2011 IEEE Intelligent Vehicles Symposium (IV)
Baden-Baden, Germany, June 5-9, 2011

Large Scale Sign Detection using HOG Feature Variants

Gary Overett and Lars Petersson
NICTA, Locked Bag 8001, Canberra, Australia
{gary.overett,lars.petersson}@nicta.com.au

***Abstract*— In this paper we present two variant formulations of the well-known Histogram of Oriented Gradients (HOG) features and provide a comparison of these features on a large scale sign detection problem. The aim of this research is to find features capable of driving further improvements atop a preexisting detection framework used commercially to detect traffic signs on the scale of entire national road networks (1000's of kilometres of video). We assume the computationally efficient framework of a cascade of boosted weak classifiers. Rather than comparing features on the general problem of detection we compare their merits in the final stages of a cascaded detection problem where a feature's ability to reduce error is valued more highly than computational efficiency.**

Results show the benefit of the two new features on a New Zealand speed sign detection problem. We also note the importance of using non-sign training and validation instances taken from the same video data that contains the training and validation positives. This is attributed to the potential for the more powerful HOG features to overfit on specific local patterns which may be present in alternative video data.

I. INTRODUCTION

Automatic traffic sign detection is an important problem for driver assistance systems and automatic mapping applications. In this paper we aim to improve upon a preexisting sign detection framework used commercially in large scale[1] mapping applications. A key challenge in this work is producing classifiers which are able to scan large video datasets quickly, while achieving *high detection rates*[2] ($\approx 99\%$) and *minimal false positive rates* ($\approx 10^{-9}$) despite the huge volume of input image data and the relative sparsity of traffic signs within a typical road scene.

The authors of this paper have previously promoted [1], [2] a viewpoint that both *error rates* and *time to decision* determine the value of a given detection solution. This research has contributed to the development of a highly robust traffic sign detection platform used commercially on 1000's of kilometres of on road video. This includes the histogram feature [1] and the LiteHOG+ feature [2] which are both extremely computationally efficient, even when compared to other efficient features such as Haar features [3]. While the Histogram and LiteHOG+ features can achieve excellent detection rates in combination with extremely low false positive rates ($< 10^{-9}$) for many sign detection problems, the *large volume* of data we are processing in tandem with the *extreme sparsity* of traffic signs within a road scene mean that we still have need of more powerful features able to *further reduce* false positive rates when our currently available features are no longer able to.

If we accept that more powerful and complex features will generally be more computationally intensive [2] we must also accept that they may not be suitable for the *entire* evaluation chain of a classifier. However, if we maintain our use of fast-to-evaluate features at the head of a detection cascade and use more powerful and more complex features at the tail end of a cascade we are able to maintain fast average evaluation times. By using these powerful features only to resolve the class/non-class status of a *tiny minority* (see Figure 1) of the input imagery we are able to benefit from their discriminative power at negligible computational time cost.

Fig. 1. New Zealand speed signs (top three rows) and challenging speed sign false positives (bottom three rows). The speed sign false positives shown here represent a tiny minority (5×10^{-10} of windows scanned) of non speed signs remaining after a four stage classifier. Logos, wheels and other signs make up a large portion of difficult false positives. These challenging instances require more powerful features to further reduce the error rate of the classifier.

This paper is organised as follows: Initially we present some background information covering prior work. This is followed by an explanation of our baseline sign detection framework. Next we outline two previously unpublished Histogram of Oriented Gradient based features which are used to extend the baseline system to further reduce the detector false positive rate. Finally we include an analysis of the results using speed sign training and validation tasks.

[1]Results shown in this paper consider a video dataset of 15 million 1024×768 image frames taken at 10 meter intervals from a vehicle mounted camera system deployed in New Zealand.

[2]All false positive rates indicated in this paper are calculated per-classifier-inspected-window. A typical single frame of video may be inspected more than a million times as the detector is run over multiple scales and locations in the frame.

978-1-4577-0891-6/11/$26.00 ©2011 IEEE

II. Background

Given the importance of sign detection in several applications there is a significant amount of previous literature dedicated to the subject [4], [5], [6], [7], [8], [9]. Unfortunately, very little of this research deals with high volumes of video data exposed to real world problems.

Many previous sign detectors have employed the use of color sensitive features to take advantage of the strong color cues used in traffic signs [4], [5], [8]. Broggi et al. [8] present a real time road sign detector using color segmentation, shape recognition and neural network learning. Lafuente-Arroyo et al. [4] employ color and shape information within a Support Vector Machine (SVM) based classifier. Bahlmann et al. [5] apply a standard Haar feature based approach to independent color channel information. Paulo and Correia [6] have used red and blue color information as an initial cue for a sign detection system. Signs are then further classified using shape information into several broad subcategories such as 'danger' and 'information'. While the approach and features adopted in our research have been applied to color information [10] we will limit the scope of our experimental results in this paper to those obtained using grayscale imagery in order to maintain a simpler platform for comparisons.

Alefs et al. [7] present a road sign detection framework using edge orientation histograms as the main feature. More recently, Timofte et al. [9] detailed a full sign detection algorithm employing Haar-like features which automatically acquires a 3D localisation (geo-location) of the detected signs. While this method does commit to some sign specific shape based techniques it is shown to produce excellent results across a range of sign types and circumstances.

A major distinguishing feature of this research from prior work by others is the use of large scale training and validation real world video data (see Figure 1). This enables us to perform a direct performance comparison of alternative detectors at the scale in which the system will be deployed, on video of entire national road networks. Training and validation is performed using a 15 million 1024×768 image frames collected at a geospatial frequency of 10 meters across New Zealand. To our knowledge this is the largest validation task ever presented for a sign detection problem.

III. A Baseline Detection System

In this section we introduce our baseline experimental detector against which we can test our two prospective HOG features. Since the construction and design of this preexisting detector differs somewhat to that found in prior literature its implementation warrants both description and justification.

The aim of our detector creation process is ideally to yield a three to five stage classifier. The final classifier should aim to achieve a 99% detection rate with a false positive rate below 10^{-9}. Some applications require false positive rates as low as 10^{-11} to 10^{-12}.

The following elements are used in the creation of our baseline detector:

A *short* cascade structure: While several other detectors found in literature have used a cascade detector structure [3], [11], [12], [13] all of these have tended to consist of a large number of cascade stages (more than 20 in [12] and [13]). In contrast, we employ a short cascade with just five stages. Since each stage will likely reduce the detection rate, many stages can become a liability to the maintenance of a high detection rate. Even if very ambitious per stage detection rates are specified these must be set against a validation population which may not be a reliable indicator of the actual detection rate which will be a achieved.

The *LogitBoost* [14] Learning Algorithm: The time-constrained nature of our final application means that we require a learning algorithm providing efficient classifiers, boosting meets this criteria. Of the large number of boosting algorithms available [14], [15], [16] we have selected LogitBoost [14] as our chosen approach. The motivation for this is that we have found it to be the most consistently superior on a sign detection task when compared to others we have considered (Including AdaBoost, RealBoost [15] and GentleBoost [14]).

LiteHOG+ [2] as the Default Feature: Again the time-constrained nature of our application means that we cannot move away from an extremely fast feature such as LiteHOG+ for the majority of the cascade evaluation work.

Speed Signs as the Object Class: This paper makes no specific sign type assumptions although we have chosen to use the New Zealand speed sign as a test case. Of the many traffic sign types we consider to be of interest, speed signs are among the most abundant within a typical road scene. By dividing our New Zealand video according to the natural geographic division between the country's north and south islands we get a training population of 57377 speed signs from the north island with a validation population of 21500 signs from the south island. By merging all speed signs into a single sign type we create a sign class which is relatively challenging, since a general speed sign detector must capture the properties of a variety of face values in a single detector.

IV. Previous HOG Implementations

In the literature, the name "Histograms of Oriented Gradients" or HOG has been used to refer to a number of unique but related features. Differences vary in terms of the shape and size of the local region over which a histogram may be calculated, the range of orientations allocated to a single histogram bin, the normalisations (if any) applied to the gradients, and the manner in which the descriptor responses are combined by a machine learning method or other approach.

The power of HOG features came to prominence via the work of Dalal and Triggs [17] who demonstrated its superiority above a number of features in a basic pedestrian detection problem. This original work provided a plethora of closely related HOG variants which it compared on its own INRIA pedestrian dataset. The authors follow the approach of varying a number of parameters in order to find some degree of "optimal" HOG feature. Notably, they compare

both rectangular and circular "cells" over which to sum gradients, and compare the merits of increasing the number of orientation bins. At the learning level they apply Linear SVM learning to construct the final pedestrian detector.

Despite its successes on a challenging pedestrian detection problem the evaluation of the Dalal and Triggs SVM classifier is quite slow. In response Zhu et al. [11] created a much faster HOG based pedestrian detector using an AdaBoost trained classifier. The resulting HOG classifier yielded a greater than 60 times speedup with comparable detection performance. Two key differences in the work of Zhu et al. as compared to the original Dalal and Triggs implementation is the use of HOG features at a wide variety of scales and the per-histogram linearisation using individually trained SVM classifiers for each feature.

Our previously published feature, LiteHOG+ [2], is an even more computationally efficient HOG feature. Limiting the cell size over which gradients are summed to a 4×4 image region, greatly reduces the computational complexity of the feature.

The two HOG features presented in this paper are most closely aligned to those presented in [11] in that histograms are taken over multiple shapes and scales and are combined using a boosting approach. Primarily what makes our two features distinct from those of Zhu et al. is the manner in which we create a linear class/non-class weak hypothesis from the histogram response.

V. Formulating New HOG Feature Variants

As an ensemble method [18], boosting acts on a population of weak hypotheses or models rather than the feature responses themselves. Where Zhu et al. employ the use of SVM to create the weak hypotheses we chose to use the Smoothed Response Binning (SRB) learner as used in [2], [19], [20]. An advantage of the SRB learner is its ability to learn multi modal distributions with high accuracy and a low degree of overfitting when training populations are small. However, no direct comparison has been made between SRB learners and SVM applied as a weak learner.

A. Single Bin HOG Feature (S-HOG)

Rather than applying any form of SVM linearisation, single bin HOG features are constructed by taking the individual histogram bins separately. By pairing the one dimensional S-HOG feature responses with an SRB learner we create the weak hypotheses required by boosting.

Therefore the S-HOG feature response is calculated by simply taking the sum of gradient magnitudes (aligned to a single orientation bin) within a rectangular area to be the feature response. Equation 1 shows the formula for calculating the feature response $f(x)$ of a single S-HOG feature.

$$f(O, R, x) = \sum_{\forall (i,j) \in R} G_O(i, j) \tag{1}$$

where $O \in \{0, 1, \ldots, 7\}$ defines the orientation according to Figure 2, R defines the set of points within a given rectangular region, and $G_O()$ provides the L1 gradient magnitude at coordinates (i, j) in the input image x.

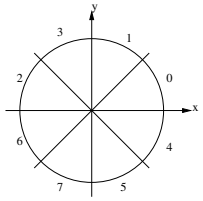

Fig. 2. The orientation space is divided into 8 bins. Each pixel in the gray image is assigned to one of these orientations.

In similar fashion to Zhu et al. we employ the use of the integral image representation [3], [21] to optimise the summing of gradient magnitudes over various rectangular shapes R. Figure 3 shows the basic flow of computation from the input image to the creation of the integral images.

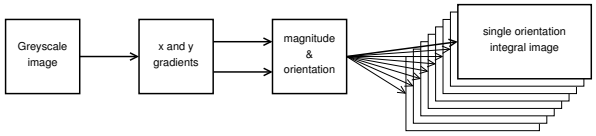

Fig. 3. Computing the Histogram Integral Images. Given a grayscale image, the horizontal (x) and vertical (y) gradients of an image are computed at every pixel using a $[-1, 0, 1]$ kernel. This is then used to compute an L1 magnitude and discretised orientation for each pixel. The magnitude values are subsequently separated according to 8 orientations (see Figure 2) which are used to create 8 separate integral images.

Given a positive and negative class population of scalar feature responses $f()$ we then train a weak hypothesis using the SRB learner. The trained S-HOG features can then be used in any ensemble learning algorithm. The performance of S-HOG features is given in Section VI.

B. FDA Linearised HOG Feature (FDA-HOG)

The second HOG feature we will deal with in this paper is the Fishers Discriminant Analysis (FDA) linearised HOG feature (FDA-HOG). The idea here is to replace the SVM learner used by Zhu et al. [11] with a linearisation obtained using Fishers Discriminant Analysis [22]. This scalar response can then be combined with an SRB learner as in the S-HOG feature. The motivation for this is simple, FDA can supply a linearisation in less time than it would take to train a local SVM classifier for each prospective feature. Furthermore, the evaluation of the final feature response is simpler than evaluating an SVM classifier at run-time.

As with the S-HOG feature we use 8 orientation bins as shown in Figure 2 and use L1 gradient magnitudes. In order to maximise the potential of the featurespace we allow masking of the 8 dimensional histograms to allow FDA to work on subsets of N histogram bins. We find the projection w using the canonical variate of FDA [22] via Equation 2.

$$w = S_w^{-1}(m_1 - m_2) \tag{2}$$

where w is the N dimensional projection matrix, S_w is the within class scatter matrix and m_1, m_2 are the means of the positive and negative classes respectively.

This gives us the feature response:

$$f(\bar{O}, R, x) = w \cdot \Phi(\bar{O}, R, x) \tag{3}$$

where $\bar{O}$ is a vector containing the selected orientations from $\{0, 1, \ldots, 7\}$, R defines the set of points within a given rectangular region, and $\Phi()$ supplies the histogram vector for the gradient magnitudes within rectangle R in image x for the required orientations $\bar{O}$.

As with the S-HOG feature the weak hypothesis is calculated using the SRB learners trained with a population of positive and negative training data.

VI. Experimental Evaluation

In accordance with our aim of finding more powerful features to improve our classifier performance at the tail end of the detection problem we trained a baseline five stage LiteHOG+ classifier using LogitBoost (see Section III). Tables I and II provide some detailed information about this preexisting cascade.

TABLE I
BASELINE CLASSIFIER INFORMATION

Sign Type	New Zealand Speed Signs
# Training Positives	57377
# Validation Positives	21500
# Training Negatives	160K
# Validation Negatives	40K
Negative Source	NICTA Road Scene DataBase

With regard to the training and validation positives shown in Table I we note that not all positive instances will pass through to the later stages of the cascade since they will be lost in the earlier stages in accordance with the false negative rates shown in Table II. Conversely, the number of training *negatives* supplied in each training iteration is held constant by bootstrapping negative examples from the NICTA Road Scene DataBase (NICTA RSDB). This database includes a collection of road scene videos collected in a diverse range of settings around the world, including numerous locations in Australia, the United States, Europe and China. The total amount of data which can be scanned for negatives includes over 10 Million video frames, primarily with a resolution of 960×540. Bootstrapping from this database proceeds for each stage with a calibration step aimed at calculating the sparsest sampling pattern which will yield sufficient samples to make up the negative training set. That is, for the initial stages where false positive rates are still significant, the bootstrapping process will use a sparse frame sampling method with a larger step size and scaling factor[3] in a search for negative road scene instances. Since this data is collected from a road scene that may contain significant true sign instances we must also perform a manual 'cleaning' step to ensure that the negative training and test images are free from actual speed signs.

[3]According to the usual sliding windows and scale-space pyramid approach seen in other literature [3].

TABLE II
PER-STAGE BASELINE CLASSIFIER INFORMATION

Stage #	1	2	3	4	5
# Features	35	200	400	600	1000
# Feature Pool	4300	2048	2048	2048	2048
Per-Stage False Neg.	0.02%	0.58%	0.24%	0.24%	0.24%
Accum. Hit Rate	99.98%	99.40%	99.16%	98.92%	98.68%
Per-Stage False Pos.	0.1%	0.1%	1%	5%	20%
Accum. False Pos	10^{-3}	10^{-6}	10^{-8}	$5{\times}10^{-10}$	10^{-10}

Table II shows the per stage training parameters. The number of features in each stage is gradually increased with just 35 features in the initial stage. Essentially this is in accordance with a tried and tested formula used in many of our other classifiers. The feature pool in Table II refers to the number of LiteHOG+ features which are made available for selection in each training iteration during boosting. This feature pool is replaced with a random subset of the total featurespace at each iteration, increasing the total number of features available to the classifier. Experimentation shows that no improvements are made to classifier performance if the feature pool is increased.

The accumulated false positive rate shown in Table II is calculated using the NICTA Road Scene DataBase (NICTA RSDB). This statistic is known to vary significantly between videos from different scenes.

In our first pair of experiments we replace the fifth stage LiteHOG+ classifier with either an S-HOG or FDA-HOG classifier, maintaining all other experimental parameters as shown in Tables I and II. In order to compare the fifth stage LiteHOG+ feature with the S-HOG and FDA-HOG classifier we produce ROC (Receiver Operating Characteristic) curves for each of the three classifiers. Rather than producing an ROC curve for the entire range of false positive rates and hit rates we limit our calculation to those values achievable given the preexisting four stages. This 'locked in' false negative rate is indicated by the gray region of the ROC plot shown in Figure 4.

A. Native vs. Introduced Negatives

Statistical machine learning algorithms tend to learn patterns in training data which can be considered to be of two different categories, *general patterns* and *specific patterns*. General patterns are those that are represented in the training data which can be said to describe an actual property in that class of data in the real world. Specific patterns are those patterns which are present in the training data but may not occur to the same degree for a given target class in the real world.

The surprising observations noted in Figure 4 suggested that overfitting on specific patterns may have been an issue for the S-HOG and FDA-HOG features. Furthermore, a close examination of the bootstrapped negatives reveals that various video datasets within the NICTA Road Scene DataBase

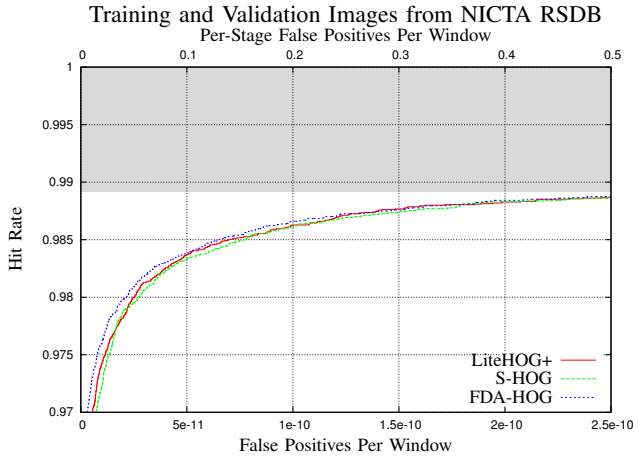

Fig. 4. Fifth stage ROC performance using training and validation images obtained using the NICTA Road Scene DataBase. Surprisingly the performance of each of the three features is almost equivalent. Upon closer inspection of the classifiers we found that the S-HOG and FDA-HOG features achieved much higher detection rates than LiteHOG+ on **training** data. This suggested that the S-HOG and FDA-HOG features were overfitting patterns in the training data more severely. See Section VI-A. The gray shaded area at the top of the plot indicates the accumulated false negative rate incurred by the previous four stages (see Table II).

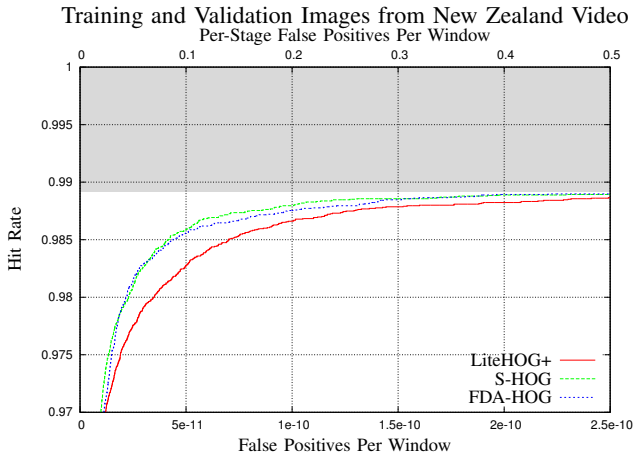

Fig. 5. Fifth stage ROC performance using training and validation images obtained using the New Zealand Video. Results indicate that the S-HOG and FDA-HOG features are indeed able to reduce *additional* false negatives significantly. For the S-HOG feature we observe a 50% reduction in the per-stage false negative rate for a given per-stage false positive rate of 20%. The gray shaded area at the top of the plot indicates the accumulated false negative rate incurred by the previous four stages (see Table II).

tend to exhibit differing volumes of false positives belonging to locally specific negative instances. For example, false positives that may be common in an urban city in Asia may be significantly different to those found in rural Australia. Thus we suspected that the S-HOG and FDA-HOG features in question may simply be exhibiting a greater degree of overfitting to the specific patterns in the *introduced* video data from the NICTA RSDB than LiteHOG+[4].

A solution to this problem is to use *native* negatives from our New Zealand video rather than *introduced* images sourced from other video. That is, training negatives should be taken from the *same* video data as the corresponding positive training instances. This ensures that there is no undue divergence between positive and negative class properties due to effects such as camera noise and country specific patterns (logos, signage, clutter, weather, etc.). Figure 1 shows some false positives which are typical in New Zealand.

In light of this we repeat the experimentation shown in Figure 4 using native negatives sourced from within the *corresponding* New Zealand north (training) and south (validation) island videos. Results are shown in Figure 5.

While both of the new HOG features perform well given native training and validation negatives we note that the simpler S-HOG feature dominates. This yields the rather surprising result that combining HOG dimensions into individual weak classifiers is not particularly more discriminative than allowing boosting to provide a linear combination itself. It would be very interesting to determine if this remains the case when comparing with the HOG features of Zhu et al. [11] who combine HOG dimensions using SVM. In context detections of the cascade are shown in Figure 6.

The use of native negative data is not yet widespread in the object detection community. For example, the pedestrian dataset of Dalal and Triggs [17] uses pedestrians from entirely different source images to those used to obtain the non-pedestrian instances. A close examination of the images shows numerous textural, shading and color differences between the pedestrian and non-pedestrian images. It may therefore be the case that several features which have been promoted to prominence via comparisons on this dataset will not perform as well given a detection task involving native negative data. We note that comparison tests using the MIT pedestrian dataset [23] consisting of native negative data yields substantially different results for several popular feature types with benchmark 'winning' performance on the INRIA dataset of [17].

B. Classifier Evaluation Speed

Timing experiments have shown that for very low false positive rates the speed of the cascade is almost entirely dependant on the computational complexity of the first stage. We find that the computational complexity of the *fifth* stage is essentially totally irrelevant in terms of the average evaluation time of a cascade. Furthermore, our implementation is such that precomputed datatypes, like the histogram image[2] and integral image, need only to be computed on those regions of the image where hits are actually found in the later stages of the cascade. All cascade classifiers shown in Figures 4 and 5 evaluate at approximately 23 frames per second using an Intel Core i7 3.33GHz processor for an image resolution of 768×580 with 420000 windows per image. If we employ our GPU implementation to further improve classifier evaluation time, speeds of up to 245 frames per second can be achieved using a single GPU on an NVIDIA GeForce GTX 295 card.

[4]This can likely be attributed to the fact that the LiteHOG+ feature is not gradient magnitude sensitive but rather uses a binary threshold representation of gradient strength [2].

Fig. 6. Example detections and false postives in context.

VII. CONCLUSIONS

In this paper we have presented two new histograms of oriented gradients (HOG) features, the S-HOG and FDA-HOG features. Both S-HOG and FDA-HOG show a significant discriminative power on a difficult tail end speed sign detection problem extending a preexisting cascade detector. With more powerful features comes a greater potential to overfit on specific local patterns in training data. Thus we make use of native negative image data rather than introduced non-class image data taken from an alternative video source. Given a large scale detection problem the S-HOG feature was able to reduce further increases in false negative error by 50%. Additionally we find that the simpler S-HOG feature performs well against its more complex FDA-HOG counterpart.

The resulting five stage speed sign detector achieves a detection rate of 98.8% in conjunction with a very low false positive rate of 10^{-10}. The evaluation time of the final classifier is 23 frames per second on 768×580 resolution video when using a CPU implementation and 245 frames per second with a GPU implementation.

VIII. ACKNOWLEDGMENTS

The authors gratefully acknowledge the contribution of NICTA. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

We would also like to thank the members of the NICTA AutoMap[5] team who have all contributed to the continued success and growth of this research.

REFERENCES

[1] N. Pettersson, L. Petersson, and L. Andersson, "The histogram feature – a resource-efficient weak classifier," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, June 2008.

[2] G. Overett, L. Petersson, L. Andersson, and N. Pettersson, "Boosting a heterogeneous pool of fast hog features for pedestrian and sign detection," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, June 2009.

[3] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1. IEEE Computer Society, 2001, p. 511.

[4] S. Lafuente-Arroyo, P. Gil-Jimenez, R. Maldonado-Bascon, F. Lopez-Ferreras, and S. Maldonado-Bascon, "Traffic sign shape classification evaluation I: SVM using distance to borders," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, 2005, pp. 557–562.

[5] C. Bahlmann, Y. Zhu, V. Ramesh, M. Pellkofer, and T. Koehler, "A system for traffic sign detection, tracking, and recognition using color, shape, and motion information," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, 2005, pp. 255–260.

[6] C. F. Paulo and P. L. Correia, "Automatic detection and classification of traffic signs," *Image Analysis for Multimedia Interactive Services, International Workshop on*, vol. 0, p. 11, 2007.

[7] B. Alefs, G. Eschemann, H. Ramoser, and C. Beleznai, "Road sign detection from edge orientation histograms," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, June 2007.

[8] A. Broggi, P. Cerri, P. Medici, P. Porta, and G. Ghisio, "Real time road signs recognition," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2007, pp. 981–986.

[9] R. Timofte, K. Zimmermann, and L. Van Gool, "Multi-view traffic sign detection, recognition, and 3D localisation," in *Applications of Computer Vision (WACV), 2009 Workshop on*. IEEE, 2010, pp. 1–8.

[10] G. Overett, L. Tychsen-Smith, L. Petersson, N. Pettersson, and L. Andersson, "Creating robust high-throughput traffic sign detectors using centre-surround hog statistics," *Submitted to: Machine Vision and Applications*, 2011.

[11] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human detection using a cascade of histograms of oriented gradients," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 2006, pp. 1491–1498.

[12] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on Riemannian manifolds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1713–1727, 2008.

[13] S. Paisitkriangkrai, C. Shen, and J. Zhang, "Fast pedestrian detection using a cascade of boosted covariance features," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 8, pp. 1140–1151, August 2008.

[14] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," *The Annals of Statistics*, vol. 28, no. 2, pp. 337–407, 2000.

[15] R. E. Schapire and Y. Singer, "Improved Boosting Algorithms Using Confidence-rated Predictions," *Machine Learning*, vol. 37, no. 3, pp. 297–336, 1999.

[16] J. Sochman and J. Matas, "Waldboost - learning for time constrained sequential detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 150– 156, 2005.

[17] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2005, pp. 886–893.

[18] H. Lappalainen and J. Miskin, "Ensemble learning," *Advances in Independent Component Analysis*, pp. 75–92, 2000.

[19] G. Overett and L. Petersson, "On the importance of accurate weak classifier learning for boosted weak classifiers," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, June 2008.

[20] ——, "Fast features for time constrained object detection," in *Feature Detectors and Descriptors: The State Of The Art and Beyond (CVPR Workshop)*, June 2009.

[21] F. Crow, "Summed-area tables for texture mapping," in *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*. ACM New York, NY, USA, 1984, pp. 207–212.

[22] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2nd Edition)*. Wiley-Interscience, November 2000.

[23] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian Detection: A Benchmark," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[5] http://www.nicta.com.au/research/projects/AutoMap