\delta_{\mathrm{C}}(u_j) \quad = \quad dist(u_j, \mu_+),$$

where $dist(u_j, \mu_+)$ is the Euclidean distance between an unlabeled data and the centroid of positive examples in the appearance space and μ_+ is the centroid of positive examples in the appearance space.

Figure 3 shows examples of canonical and unusual parking spot sub-images and their projections onto the compressed appearance space. Because of our imperfect self-labeler, the initial training data is only comprised of canonical parking spot sub-images shown in 3(a). Parking spot images with unusual appearances shown in 3(b) are manually collected. When there is only self-labeled canonical parking spot sub-images available, any learning algorithm might result in producing the solution of the parking spot classification that is biased to the centroid of the self-labeled examples. This solution might be the optimal in that it minimizes the error of canonical parking spot classification. However, this solution might not be ideal when the whole appearance space is considered, resulting in a high error rate on unusual parking spots [18]. To remedy this, it is necessary to update the biased model using data with previously-unseen appearance.

When the classification is done, our algorithm computes the score of sampling selection criteria, sorts the data in descending order of the scores, and selects the top $|B_k|$ number of examples for updating the current classification model. After receiving the true labels of samples in the sample set, B_k,

our algorithm exhaustively re-uses these samples during a retraining phase (described in Algorithm 1, steps 9–13). In step 10, the top δ number of samples, B_s, is selected again. A composite of the newly-added unlabeled data and the self-labeled examples, D, is used to retrain the current classifier, h_k. The newly-trained classifier, $h_k^{'}$, assigns the remaining examples in the sample set, B_k, with classification confidences. Due to newly-added examples, the classifier adjusts its decision boundary and its classification confidences are different. Instead of using a sample set, B_k, at once, this incremental use of samples is useful in that it allows the classifier to thoroughly investigate examples when it is uncertain about their class memberships.

4 Experimental Results

In this section, we describe experimental results that evaluate the effectiveness of our method, emphasizing on improving performance with small amounts of labeled data. We downloaded 45 parking lot images from Google Maps[3], each containing a large number of individual parking spaces. A small number (8) of these images with consistent illuminations and clear road markings were provided to the self-labeler. Our self-labeler automatically collected 532 parking spot sub-images, of which 42 contain vehicles and 32 are false positives (not parking spots). Retaining the label noise, we used these as positive examples to train the initial parking spot classifier. The same number of random sub-images are automatically collected and used as negative examples. From the rest of the parking lot images, we manually labeled 365 parking spot sub-images. The appearances of these manually collected images are quite different from those of self-labeled images in terms of the geometric shape (e.g., parking spots for the handicapped), shadows, occlusions, and the quality of road-markings (e.g., new or degraded). We term the self-collected training examples as "canonical" since they exhibit consistency in visual appearance and the manually-collected examples with varied appearance as "unusual". Figure 3(a) and 3(b) show examples of parking spots with canonical and with unusual appearances. Although parking spot images used in this paper are scale- and rotation-free, their appearances vary greatly in terms of geometric shape, intensity, color, and the quality of road-markings.

We represent a parking spot sub-image in raw-intensity values by a feature vector that is comprised of 6 different parts: the geometric shape of a parking spot (i.e., layout of road-markings in a parking spot sub-image), the distribution of Hue-Saturation-Intensity color, the values of Radon transform, the distribution of oriented gradients, the statistics of pixel intensity, and the responses of spatial filters. A combination of extracted lines and the Radon transform is used to describe the geometric shape of road-markings in a parking spot sub-image. Specifically, it is measured by the angles of intersection between lines. The Radon transform is also used to describe the geometric shape of road-markings by projecting the raw-intensity of a parking spot sub-images along a radial line

[3] `http://maps.google.com`

oriented at specific angles (i.e., 0°, 90°, 154°). We employ overcomplete methods for representing the geometric shape because our line extraction method is not 100% reliable, particularly for degraded road markings. The intensity values in a parking spot sub-image are summarized by a set of six statistics: mean, standard deviation, skewness, smoothness, uniformity, and entropy. The histogram of oriented gradients (HOG) [7] is used to summarize intensity values and geometric shapes together by approximating the distribution of oriented gradients in a parking spot sub-image. To capture the textual characteristics, we use Leung-Malik (LM) spatial filter banks that is comprised of 48 different spatial filters: 6 orientations, 6 scales, 6 center-surround, and 6 low-pass Gaussian filters [12].

For the parking spot binary classification, we use two different discriminative classifiers that can produce probabilistic classification, SVM[4] and AdaBoost in two different learning settings: Passive (random sampling) and active (uncertainty sampling) learning.

Our application of AdaBoost is motivated by Viola and Jones' approach for face detection [24] where decision stumps of individual features are linearly combined with their learned confidences for classification. A decision stump in our case is a single-layer decision tree that computes the similarity of one of the six parts in our feature representation between a parking spot image and the centroid of individual classes and assigns the label of the closest centroid to a testing image. For the actual similarity computation, we found that the histogram intersection [22] produces the most reasonable score.

Our experiments aim at verifying the usefulness of our method from two perspectives: First, how the classifier trained only using self-labeled examples improve its performance for the task of classifying the canonical and the unusual data at an incremental use of unusual data and, second, whether our incremental retraining approach provide a better result than a batch approach.

Figure 4 shows the experimental results from our incremental retraining and comparison to batch-learning. The horizontal axis represents the number (the percentage) of the unusual data (e.g., 13 (3%) out of 365 unusual data) and the vertical axis represents the classification accuracy measured by the ratio of the correctly classified testing examples to the total number of the testing examples. Because we fix the number of the training data, the number of the canonical data instances is decreased as unusual data is added. The initial binary classifiers were obtained only using the self-obtained data: 80% of the canonical data are used as the training and 20% of them are used as the testing data. The performance of the initial classifiers is shown as the first element on the graph. As predicted, the initial classifiers perform well on the canonical data but are poor at detecting the unusual instances of parking lots. This is because the large variations in visual appearance in the latter cannot be modeled as intra-class variations given the initial training set.

As unusual data is incrementally added to the training data, the performance

[4]Implemented using libsvm, which is publicly available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`

improves significantly without adversely affecting performance on the canonical set (which remains above 90%). This is important since we do not wish the introduction of automatically-labeled unusual data instances to degrade the classifier accuracy over the base set. When 28% of the training data is composed of unusual data cases (65% of the unusual data is used as the training data), SVM with uncertainty sampling outperforms all of the other approaches, including batch techniques.

At the end of graphs, there are two sets of performance measurements about a batch learning approach where all of the available data is used at once. Table 1 shows the comparison of classification accuracies between our incremental retraining approach and a batch learning approach. This observation is supported by a paired t-test at the significance level 5% with p-value of 0.0129.

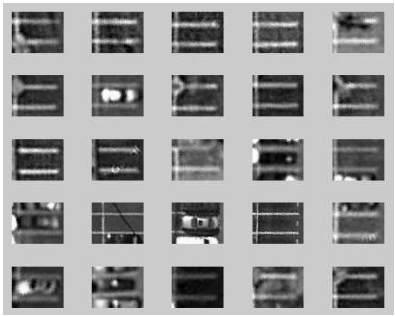
5 Conclusions and Future Work

This paper presents a novel approach to improving the accuracy of a classifier by exploiting unlabeled data. Specifically, given a detector that has been trained using domain knowledge to accuractely recognize only a "canonical" set of instances, we expand its ability to model unusual instances by augmenting its training set using a combination of self-labeling and active learning techniques. This enables the incrementally retrained detector to significantly improve its detection rate by better accounting for intra-class variations that were not present in its original dataset.

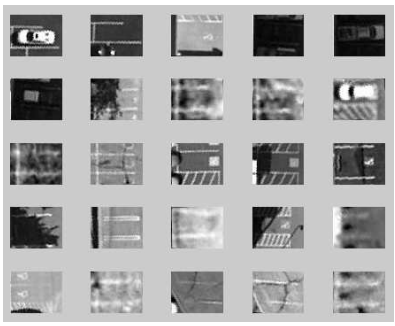
We present experiments that apply the proposed method to the problem of parking spot detection in overhead aerial and satellite images, where intra-class variations in appearance due to illumination changes, lane quality degredations and occlusions can significantly degrade detector performance. We demonstrate that our technique allows the retrained detector to mine "unusual" parking spots without adversely impacting its ability to recognize canonical instances of parking spots. While the experiments shown in this paper are limited to a visual recognition task in aerial images, the proposed technique is very general and can be applied to a broad class of detection and data mining problems.

6 Acknowledgements

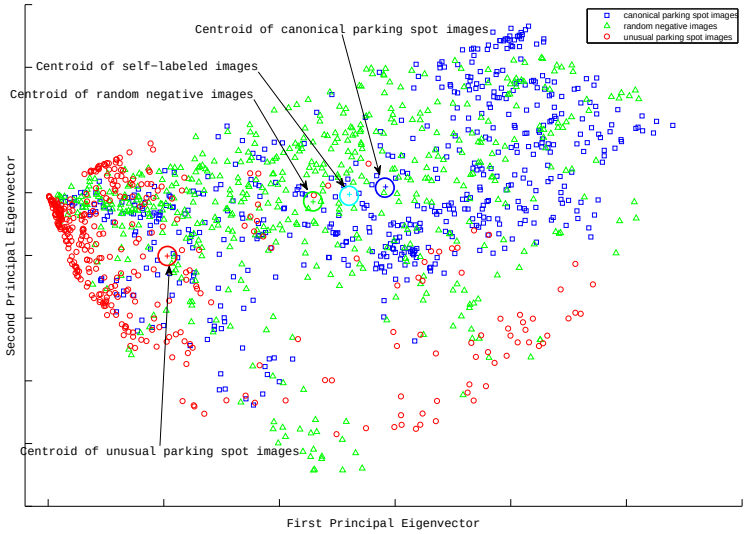
This work is funded by GM-Carnegie Mellon Autonomous Driving Collaborative Research Laboratory (CRL).



(a) Examples of parking spot sub-images with canonical appearances.



(b) Examples of parking spot images with unusual appearances.



(c) A scatter plot of (canonical and unusual) parking spot images and randomly generated negative images' 2-dimensional representation. All (blue) squares represent a canonical parking spots, (green) triangles represent negative images, and (red) circles represent unusual parking spots. There are 532 canonical parking spots, 532 negative examples, and 365 unusual parking spots. The interleaved distributions make the problem difficult.

Figure 3: A two-dimensional projection of multi-dimensional parking spot images with varying appearances. Canonical part spot sub-images are automatically obtained by our self-labeler whereas unusual parking spot sub-images are manually collected.

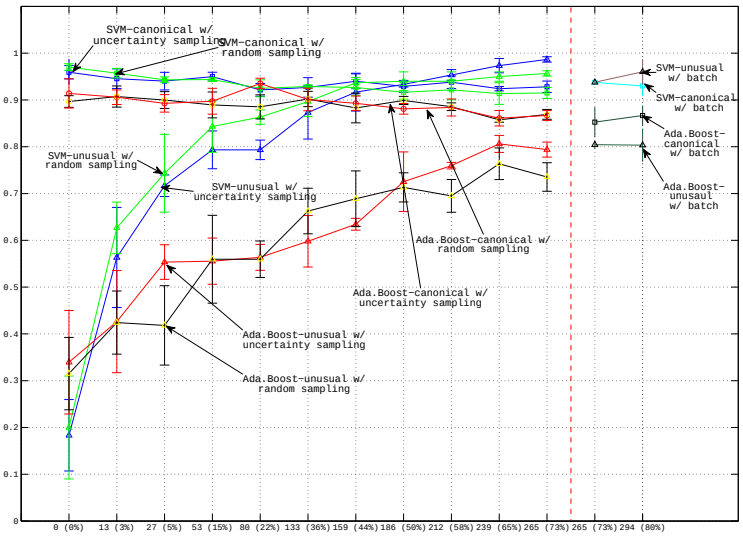


Figure 4: Experimental results of incremental retraining and comparison with batch-learning.

		186 (50%)	212 (58%)	239 (65%)	265 (73%)	265 (73%)	294 (80%)
Uncertainty	Canonical	.9290 (±.0128)	.9380 (±.0133)	.9241 (±.0046)	.9282 (±.0123)	.9376 (±.0046)	.9299 (±.0212)
	Unusual	.9333 (±.0058)	.9533 (±.0115)	.9733 (±.0153)	**.9867** (±.0058)	.9379 (±.0047)	.9602 (±.0260)
Random	Canonical	.9164 (±.0147)	.9216 (±.0212)	.9137 (±.0231)	.9149 (±.0119)	–	–
	Unusual	.9400 (±.0200)	.9400 (±.0200)	.9500 (±.0100)	.9567 (±.0058)	–	–
Uncertainty	Canonical	.8810 (±.0114)	.8838 (±.0182)	.8609 (±.0166)	.8668 (±.0105)	.8526 (±.0307)	.8667 (±.0203)
	Unusual	.7253 (±.0637)	.7596 (±.0070)	.8061 (±.0182)	.7940 (±.0160)	.8044 (±.0078)	.8034 (±.0345)
Random	Canonical	.8989 (±.0087)	.8856 (±.0082)	.8575 (±.0051)	.8692 (±.0106)	–	–
	Unusual	.7131 (±.0311)	.6950 (±.0350)	.7636 (±.0338)	.7353 (±.0305)	–	–

Table 1: Comparison of accuracy between incremental learnings and batch learnings.

References

[1] E. Baltsavias and C. Zhang. Automated updating of road databases from aerial imagery. *International Journal of Applied Earth Observation and Geoinformation*, 6:199–213, 2005.

[2] K. Brinker. Incorporating diversity in active learning with support vector machines. In *Proceedings of the International Conference on Machine Learning*, pages 59–66, 2003.

[3] O. Chapelle, B. Scholkopf, and A. Zien. *Semi-Supervised Learning*. MIT Press, 2006.

[4] C.-C. Chen, C. A. Knoblock, and C. Shahabi. Automatically conflating road vector data with orthoimagery. *GeoInformation*, (10):495–530, 2006.

[5] Y.-Y. Chiang and C. A. Knoblock. Automatic extraction of road intersection position, connectivity and orientations from raster maps. In *Proceedings of the ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2008.

[6] D. Cohn, L. Atlas, and R. Ladner. Improving generalization with active learning. *Machine Learning*, (15):201–221, 1994.

[7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings of Computer Vision and Pattern Recognition*, pages 886–893, 2005.

[8] P. F. Felzenszwalb and D. P. Huttenlocher. Pictorial structures for object recognition. *International Journal of Computer Vision*, 61(1):55–79, 2005.

[9] M. flavie Auclair Fortier, D. Ziou, C. Armenakis, and S. Wang. Automated updating of road information from aerial images. In *Proceedings of American Society Photogrammetry and Remote Sensing*, pages 16–23, 2000.

[10] D. Geman and B. Jedynak. An active testing model for tracking roads in satellite images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(1), 1996.

[11] N. Gurevich, S. Markovitch, and E. Rivlin. Active learning with near misses. In *Proceedings of the National Conference on Artificial Intelligence*, pages 362–367, 2006.

[12] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44, 2001.

[13] D. D. Lewis and J. Catlett. Heterogeneous uncertainty sampling for supervised learning. In *Proceedings of the International Conference on Machine Learning*, pages 148–156, 1994.

[14] D. Lieb, A. Lookingbill, and S. Thrun. Adaptive road following using self-supervised learning and reverse optical flow. In *Proceedings of Robotics Science and Systems*, 2005.

[15] V. Nair and J. J. Clark. An unsupervised, online learning framework for moving object detection. In *Proceedings of Computer Vision and Pattern Recognition*, pages 317–324, 2004.

[16] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng. Self-taught learning: Transfer learning from unlabeled data. In *Proceedings of International Conference on Machine Learning*, pages 759–766, 2007.

[17] C. Rosenberg, M. Hebert, and H. Schneiderman. Semi-supervised self-training of object detection models. In *Proceedings of the 7th IEEE Workshops on Application of Computer Vision*, pages 29–36, 2005.

[18] Y.-W. Seo, N. Ratliff, and C. Urmson. Self-supervised aerial image analysis for extracting parking lot structure. In *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence*, pages 1837–1842, 2009.

[19] Y.-W. Seo, C. Urmson, D. Wettergreen, and J.-W. Lee. Augmenting cartographic resources for autonomous driving. In *Proceedings of ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 13–22, 2009.

[20] B. Sofman, E. Lin, J. A. Bagnell, N. Vandapel, and A. Stentz. Improving robot navigation through self-supervised online learing. In *Proceedings of Robotics Science and Systems*, 2006.

[21] D. Stavens and S. Thrun. A self-supervised terrain roughness estimator for off-road autonomous driving. In *Proceedings of Conference in Uncertainty in Artificial Intelligence*, 2006.

[22] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[23] S. Tong and E. Chang. Support vector machine active learning for image retrieval. In *Proceedings of the ACM International Conference on Multimedia*, pages 107–118, 2001.

[24] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[25] Y. Weiss. Segmentation using eigenvectors: A unifying view. In *Proceedings of International Conference on Computer Vision*, pages 975–982, 1999.

[26] Y. Wu, I. Kozintsev, J.-Y. Bouguet, and C. Dulong. Sampling strategies for active learning in personal photo retrieval. In *Proceedings of IEEE International Conference on Multimedia and Expo*, pages 529–532, 2006.