Road Boundary Detection in Challenging Scenarios

Mohamed A. Helala, Ken Q. Pu, Faisal Z. Qureshi

Faculty of Science, University of Ontario Institute of Technology, Oshawa, ON, Canada

{Mohamed.Helala, Ken.Pu, Faisal.Qureshi}@uoit.ca

Abstract—This paper presents a new approach for automatic road detection in traffic cameras. The technique proposed here detects the dominant road boundary and estimates the vanishing point in images captured by traffic cameras under a wide range of lighting and environmental conditions, e.g., in images of unlit highways captured at night, etc. The approach starts by segmenting the traffic scene into a number of superpixel regions. The contours of these regions are used to generate a large number of edges which are organized into clusters of co-linearly similar sets using hierarchical bottom up clustering. A confidence level is assigned to each cluster using a statistical approach and the best clusters are chosen. Pairs of clusters with high confidence levels are then ranked and filtered according to image perspective and activity. The top ranked pair is selected as the road boundary. The proposed technique is tested on a real world dataset collected from the Ontario 401 traffic surveillance system. Experimental results demonstrates a distinct speedup and improvement in accuracy of the proposed technique in detecting the dominant road boundary in challenging scenarios compared to the state of the art Gabor filter based technique.

Keywords-Road Boundary Detection; Hierarchical Clustering; Superpixel Segmentation.

I. Introduction

Traffic surveillance cameras are ubiquitous [1], [2]. Traffic cameras are being deployed all over the globe in an effort to 1) automate road toll collection, 2) analyze traffic, 3) enhance road safety, 4) increase situational awareness of emergency personnel, etc. Given the shear amount of video data being collected through these systems at any given time, it is concomitant to develop algorithms capable of analyzing this video data with little human input. Within this context an important enabling capability is automatic road detection. Once the road region is identified subsequent processing can focus on these regions to measure traffic flows, road incidents, erratic or dangerous driving, etc.

Automatic road detection in traffic surveillance video is challenging due to the range of environmental conditions under which these systems operate: rain, snow, sunshine, day, night, fog, etc. Several techniques have been proposed for road detection in traffic cameras. Broadly speaking these can be divided into three categories: (1) activity-driven [3], [4], (2) feature-driven [5], [6], [7] and (3) model-driven [8], [9]. The activity-driven approaches benefit from the relatively high vehicle motion activity along the roads. These generally work by extracting an activity map which divides the image into active (road) and inactive (non-road) regions.

Algorithm 1 Finding dominant road boundaries in traffic sequences: an outline

---

Input: Image sequence
Output: Dominant road boundary
 1: Divide each image into homogeneous regions through superpixel segmentation.
 2: Approximate each superpixel contour with polygons to get edges.
 3: Perform bottom up hierarchical clustering of these edges.
 4: Use statistical measures (χ^2 and Student-T Test) to identify the top-ranked clusters, each cluster represents a road boundary in the image.
 5: Construct top-ranked cluster pairs through perspective filtering and road-activity analysis.
 6: Top-ranked cluster pair is returned as the dominant road boundary.

Feature-driven approaches rely upon the extraction of image features to detect lane and road boundaries. Model-driven approaches aim to match a road model to the image.

This work develops a novel scheme for detecting dominant road boundaries in traffic cameras (see Algorithm 1). Our approach looks for stable edges in an image sequence to find the dominant road boundary. Since the proposed approach accumulates evidence over multiple images, it is immune to traffic variations that plague schemes relying upon a single frame to identify road boundaries. We compare our approach with another recent scheme that uses Gabor filters to detect dominant road boundaries [7]. We show that our approach outperforms the Gabor filter approach.

The proposed technique is fully automatic and requires no human input. It can detect road regions under challenging environmental conditions, such as day, night, rain, snow, wet roads that exhibit strong reflections from vehicle headlights, dark roads that are seen only through vehicle headlights, etc. The proposed technique is aimed at low-framerate, low-resolution imagery; however, it works equally well when presented with high-framerate, high-resolution imagery. We have evaluated the proposed technique on a manually labeled real world image sequence dataset collected from the Ontario 401 traffic surveillance system. A secondary contribution of this work is the cluster selection strategy that we have devel-
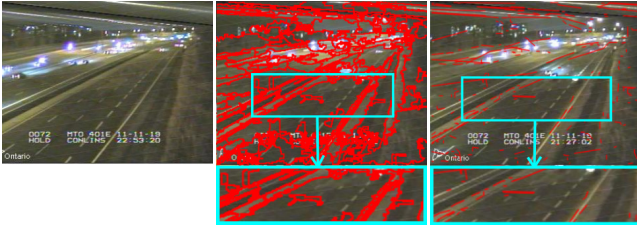
Figure 1: Superpixel segmentation. (Left) original image. (Middle) superpixel boundaries. (Right) approximating superpixel contours with polygons and showing edges that are more than 10 pixels in length.

oped for choosing the most promising clusters in hierarchical bottom up clustering.

The remainder of this paper is organized as follows. Sec. II describes the proposed technique. Next we present experimental evaluation in Sec. III. Sec. IV concludes the paper.

II. ROAD BOUNDARY DETECTION

The proposed technique consists of five steps: superpixel segmentation, contour approximation, hierarchical bottom up clustering, confidence assignment, and pairwise ranking. We discuss each of these steps below.

A. Superpixel Segmentation

Superpixel segmentation, originally proposed by Ren and Malik [10], segments an image into a collection of homogeneous regions. Several approaches have been proposed to perform superpixel segmentation [11], [12], [13], [14]. Generally speaking all of these techniques are slow and not suitable for our application. We are interested in near real-time superpixel segmentation. [15] develops a fast algorithm for performing superpixel segmentation, but it requires manual seed-point selection. Instead we employ the approach described in [16] to perform superpixel segmentation at near real-time rates (see Fig. 1). This technique starts by performing morphological open/close operations on the input image. Next Hue/Saturation/Value color quantization is applied, fusing neighboring regions having similar color values in the process. Region labeling, sorting and merging addresses the over-segmentation that typically occurs when performing superpixel segmentation. For further details we refer the kind reader to [16].

We observe that while the size of superpixels is sensitive to the lighting and environmental conditions under which the image was captured, superpixel boundaries are still able to identify stable edges in the scene. By accumulating these edges over a sequence of images, road region can be identified even in the presence of extreme lighting and environmental conditions.

B. Contour Approximation

Edges that feed into the hierarchical clustering are computed by approximating the boundary of each superpixel with a polygon. Superpixel boundaries are approximated by polygons through adaptive sampling [17]. Initially each superpixel boundary is sampled at three points (start, middle, and end). Tangent values at these points are computed to determine the degree of colinearity of the three points. The sampling continues as long as the degree of colinearity is below a certain threshold. This approach can sometimes lead to over-sampling, i.e., a straight section of the superpixel boundary is represented by three or more line segments. We resolve this be deleting the intermediate vertices (sample points) that lie on a straight segment. Fig. 1(c) shows the polygon edges that are at least 10 pixel long.

C. Hierarchical Bottom up Clustering

Hierarchical bottom up clustering with average linkage is applied to the segments of the polygons approximating superpixels from an image sequence. We represent each segment $\mathbf{s}$ in polar coordinates giving rise to 2D vectors of the form (ρ, θ). Let $\{\mathbf{s}_i | i = 1 \cdots n\}$ be the set of all segments for a given image sequence. We normalize the segments such that these have zero mean and unit variance. Henceforth we use $\mathbf{s}_i = (\rho_i, \theta_i)$ to refer to the ith normalized segment. The set $\mathcal{S}$ of normalized segments are fed into the bottom up hierarchical clustering as inputs.

Figure 2: Segments from three clusters along a single path in the cluster hierarchy tree. (Left) root cluster. (Middle) A cluster with large variance and a large number of samples. (Right) A cluster with small variance and a large number of samples, representing a road boundary.

Hierarchical bottom up clustering starts by creating a cluster c_i for each segment $\mathbf{s}_i$. The closest pair of clusters is merged to form a single cluster as one moves up the hierarchy. The process continues until only a single cluster is left (root). Distance between two clusters is defined to be the Euclidean norm in the Hough space. Let n_i be the number of segments in cluster c_i. Furthermore we define $\mathcal{S}_i$ to be the set of segments belonging to cluster c_i. Then the average distance between two clusters c_i and c_j is defined as,

$$d(c_i, c_j) = \frac{1}{n_i n_j} \sum_{\mathbf{s}_i \in \mathcal{S}_i} \sum_{\mathbf{s}_j \in \mathcal{S}_j} |\mathbf{s}_i - \mathbf{s}_j|. \tag{1}$$
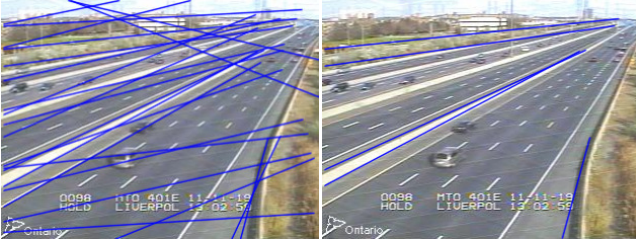
Figure 3: Clusters with high confidence value using a 25 frame sequence. (Left) Cluster means corresponding to clusters containing at least 30 segments. (Right) Cluster means corresponding to clusters containing at least 50 segments.
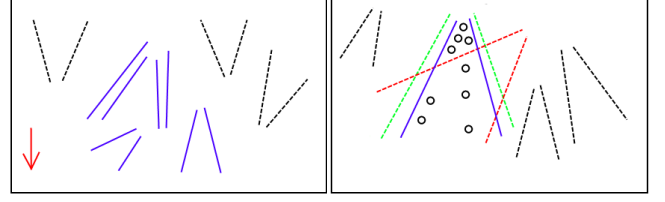

Figure 4: Perspective filtering and image activity based ranking cluster pairs. The solid lines indicate mean lines corresponding to cluster pairs that survive the perspective filtering and activity ranking. (Left) Perspective filtering with respect to the red arrow indicating the downward direction. (Right) Activity based filtering. Here circles represent vehicular traffic.

D. Confidence Assignment

A confidence level is assigned to each cluster in order to penalize clusters with high variance or small number of segments. For each cluster, we have the cluster mean $\overline{x}$ and variance s^2 to be unbiased estimation of the unknown true mean x^* and true variance σ^2. Intuitively a cluster with a small variance and a large number of segments should have a higher chance for the estimated mean to be closer to the true mean. This can be stated as $P\left[|x^* - \overline{x}| \leq \varepsilon\right]$, for a small error ε. The expectation of this probability is

$$
E\left(P\left[|x^* - \overline{x}| \leq \varepsilon\right]\right) = \\
\int_{-\infty}^{\infty} P_{\overline{x}}(u) \int_{0}^{\infty} P_{s^2}(v) \int_{\overline{x}-\varepsilon}^{\overline{x}+\varepsilon} \mathcal{N}(w|\mu = u, \sigma^2 = v)dw\, dv\, du,
$$
(2)

where the cluster is modeled as a multivariate normal distribution $\mathcal{N}(\mu, \sigma^2)$ evaluated over a confidence interval $\pm\varepsilon$ around the cluster estimated mean $\overline{x}$. The first and second integrals represent the likelihood of various assignments to the random variables $\overline{x}$ and s^2 around the cluster true mean x^* and true variance σ^2. This integral is difficult to solve for each cluster. Instead we can factor this integral and approximate the likelihood terms over confidence intervals as

$$
\int_{-\infty}^{\infty} P_{\overline{x}}(u)du \int_{0}^{\infty} P_{\sigma^2}(v)dv \approx \int_{\overline{x}-\alpha}^{\overline{x}+\alpha} P_{\overline{x}}(u)du \int_{s^2-\beta}^{s^2+\beta} P_{s^2}(v)dv,
$$
(3)

where α and β define confidence intervals around the true mean and true variance, respectively. Here k is the degree of freedom which equals $n-1$, n being the number of samples in the cluster. The first likelihood term can be modeled by Student-T distribution $\text{Stud}_x(\overline{x}, s^2, k)$ [19], where the likelihood is calculated over a confidence interval α and has a high value when a cluster has low variance and large number of samples. The second likelihood term can be modeled by a Chi-square (χ^2_k) distribution over a confidence

interval β [19], where the likelihood has a high value when the number of samples is large, indicating good estimation of cluster variance. We can then define the confidence value for a cluster c_i as

$$
\text{Conf}(c_i) = \int_{\overline{x}-\alpha}^{\overline{x}+\alpha} P_T(u|s^2, k)du \int_{s^2-\beta}^{s^2+\beta} P_{\chi^2}(v|s^2, k)dv \quad (4)
$$

where the interval α is calculated from the Student-T statistics to be $\pm 0.05\frac{s}{\sqrt{n}}$ and the interval β is calculated from the (χ^2_k) statistics to be $(k, 1.2k)$. This assigns low confidence values to 1) clusters with large variances and to 2) clusters with low variances but having only few samples. Clusters that have large number of samples and low variances will be assigned high confidence values. These clusters correspond to dominant edges in the image. Since we integrate information from multiple frames, these edges also need to be persistent across multiple images. Fig. 3 shows the top 20% high confidence clusters with different values for the number of segments per cluster.

Lines corresponding cluster means are constructed by projecting cluster segments onto the cluster means, which determines the extent of support provided by cluster members for its mean. This serves two purposes: 1) clusters can be rejected if the length of its mean line segment is less than some threshold and 2) detect road boundaries that may have gone undetected between frames.

E. Pairwise Ranking

Although the means of the high confidence clusters can determine the dominant lines in a frame sequence, several lines can be misclassified as the the road boundary. This is resolved by constructing cluster pairs and ranking these pairs based upon perspective and image-activity cues. Lines corresponding to top ranked clusters are paired with each other to find the top-ranked pair, representing the dominant road boundary. Specifically, for each pair, the vanishing point

is calculated. The pairs are then filtered toward the camera perspective view of roads which makes the road boundary lines (parallel in real world) have an angle of 90 degrees at max with a vector that point downwards in the image plane (Fig. 4(a)). Pairs that do not meet this requirement are rejected. Confidence score r_{Conf} and activity score r_{Activity} are computed for each of the remaining pairs. The overall rank of a pair (i, j) is

$$r(i,j) = r_{\text{Conf}}^{(i,j)} \times r_{\text{Activity}}^{(i,j)},$$

where

$$r_{\text{Conf}}^{(i,j)} = \text{Conf}(c_i) \times n_i \times \text{Conf}(c_j) \times n_j$$

and

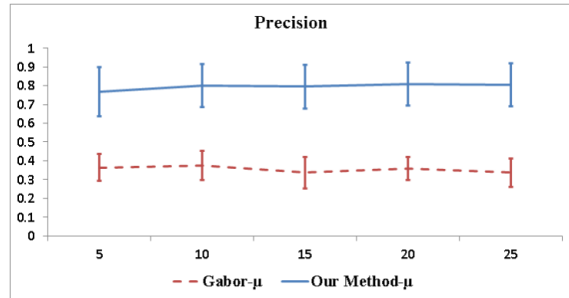$$r_{\text{Activity}}^{(i,j)} = \frac{\#\text{objects in the region}}{\text{area of the region}}.$$

The region is defined to be the image region enclosed by the lines representing cluster means. $r_{\text{Conf}}^{(i,j)}$ prefers cluster pairs containing large number of segments and having stable means; whereas, $r_{\text{Activity}}^{(i,j)}$ prefers cluster pairs that enclose image activity.[1] The image activity is extracted by segmenting the moving vehicles using background subtraction [20] and calculating object centroids. Fig. 4(b) illustrates perspective and image activity based ranking of cluster pairs.
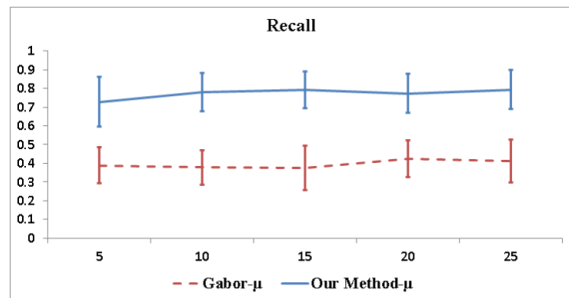
III. EXPERIMENTAL RESULTS

We have tested the proposed road boundary detection strategy on a dataset collected from 14 cameras installed on the Ontario 401 highway. These cameras are part of the 401 traffic surveillance system. We have manually labeled 50 frames from each location to serve as ground truth. 25 of these frames are taken during the day time; where as, the rest are taken during the night time. Out of 14, two cameras are installed at locations where the highway is not lit at night. Here the road can only be seen through vehicle headlights. The dataset shows large variations in color, texture, illumination and the surrounding environment. The frames are recorded 15 to 20 minutes apart to ensure coverage of different environmental conditions. Frame resolution is 320×240. We have intentionally avoided high-resolution/high-frame rate images to demonstrate that the proposed road boundary detection strategy works well under resource constrained settings. For our experiments, the number of detected superpixels are at most 100 depending on the scene details.

The proposed technique is capable of detecting the dominant road boundary that surrounds the active regions in the image. The technique is compared to the Gabor filter based technique that appeared in [7], which uses Gabor filter outputs to vote for the vanishing point corresponding to the
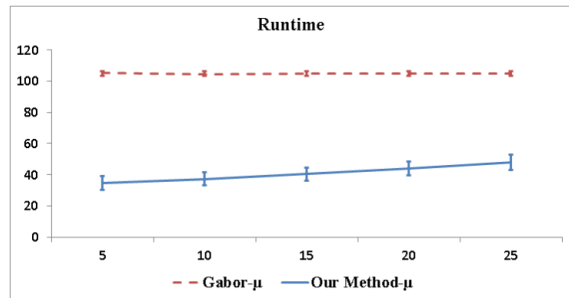

(a) Precision

(b) Recall

(c) Runtime Performance (in seconds)

Figure 5: Statistical comparisons between the proposed method and the Gabor filter based method [7]. X-axis represents the frame number.

most dominant road boundaries. Gabor filters are also used in [9] to estimate the vanishing point.

Figure 6 shows sample results in 8 different locations. For each location, the first image is the ground truth, the second is our result computed over 25 frames, and the third is the Gabor filter based result. For Gabor filter based method, we randomly selected one of the 25 frames to find the dominant road boundary. The Gabor filter based method is sensitive to noise, illumination changes, light reflections, and unlit regions of the roads. However, Our technique is able to return a reasonable estimate of the dominant road boundaries under severe lighting and environment conditions. The Gabor filter based method, being memory-less, is also affected by the current traffic conditions.

The results demonstrate that our method is able to detect

---

[1]The assumption being that dominant image activity is the result of vehicular traffic in traffic surveillance images.
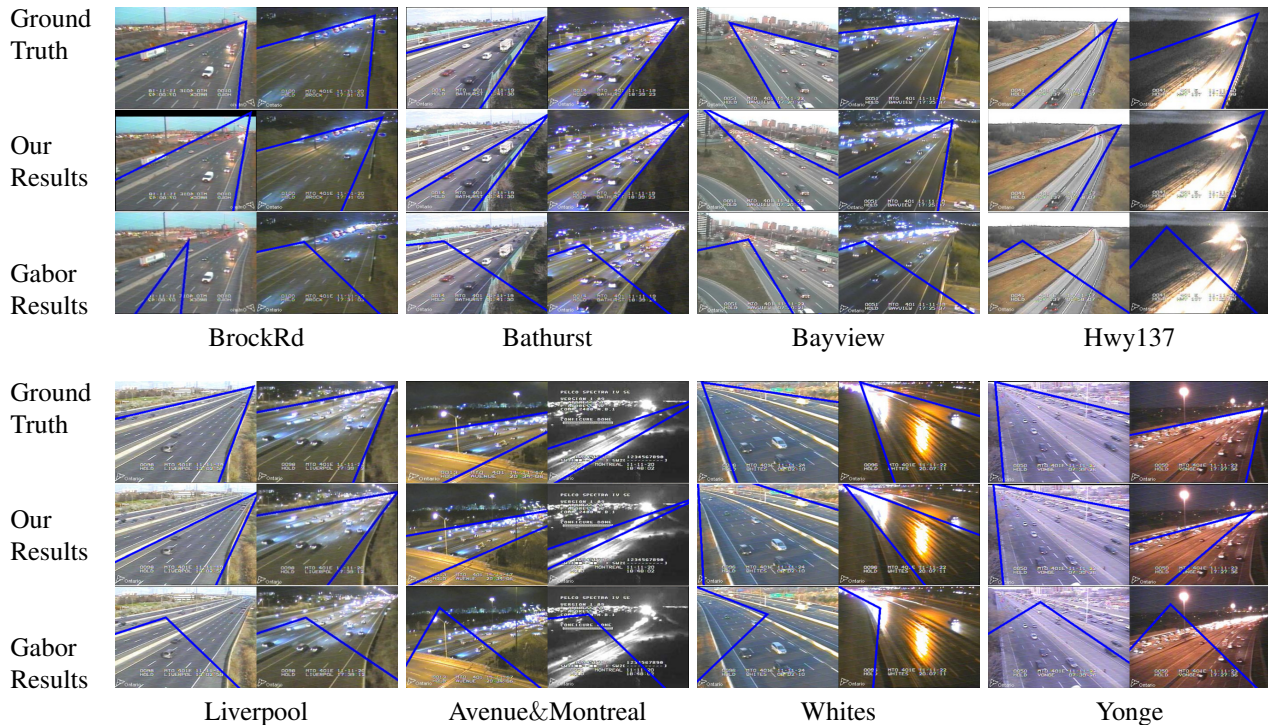
Figure 6: Results of the proposed technique compared to the Gabor-based approach [7] in 9 different locations in day/night lighting. The first image in each location is the ground truth, the second is our results and the third is Gabor-based results.

the dominant road boundary correctly under a wide-variety of challenging lighting and environmental conditions. The BrockRd, Yonge and LiverPool locations provide example of day and night lighting conditions. A side lane is visible in the Bayview location. The Hwy137 location is unlit. Here at night the road is only seen through vehicle headlights. The Whites location illustrates how vehicle headlight reflections can be nuisance under wet road conditions. The Bathurst and Bayview location shows reflections of vehicle lights that disrupt the road region. At the Avenue location the camera's view of the dominant road is blocked by another road.

The precision and recall comparisons [21] have been used to evaluate the performance of the proposed approach. Manually labeled ground truth road regions are defined for each location. The precision is defined as the ratio of the intersected area of the estimated road boundary and the ground truth to the area of the estimated boundary. The recall is the ratio of the intersected area of the estimated boundary and the ground truth to the area of the estimated truth.

Figure 5a shows the precision comparison between our method and the Gabor filter-based technique. It can be seen that our method achieved 80% precision on average with $23\%\sigma$ after 25 samples. The Gabor filter based technique achieved 35% on average with $16\%\sigma$. Figure 5b shows the recall comparison where the proposed method achieved 77% on average with $20\%\sigma$ compared to the Gabor filter based

technique which achieved 40% on average with $23\%\sigma$.

Runtime comparison of our approach with the Gabor filter based approach is shown in Figure 5c. Experiments were executed on an 2.9GHz Quad-core AMD Athlon II 635 machine with 6GB RAM. Our method achieved 40sec on average over 25 image sequence with a standard deviation of 8.7 sec, while the Gabor filter based technique achieved 105sec on average with a standard deviation of 1.8 sec. The proposed technique clearly outperforms the Gabor filter based method.

These results confirm the effectiveness of the proposed technique in estimating the dominant road boundary in several difficult lighting and environmental conditions. Additionally, our method outperforms the Gabor filter based state of the art technique in terms of precision, recall and runtime.

IV. CONCLUSION AND FUTURE WORK

This paper presents a new technique for dominant road boundary detection. The proposed method is evaluated on a realistic dataset gathered from traffic surveillance cameras installed on the Ontario 401 highway. The dataset exhibits a large variation in lighting and environmental conditions. We compare our method against a state-of-the-art Gabor filter based method, and our method outperform the exiting method in terms of precision, recall and runtime

performance. We plan to improve the performance of our algorithm by means of incremental hierarchical algorithms implemented in a distributed map-reduce framework. We also wish to explore applying GPU-based techniques to further improve the proposed algorithm so it can be applied to traffic video sequences in real-time.

REFERENCES

[1] V. Kastrinaki, M. Zervakis, and K. Kalaitzakis, "A survey of video processing techniques for traffic applications," *J. Image and Vision Comp.*, vol. 21, no. 4, pp. 359–381, April 2003.

[2] N. Buch, S. Velastin, and J. Orwell, "A review of computer vision techniques for the analysis of urban traffic," *J. IEEE Trans. on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 920–939, Sept. 2011.

[3] B. Stewart, M. T. I. Reading, T. Binnie, K. Dickinson, and C. Wan, "Adaptive lane finding in road traffic image analysis," in *in Proc. 7th IEEE Int. Conf. Road Traffic Monitoring and Control*, Napier Univ. Edinburgh, Apr. 1994, pp. 133–136.

[4] J. Melo, A. Naftel, A. Bernardino, and J. Santos-Victor, "Detection and classification of highway lanes using vehicle motion trajectories," *J. IEEE Trans. on Intelligent Transportation Systems*, vol. 7, no. 2, pp. 188–200, June 2006.

[5] A. Soumelidis, G. Kovacs, J. Bokor, P. Gaspar, L. Palkovics, and L. Gianone, "Automatic detection of the lane departure of vehicles," in *in Proc. 8th IFAC Symposium on Transportation Systems*, Chania, Greece, June 1997, pp. 1045–50.

[6] M. Aly, "Real time detection of lane markers in urban streets," in *in Proc. IEEE Intelligent Vehicles Symposium*, Eindhoven, The Netherlands, June 2008, pp. 7–12.

[7] H. Kong, J. Audibert, and J. Ponce, "Vanishing point detection for road detection," in *Proc. IEEE CVPR*, Miami, FL, June 2009, pp. 96–103.

[8] S. Zhou, J. Xi, J. Gong, G. Xiong, and H. Chen, "A novel lane detection based on geometrical model and gabor filter," in *in Proc. Intelligent Vehicles Symposium*, San Diego, CA, June 2010, pp. 59–64.

[9] Y. Wang, E. K. Teoh, and D. Shen, "Lane detection and tracking using b-snake," *J. Image and Vision Comp.*, vol. 22, no. 4, pp. 269–280, April 2004.

[10] X. Ren and J. Malik, "Learning a classification model for segmentation," in *in Proc. ICCV*, Nice, France, 2003, pp. 10–17.

[11] S. Paris and F. Durand, "A topological approach to hierarchical segmentation using mean shift," in *in Proc. IEEE CVPR*, Minneapolis, MN, June 2007, pp. 1–8.

[12] M. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *in Proc. IEEE CVPR*, Colorado, June 2011, pp. 2097–2104.

[13] C. Chefd'Hotel and A. Sebbane, "Random walk and front propagation on watershed adjacency graphs for multilabel image segmentation," in *in Proc. IEEE ICCV*, Rio de Janeiro, Oct. 2007, pp. 1–7.

[14] S. Yu and J. Shi, "Multiclass spectral clustering," in *in Proc. IEEE ICCV*, vol. 1, Nice, France, Oct. 2003, pp. 313–319.

[15] A. Levinshtein, A. Stere, K. Kutulakos, D. Fleet, S. Dickinsonl, and K. Siddiqi, "urbopixels: Fast superpixels using geometric flows," *J. IEEE TPAMI*, vol. 31, no. 12, pp. 2290–2297, Dec. 2009.

[16] M. Helala, "Quantitative image search based on feature integration," in *master thesis*. Cairo, Egypt, Dec. 2010.

[17] L. Figueiredo, "Adaptive sampling of parametric curves," in *in Graphics Gems V*. Academic Press, 1995, pp. 173–178.

[18] J. Han and M. Kamber, "Data mining : concepts and techniques." San Francisco : Morgan Kaufmann Publisher, 2001.

[19] W. Bolstad, "Introduction to bayesian statistics." John Willy & Sons Publisher, 2007.

[20] K. Kyungnam, T. Chalidabhongse, D. Harwood, and L. Davis, "Background modeling and subtraction by codebook construction," in *in Proc. Int. Conf. ICIP*, vol. 5, Singapore, Oct. 2004, pp. 3061–3064.

[21] R. Baeza-Yates and B. Ribeiro-Neto, "Modern information retrieval." New York: ACM Press, Addison-Wesley, 1999.