Occupant Monitoring System for Traffic Control Based on Visual Categorization

J. Javier Yebes, Pablo F. Alcantarilla, Luis M. Bergasa

Abstract— **This paper presents the basics of *Bag of visual words* method, which will be used for an occupant monitoring system that integrates a small onboard camera inside vehicles. It is intended to detect passengers' faces because it is the most appealing characteristic of occupants in a vehicle. This work proposes the implementation of visual categorization by means of two classification methods (Naïve Bayes and Multi-class SVM) that build multi-category image models using the invariant descriptors (SIFT and SURF) extracted from the images under analysis. Bag of visual words approach requires training in order to cluster invariant descriptors and learn the data distribution depending on the classification algorithm. Once the model is created, the category of every test image can be determined by querying a visual dictionary like searching a word in a text dictionary. The performance of the classifiers will be evaluated doing several comparative tests and using standard multi-category image databases. Experimental results and the conclusions are presented.**

I. Introduction

The monitoring and control of traffic volume is becoming a constant social, economic, and environmental pressure in the industrialized countries because of current infrastructure strain under an increasingly mobile population. The viability of high-occupancy vehicle (HOV) lanes for easing traffic congestion, and hence maximising traffic flow, has been proven in countries worldwide. To date, all enforcement has been manually applied by a police officer and some studies have concluded this is typically only 65% accurate due to several variables (e.g. environmental conditions, alertness of the officer, etc.) affecting the accuracy of the collected data. An automatic vehicle occupant counting system could replace human counters and facilitate the gathering of statistical data for traffic operations management, transportation planning, and construction programming. Also, it could provide the technical means to perform the HOV lane monitoring task more effectively, as well as facilitate enforcement to allow single-occupant vehicles to use the HOV lane for a fee [1]. This concept can be applied to parking lots in city centres too.

This paper presents an occupant monitoring system for traffic control in HOV lanes and city centres, based in computer vision techniques using a cheap, VGA camera with a wide angle lens, located inside vehicles. Furthermore, it is non-intrusive because it does not distract the driver and either requires manual operation. A communication module will receive the captured frames which will be wirelessly submitted to a server at the infrastructure side in order to process them and collect data about onboard occupants. The most appealing characteristic to detect and count vehicle occupants is the visual appearance of the face. However, due to the large variability in face appearances under driving conditions (changes in viewing direction, lighting...) we need a smart recognition scheme that can scale efficiently to a large number of cases as depicted in Fig. 1. Thus, we propose to do a visual categorization with well-known *Bag of visual words* method [2] [3].

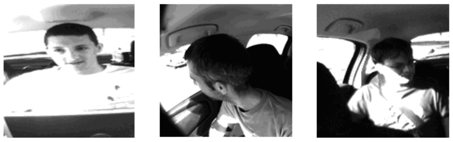


Fig. 1. Faces appearance under normal driving conditions.

In the remainder of the paper, we review vehicle occupancy technologies and other related works in Section II. Section III describes bag of visual words method [3] for diferent image features (SIFT [4] and SURF [5]) and two classifiers: Naïve Bayes [6] and Multi-class SVM [7]. The occupant monitoring system inside a vehicle is presented in Section IV and finally in Section V we show several perfomance results using a standard image dataset and another one built upon the images from our occupant monitoring system. We remark final conclusions and future works guidelines in Section VI.

II. Related Work

Most of the automatic systems for occupancy detection in the last years, are based on cameras placed on the infrastructure [8], [9], [10]. However, recently, the focus has shifted to in-vehicle sensing. The need for this technology has emerged out of legislation dealing with passenger airbag deployment and baby seats. The study [1] considers in-vehicle detection more suitable for enforcement in HOV lanes.

Four in-vehicle occupant detection systems are generally used today and are outlined in the synthesis

J. Javier Yebes, Pablo F. Alcantarilla and Luis M. Bergasa are with Department of Electronics, University of Alcalá. Alcalá de Henares, Madrid, Spain. e-mail: `javier.yebes, pablo.alcantarilla, bergasa@depeca.uah.es`

report [11]. In the last years, researchers have focused on optical sensing systems using cameras. Some works in this line are [12] [13]. They are intended to detect and even classify the occupants inside a vehicle in order to contribute adaptive restraint systems in cars.

A previous work by Yebes et. al. [14] revealed the inconvenients of using Viola&Jones algorithm [15] to detect passengers' faces for an occupant monitoring system. The wide angle lens radially distorts the image so that faces appearance is affected and Viola&Jones detector does not perform well. Additionally, occlusions, multi-view faces and illumination changes penalize the detector performance. On the other hand, undistortion has been proved unuseful too. The bag of visual words approach deals with these drawbacks and the work [14] introduces preliminar tests based on hierarchical clustering techniques [16]. The results demonstrate great improvements in occupants detection using visual categorization.

As a matter of fact, several image recognition algorithms have benefited from the great advances in invariant descriptors such as SIFT [4], SURF [5] and M-SURF [17], etc. Smart dictionaries or bag of visual words methods has become very popular thanks to its simplicity and good performance. There are applications from loop closure detection [18] in Simultaneous Localization and Mapping (SLAM) to CD cover recognition [19]. We have a great interest on this researching area and our occupant monitoring approach based on visual categorization is intended to give valuable information about the occupants inside the vehicle.

III. Visual categorization with Bag of visual words method

Bag of visual words approach is a supervised learning method similar to text categorization techniques [20]. In these works, text documents are represented as *bags of words* whose meaning will allow to determine the kind of document. Visual categorization using a similar approach implies the extraction of invariant descriptors (keypoints) from local image regions in order to create a set of keypoints. Then, these descriptors are quantized into visual words through the use of a K-Means clustering [21]. This *bag of visual words* allows a classifier to be trained. Hence, a model of the training data is created to implement a complex multi-class classifier. Figure 2 depicts the required steps in the training and testing stages of the proposed method [3].

In first place, training and testing images datasets are split according to a pseudo-random process and a training percentage over the total number of images in all the categories. Next steps are summarized in the following subsections.

A. Invariant descriptors

Visual categorization use local invariant descriptors as semantic features of images because of the invariance itself. These descriptors should be invariant to some variations, e.g. image affine transformations, blur and lighting variations. But they should carry enough information to be discriminative between images categories. Besides, they are robust to partial visibility and occlusions. Such tasks, require descriptors that are repeatable in the sense that if there is a transformation between two instances of an object, corresponding points are detected and identical descriptors are obtained around each. These properties deal with some inconvenients of Viola&Jones face detector [14] as explained in previous section.

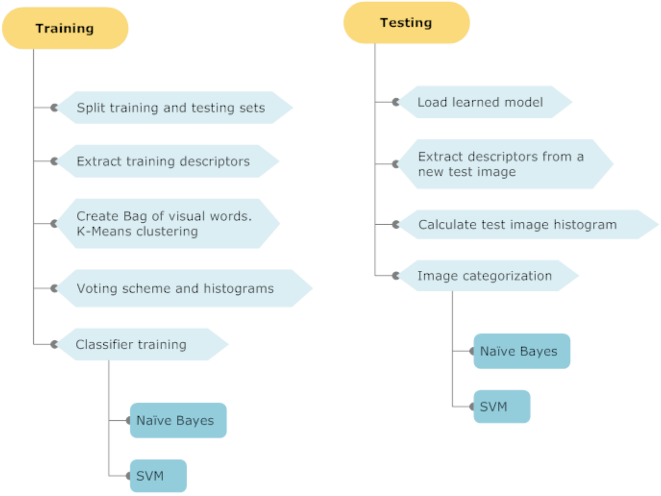

Training
Split training and testing sets
Extract training descriptors
Create Bag of visual words. K-Means clustering
Voting scheme and histograms
Classifier training
Naïve Bayes
SVM
Testing
Load learned model
Extract descriptors from a new test image
Calculate test image histogram
Image categorization
Naïve Bayes
SVM

Fig. 2. Bag of visual words categorization. Training and testing diagrams.

The extraction of local invariant descriptors from images comprises two algorithms. Firstly, a keypoint detector search for interesting features (high contrasted areas, blobs, corners, etc) where every detected keypoint keeps information about pixel position, scale and orientation. Secondly, descriptor vectors are computed in an area around each keypoint. In this work, three combinations of keypoints detector plus descriptors are considered in order to get a trade-off between classification and time performances:

- Difference of Gaussians (DoG) detector and Upright Scale Invariant Feature Transform (USIFT) descriptors [4], [22].
- Fast Hessian detector and Modified Upright Speed Up Robust Features (M-USURF) descriptors [5], [17], [23].
- STAR detector [24] and M-USURF descriptors [17].

All the descriptors in this work are obtained considering its upright mode. Rotation invariance is not considered because the object categories under analysis do not present great orientation changes.

B. Vector quantization.

Once a list of keypoints (all the invariant descriptors) have been extracted from the training images, they are clustered with K-Means [21] in order to quantize them into visual words (cluster centers).

The K value has to be large enough to distinguish relevant changes in image parts, but no so large to include irrelevant variations such as noise. We perform several tests to determine the best trade-off of accuracy and computational performance. Besides, ten iterations of K-Means with different initial seeds are done for chosing the set of clusters with the lowest compactness measure (Eq. 1). Cluster centers initialization is based on Arthur and Vassilvitskii algorithm [25].

$$compactness = \sum_i \left\| samples_i - centers_{labels_i} \right\|^2 \quad (1)$$

In the equation above, *samples* is a vector containing all the training descriptors, *centers* is the vector of the cluster centers determined by K-Means and *labels* assigns the corresponding center identifier to every descriptor.

C. Voting scheme and histograms

The clustering process builds a dictionary or bag of visual words representation. Next step computes the votes of every visual word given an image. The voting scheme is used for both training and testing stages, and it involves the computation of an histogram of descriptors for every image. This histogram is a vector of K bins that measures the number of occurrences of particular image patterns. Given the extracted descriptors from an image and the bag of visual words, it is selected the nearest j^{th} cluster center based on euclidean distance. Then, one vote is added to the j^{th} element of the histogram vector.

During the training of the multi-class classifier a NxK matrix is built, where N is the number of training images. On the other hand, the classification step only requires one histogram computation for each testing image.

D. Naïve Bayes classifier

As a first approach, we assume independance between visual words of different categories. Thus, Naïve Bayes classifier [6] is a simple and fast algorithm that estimates the maximum *a posteriori* probability for a generative model in which:

- A category depends on class prior probabilities,
- Every descriptor in an image is chosen independently from a multinomial distribution over descriptors specific to that class.

Given a training images set $I = \{I_i\}$ labeled according to categories $C = \{C_j\}$, a bag of visual words $V = \{v_k\}$ and the histogram $N(i,k)$ of descriptors of each image; to classify a new image into the available categories, Bayes' rule is applied and the largest *a posteriori* score is taken as the prediction:

$$P(C_j|I_i) \propto P(C_j)P(I_i|C_j) = P(C_j) \prod_{k=1}^{|V|} P(v_k|C_j)^{N(i,k)} \quad (2)$$

The probability of each visual word v_j given the category C_j is computed using the following formula:

$$P(v_k|C_j) = \frac{1 + \sum\limits_{\{I_i \in Cj\}} N(i,k)}{|V| + \sum\limits_{s=1}^{|V|} \sum\limits_{\{I_i \in Cj\}} N(i,s)} \quad (3)$$

E. Multi-class SVM classifier

As a second approach we use a kernel based method in order to match the histogram of an image to a specific category to solve the classification problem. Support Vector Machine (SVM) [7] creates a point representation in a space of n-dimensions and finds a (n-1)-dimesional hyperplane which separates two-class data with maximal margin. The solution is optimum in the sense that the distance between the hyperplane and the nearest points of each class (“support vectors“) is maximum.

In our case, the Multi-class SVM uses several binary SVM classifiers to discriminate point sets in multiple categories. On the other hand, we choose a linear kernel because of the high dimension space given the K bins (500 - 2.000) of the histograms, which are built upon bag of visual words representation. In addition, preliminar tests showed a good categorization performance.

During the training stage, Multi-class SVM creates the model from the input data: the histograms vectors for every training images and the category labels vector. For classifying a new image, its histogram is checked against the model in order to estimate the category among the availables.

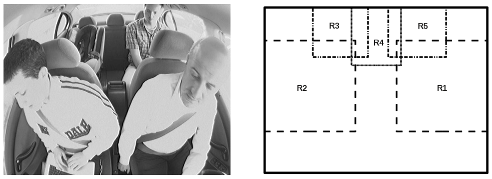
IV. Occupant monitoring system

Our occupant monitoring approach proposes the use of a wide angle lens mounted on a cheap digital camera, with low power consumption and small size. The field of view of the lens allows to frame all the onboard passengers in one image. The best camera location is attached to the roof, some centimeters in front of the rear-view mirror, with a small inclination angle (less than 45°) to frame all the passengers [14].

The captured frames will be lately processed at the infrastructure side as indicated in Section III in order to detect the presence of passengers in the vehicle. Furthermore, Bag of visual words does not consider geometric object information, but only visual appearance. Hence, for an accurate detection of each occupant face, every frame from the onboard camera is empirically split in 5 regions of interest surrounding each passenger as depicted in Fig. 3. Then, the classifier processes every region to label it as face or background.

V. Experimental Results

In this section we present classification rates and time performances of the method in order to choose the best combination of detector, descriptor and classifier. Two measurements are used: confusion matrix (Eq. 4) and overall error rate(Eq. 5).



(a) (b)

Fig. 3. (a) Sample frame from the onboard camera, (b) Five regions of interest.

$$M_{ij} = \frac{|\{I_k \in C_j : h(I_k) = i\}|}{|C_j|} \quad (4)$$

where i, j ∈ {1, ..., N_c}, C_j are the set of images from j^{th} category and $h(I_k)$ is the estimated category for image I_k.

$$ovr = 1 - \frac{\sum_{j=1}^{N_c} |C_j| M_{ij}}{\sum_{j=1}^{N_c} |C_j|} \quad (5)$$

We firstly make experiments using a standard dataset of images. After that, similar results are exposed for our image dataset from the occupant monitoring system.

A. *Classification results using a standard image dataset*

Tests in this section use a dataset of 3,150 images of different pixel resolutions divided into 7 categories of 450 items. They are: airplanes (AI), cars (CA), faces (FA), google things (GO), guitars (GU), houses (HO) and motorbikes (MO) [26]. A pseudo-random set of 784 images (25% of the whole dataset, 112 per category) are chosen for training, the rest of them for testing.

In the first experiment (Fig. 4), overall error rate and K number of clusters are compared for different invariant descriptors and sizes during the testing stage of Naïve Bayes classifier.

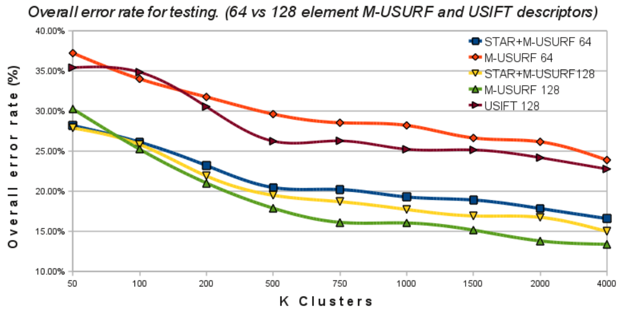


Fig. 4. Overall error rate vs K clusters.

Execution times for training stage considering a 2GHz CPU core, are divided into 4 groups:

- Extraction of keypoints from all the training images introduces a mean delay of: 150ms for STAR+M-USURF 64- and 128-elements descriptors, 290ms for FH+M-USURF 64- and 128-elements descriptors and 600ms for DoG+USIFT 128-elements descriptors.
- K-Means execution times are displayed in Fig. 5.
- Naïve Bayes training CPU times are shown in Fig. 6.
- Keypoint extraction plus Naïve Bayes classification of a new test image. (K=2000). See Table I.

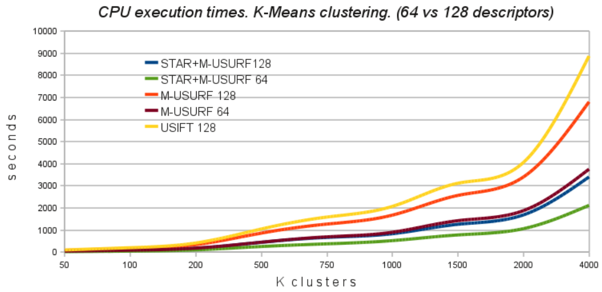


Fig. 5. K-Means CPU times vs K clusters.

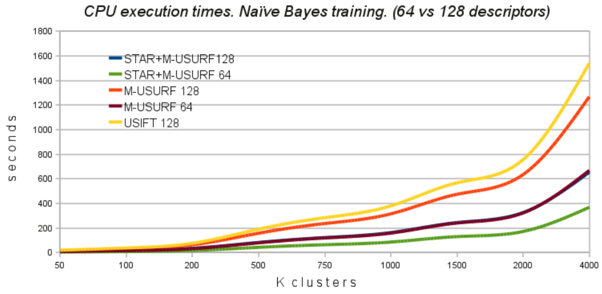


Fig. 6. Naïve Bayes training CPU times vs K clusters.

TABLE I

Naïve Bayes classification delay ranges for a new test image.

	64-descriptor	128-descriptor
DoG+USIFT	—	800 - 2800ms
FH+M-USURF	500 - 800ms	600 - 1500ms
STAR+M-USURF	300 - 700ms	500 - 1300ms

Looking at Fig. 4 - Fig. 6, the best trade-off between classifier performance and CPU times is K = 2000 clusters. For this value, Fast Hessian plus M-USURF 128-elements descriptors lead to the lowest overall error rate (13.78%). Hence, Table II summarizes the Naïve Bayes classification results for this configuration. The highest missclasification rate is for the "google things" category [26], due to its images overlaps with the remaining categories of the experiment.

Considering the second classifier approach, i.e. Multi-class SVM, the overall error rate evolution as K increases is very similar to Fig. 4. However, Multi-class SVM classifier yields better results than Naïve Bayes.

TABLE II

Confusion matrix for Naïve Bayes classification. K = 2000, FH+M-USURF 128.

True category						
AI	CA	FA	GO	GU	HO	MO
90.82	0	0.88	2.95	0	6.21	0.88
0.29	**97.92**	0	0.29	0	0.88	0
3.25	0	**90.53**	13.01	4.43	0	0.59
1.18	0	8.28	**50**	7.10	1.47	1.18
0	0	0.29	13.31	**87.28**	0	0.29
3.55	0.88	0	5.32	0.29	**89.94**	0
0.88	1.18	0	15.08	0.88	1.48	**97.04**

Table III is the resultant confusion matrix for K=2000 and FH+M-USURF 128-elements descriptors. For this test, the corresponding overall error rate is 8.81%.

Time delays during Multi-class SVM training are slightly higher (about 30s more) than Naïve Bayes. However, the required time to classify a new test image shows very similar values compared to Naïve Bayes (see Table I).

TABLE III

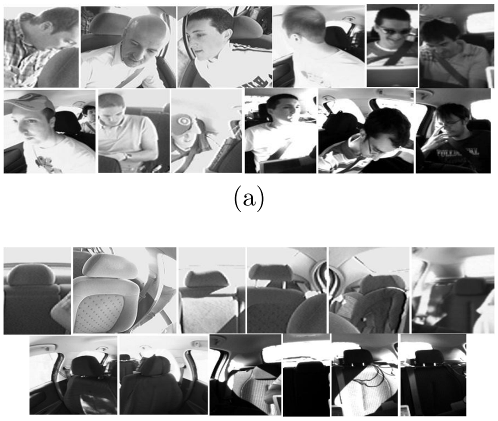
Confusion matrix for Multi-class SVM classification. K = 2000, FH+M-USURF 128.

True category						
AI	CA	FA	GO	GU	HO	MO
96.15	0	0.29	1.47	0.29	4.14	1.47
0	**98.82**	0	2.36	1.18	2.36	0.88
0	0	**98.81**	4.73	1.18	0	0.59
0.88	0.88	0.59	**81.65**	11.54	6.21	4.44
0.59	0	0.29	1.47	**85.21**	0	0
0.59	0.29	0	4.73	0.29	**86.39**	1.47
1.77	0	0	3.55	0.29	0.88	**91.12**

B. Classification results using the occupant monitoring image dataset

Occupants' face detection based on visual categorization divides images into two classes: faces and background. We have built a dataset composed of 1,400 image patches equally divided for each class. These patches corresponds to the region splitting process depicted in Fig. 3. The original frames of 640x480 pixels belong to some recorded videos inside a vehicle, at daytime and under a variety of environment conditions. Dataset images have been manually selected in order to include typical cases: illumination changes, partial illuminated areas and shadows over the seats and passengers, several users including people with glasses, sunglasses, caps, changes in face direction and orientations, etc. Fig. 7 displays a random subset of patches from each category.

Tables IV and V summarizes the best classification rates for Naïve Bayes and Multi-class SVM classifiers respectively. After several tests and considering soft real time requirements for the occupant monitoring system, we have chosen a trade-off scenario in which 30% of images are for training. Extraction of descriptors is performed with STAR detector and M-USURF 64-elements descriptors and K = 2,000 clusters. The obtained overall error rates are 19.2% and 12.5% respectively. The total CPU times for training are about 55 and 63 seconds, while a query to classify a new image patch is in the range of nearly 100 ms in Naïve Bayes case and 50 - 180 ms in Multi-class SVM.


(a)

(b)

Fig. 7. Occupant monitoring dataset: (a) Faces (b) Background.

TABLE IV

Occupant monitoring confusion matrix. Naïve Bayes classifier

		True category	
		Background	Faces
Estimated category	Background	**93.47**	31.83
	Faces	6.53	**68.17**

TABLE V

Occupant monitoring confusion matrix. Multi-class SVM classifier

		True category	
		Background	Faces
Estimated category	Background	**78.77**	3.67
	Faces	21.23	**96.33**

Naïve Bayes creates an independent feature model that do not deal with bias between different classes. As a consequence, in Table IV the highest classification rate corresponds to the background category due to the presence of partial seats and other car parts in the images of the two classes. Thus, nearly 1/3 of faces' patches are classified as background. On the other hand, Multi-class SVM is focused on the distribution of votes over the visual words, using a linear kernel to match an image to a category. Hence, in the model face features are easier to separate from common visual features in seat covers, then image patches of faces are more distinguishable too. Nevertheless, a linear boundary between classes might not be the best approach and false face detection grows. However, a non-linear kernel will increase the model complexity and CPU execution times.

VI. Conclusions and Future Work

This paper has presented a visual categorization approach with bag of visual words method, which has been applied to an occupant monitoring system for traffic control in HOV lanes. Several comparative results using two image datasets have been shown.

Fast Hessian detector and M-USURF descriptors of 128 elements produces the best result in terms of overall error rate for both Naïve Bayes and Multi-class SVM classifiers. Nevertheless, it comprises higher CPU delays than other descriptors (except USIFT which shows the worst performances). Combination of STAR detector plus M-USURF descriptors of 64 elements leads to the fastest results in terms of computation and memory consumption, while overall error rate presents mean values over the rest. It is a good choice for soft real time applications like the occupant monitoring system proposed in this paper. Besides, we chose K = 2,000 clusters because it is a good trade-off between classification rates and computation times.

Additionally, the set of experiments carried out in this work indicates that Multi-class SVM yields a reduction of 5% in the overall error rate compared to Naïve Bayes, whilst CPU times for training and testing remain similar.

As future guidelines we are interested in further optimizations of the bag of visual words method in order to reduce false positives in face detection. As a solution, we propose the implementation of a multi-frame approach in order to process real-time videos. Another future work is the integration with the communication module and systems at the infrastructure side for traffic control tasks.

Furthermore, we are also interested in advance research of visual categorization methods improving the voting scheme and classification algorithms.

VII. Acknowledgments

This work was supported in part by the Spanish Ministry of Education and Science (MEC) under grant TRA2008-03600 (DRIVER-ALERT Project), by the multinational corporation FICOSA INTERNATIONAL and by Comunidad de Madrid through the project RoboCity2030 II-CM (CAM - S2009/DPI-1559).

References

[1] McCormick Rankin Corporation, “Automated Vehicle Occupancy Monitoring Systems for HOV/HOT Facilities,” 2004, Ontario Ministry of Transportation. Canada.

[2] B. Fulkerson, A. Vedaldi, and S. Soatto, “Localizing objects with smart dictionaries,” in *Eur. Conf. on Computer Vision (ECCV)*, 2008.

[3] G. Csurka, C. Bray, C. Dance, and L. Fan, “Visual categorization with bags of keypoints,” *Workshop on Statistical Learning in Computer Vision, ECCV*, pp. 1–22, 2004.

[4] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *Intl. J. of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[5] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, “SURF: Speeded up robust features,” *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[6] D. D. Lewis, “Naive Bayes at Forty: The Independence Assumption in Information Retrieval,” in *ECML*. Springer Verlag, 1998, pp. 4–15.

[7] “LibSVM. A Library for Support Vector Machines,” http://www.csie.ntu.edu.tw/ cjlin/libsvm/.

[8] I. Pavlidis, P. Symosek, B. Fritz, M. Bazakos, and N. Papanikilopoulos, “Automatic detection of vehicle occupants: the imaging problem and its solution.” *Machine Vision and Applications*, vol. 11, no. 6, pp. 313–320, 2000.

[9] J. R. Tyrer and L. M. Lobo, “An optical method for automated roadside detection and counting of vehicle occupants,” in *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 222, no. 5, January 2008, pp. 765–774.

[10] A. Pérez-Jiménez, J. Guardiola, and J. Pérez-Cortés, “High occupancy vehicle detection,” vol. 5342, pp. 782–789, 2008, 10.1007/978-3-540-89689-0_82. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-89689-0_82

[11] J. Wikander, “Automated vehicle occupancy technologies study: Synthesis report,” Texas Transportation Institute. The Texas A&M University System, Tech. Rep., 2007.

[12] Y. Yang, G. Zao, and J. Sheng, “Occupant Pose and Location Detect for Intelligent Airbag System Based on Computer Vision,” *International Conference on Natural Computation*, vol. 6, pp. 179–182, 2008.

[13] A. Makrushin, M. Langnickel, M. Schott, C. Vielhauer, J. Dittmann, and K. Seifert, “Car-seat occupancy detection using a monocular 360° nir camera and advanced template matching,” in *DSP’09: Proceedings of the 16th international conference on Digital Signal Processing*. Piscataway, NJ, USA: IEEE Press, 2009, pp. 1044–1049.

[14] J. J. Yebes Torres, P. Fernández Alcantarilla, L. M. Bergasa Pascual, and Á. González, “Occupant monitoring system for traffic control in hov lanes and parking lots,” in *Workshop on Emergent Cooperative Technologies in Intelligent Transportation Systems*, 2010.

[15] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2001.

[16] A. Vedaldi and B. Fulkerson, “Bag of visual words implementation,” http://www.vlfeat.org/ vedaldi/code/bag/walk.html, 2008.

[17] M. Agrawal, K. Konolige, and M. R. Blas, “CenSurE: Center Surround Extremas for realtime feature detection and matching,” in *Eur. Conf. on Computer Vision (ECCV)*, 2008.

[18] A. Angeli, D. Filliat, S. Doncieux, and J. A. Meyer, “Fast and Incremental Method for Loop-Closure Detection using Bags of Visual Words,” *IEEE Trans. Robotics*, vol. 24, pp. 1027–1037, 2008.

[19] D. Nistér and H. Stewénius, “Scalable recognition with a vocabulary tree,” in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2006.

[20] T. Joachims, “Text categorization with support vector machines: Learning with many relevant features,” in *ECML*. Springer Verlag, 1998, pp. 137–142.

[21] A. Moore, “K-means and Hierarchical Clustering - Tutorial Slides,” http://www-2.cs.cmu.edu/ awm/tutorials/kmeans.html.

[22] A. Vedaldi and B. Fulkerson, “Lightweigth c++ implementation of sift detector and descriptor,” http://www.vlfeat.org/ vedaldi/code/siftpp.html, 2006.

[23] C. Evans, “Notes on the OpenSURF library,” University of Bristol, Tech. Rep. CSTR-09-001, January 2009. [Online]. Available: http://www.cs.bris.ac.uk/Publications/Papers/2000970.pdf

[24] “STAR detector OpenCV. Willow Garage wiki.” http://pr.willowgarage.com/wiki/Star_Detector.

[25] D. Arthur and S. Vassilvitskii, “k-means++: the advantages of careful seeding,” in *SODA ’07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 2007, pp. 1027–1035.

[26] “Visual Geometry Group,” http://www.robots.ox.ac.uk/∼vgg /data3.html.