Video-based Road Detection Using Evolving GMMs and Region Enhancement

Ioannis Kaloskampis *†‡ Mahmud Abdulla Mohammad †‡ Yulia Hicks ‡

Abstract

In this article we propose a new online method for road detection which uses as input video captured by a single video camera. Our method consists of two stages. In the first stage we build a statistical road model using training data and in the second we detect the road area in new video frames. The road model is based on video segmentation with evolving GMMs. After the initial detection of the road area, the result is improved with post-processing, which caters for inaccuracies in the detected road region caused by shadows, illuminations and unusual road shapes. Experimental results for the established, publicly available CamVid dataset show that the proposed method achieves high accuracy in road detection.

1. Introduction

Recent advances in autonomous vehicles have resulted in intelligent automobiles which sense the environment using a variety of sensors, such as GPS, radars and cameras. By processing the information acquired by these sensors, they are capable of determining the travel route and identifying important scene objects, such as traffic signs and obstacles. An important problem in the design of autonomous vehicles is road detection, as it provides an important information cue to sense the environment and eases applications such as path finding and planning, object tracking, anomaly detection and situation assessment.

Vision-based approaches typically use colour as the main feature for road detection, as texture is dependent on reliable shape patterns parallel to the road direction and increases computational cost [4]. However, environmental challenges such as colour variation, shadows and lighting conditions pose problems to colour-based road detectors [17], hence additional information is required to improve the detection accuracy.

*This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/K014307/1 and the MOD University Defence Research Collaboration in Signal Processing.

†Contributed equally.

‡The authors are with Cardiff University, UK. E-mails: {kaloskampisi, hicksya, mohammadma}@cardiff.ac.uk

For roads that are designed in accordance with design guidelines and standards, road structure can be used as a cue to improve the system's performance [9]. The drawback of such approaches is that they cannot operate reliably in unstructured road scenes.

Typical road geometries can also be exploited to enhance the performance of road detection, *e.g.* left turn, straight, and t-like junction as in [3]. Such approaches lose accuracy in certain circimstances, e.g. in cluttered scenes [4].

Other methods employ a combination of techniques to enhance the performacnce of road detection. Colour plane fusion and convolutional neural networks were combined in [2]. This work assumes that the bottom part of the video frame captures the road region, which cannot be always guaranteed in practice [5].

Prior knowledge regarding the road shape has also been used to improve the performance of road detection, *e.g.* [10]. Such methods may face problems when the road region in the input frame is significantly different than the models in the learned road shape database.

The shortcomings of vision-based approaches have led to the inclusion of additional sensors alongside traditional cameras, such as stereo cameras [8], LIDAR [15], and GPS/GIS [4]. Although the additional sensors improve the road detection accuracy, due to high cost, complex installation and high computational load they are currently not close to becoming standard for vehicles. Moreover, certain sensors may additionally suffer from interference problems.

In this article, we propose a novel on-line method for road detection which utilises only vision-based features. We rely on vision-based features as the visual sensor is gradually becoming standard for modern vehicles, with an increasing number of vehicles being equipped with dashboard cameras. Our method first builds a road model using training data and then uses this model for road detection. The road model is based on video segmentation. After the initial detection of the road area, the result is improved with post-processing, which caters for inaccuracies in the detected road region caused by shadows, illuminations and unusual road shapes.

We assess our method on the CamVid dataset [7] and show that it achieves high accuracy in road detection.

The remainder of this article is structured as follows: the process of building the road model is presented in Section

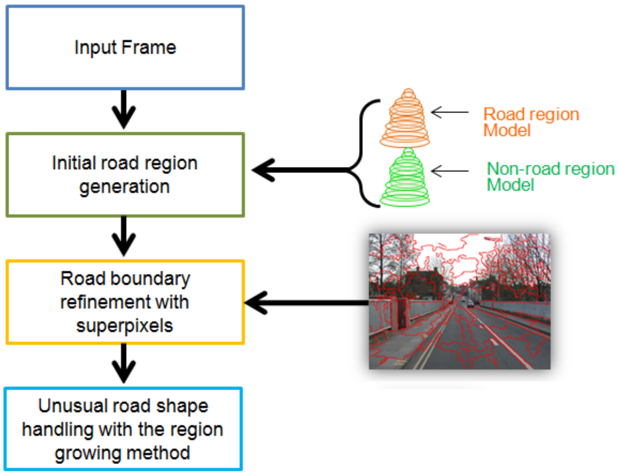

Figure 1. The road detection pipeline.

2. The road detection pipeline is described in Section 3. Experiments and results are presented in Section 4 and the article is concluded in Section 5.

2. Building the road model

This section describes the process of building the off-line model of our framework, which will be used in the later on-line road detection phase (Section 3). The proposed model discriminates between road and non-road regions and is learned automatically from training data.

Our model is based on the EvoGMM video segmentation algorithm [11]. It acquires colour and spatial information from training video frames, which will enable discrimination between road and non-road regions. As we would like our method to work as close as possible to real time, this information should be handled efficiently. For each frame in the training set we extract features from each of its pixels. We then build a GMM using the features of the frame; after building the GMM, all features extracted from the pixels are discarded. Each frame is therefore represented by a GMM rather than its pixel features, saving a significant amount of computer storage space and memory. Note that EvoGMM estimates the GMM's number of components automatically.

To build a model for the region class road or non-road, we could simply concatenate the components of the GMMs corresponding to this specific region class from all the training frames. However, this would lead to a complex model with a large number of overlapping components. This problem is handled by the EvoGMM algorithm by merging similar components after the initial concatenation, using a modified version of the EM algorithm. Thus, a compact model is produced with no overlapping components. This compactness reduces the computational cost of the generation of the initial road region.

To build the statistical discriminative model, we use a set S of m training image sequences:

$$S = \{I^{(1)}, I^{(2)}, ..., I^{(m)}\} \tag{1}$$

where $I^{(n)}, n \in \{1, 2, ..., m\}$ an image sequence.

Then, visual features from every frame of each image sequence in the set S are extracted. Following [11] this is achieved by representing each pixel in each frame with a five-dimensional vector which includes the pixel's colour value in the Lab space and the pixel's spatial coordinates. We denote by $F^{(n)}$ the feature representation of $I^{(n)}$ and define the set of feature representations S' as:

$$S' = \{F^{(1)}, F^{(2)}, ..., F^{(m)}\} \tag{2}$$

Then the EvoGMM algorithm from [11] is applied to all elements of S'. This algorithm converts each frame to a GMM by grouping the pixels on the basis of their feature similarities in the five-dimensional feature space. Each Gaussian component of the GMM corresponds to a homogeneous region of the frame. After that, all GMMs are labeled manually into two classes, road and non-road. Finally, all resulting GMMs illustrating the same class are concatenated into a unified model. The model M_i, for class i, is given by the equation:

$$M_i = \{L_{ik}\}_{k \in \{1,2,..,N_i\}} \tag{3}$$

where L_{ik} is the k^{th} Gaussian in M_i, and N_i is the total number of Gaussians in M_i.

3. Road detection

In this section we describe the process of detecting the road area, using the model of Section 2. The pipeline of the method is shown in Fig. 1. Note that the road detection process is fully on-line.

3.1. Initial road region generation

To detect the road region in an input frame f, we first build the frame's model, M_f using the EvoGMM algorithm and then correlate its components to the GMMs of the trained model M_i representing road and non-road areas obtained during the training stage (Section 2). The model M_f at input frame f is a GMM, given by the equation:

$$M_f = \{U_{fj}\}_{j \in \{1,2,..,N_f\}} \tag{4}$$

where U_{fj} is the j^{th} Gaussian in U_f and N_f the total number of Gaussians in the model M_f.

The next step is to estimate the distance between each Gaussian U_f from the segmented frame f and the models of road and non road regions, using the Bhattacharyya distance [6]. We denote by $B\left(U_{fj}, L_{ik}\right)$ the Bhattacharyya distance between the j^{th} Gaussian of the GMM of f and the k^{th} Gaussian of the model M_i. Then we find the minimum distance, β_{fij}, between the j^{th} Gaussian in frame f, U_{fj} and the Gaussians in the model M_i:

$$\beta_{fij} = min\left\{B\left(U_{fj}, L_{ik}\right)\right\} \tag{5}$$

The classification of Gaussians to road and non-road is given by the equation:

$$D_{fj} = \arg\min_{i} \|\beta_{fij}\| \tag{6}$$

where D_{fj} is the classification outcome for the j^{th} Gaussian of frame f which can be road or non-road.

Having classified the Gaussians of input frame f, the initial road region, I^G, is generated, by merging the regions corresponding to the Gaussians classified as road region.

$$I^G_{xy} = \begin{cases} 1, & \{xy\} \in \{R_{fj}\}_{(D_{fj}=r)} \\ 0, & \{xy\} \in \{R_{fj}\}_{(D_{fj}=n)} \end{cases} \tag{7}$$

where I^G_{xy} is the output for the xy^{th} pixel of the frame, R_{fj} is the segmented region in frame f corresponding to the j^{th} Gaussian of frame f, r and n are the road and non-road region classes, respectively.

3.2. Road boundary refinement with superpixels

The initial road boundaries are rough. We refine and smoothen them with entropy rate superpixels [14]. We first oversegment each video frame to θ superpixels, and then merge the superpixels that have an overlap rate with the initial road region. We denote the overlap rate between superpixel q and the initial road region by $\alpha_q, q \in \{1, 2, .., \theta\}$. The merging process is given by the equation:

$$I^S_q = \begin{cases} 1, & \alpha_q \geq \tau \\ 0, & \alpha_q < \tau \end{cases} \tag{8}$$

where I^S_q is the value of superpixel q and $\tau \in (0, 1)$ is the threshould which controls the region merging.

3.3. Region enhancement with region growing

Scene complexity, illuminations, light direction and different levels of shadows might have a negative impact on the detected road area. This can be improved by region growing [1], a pixel-based classification method, which utilises the homogeneity between neighbourhood pixels to classify them into regions. Although region growing can suffer from shadow and illumination effects, since both problems can be mitigated at earlier stages of our pipeline, its use at this point is justified. In summary, initial seeds are first selected and the growing process commences with the comparison between the initial seed point and its pixel neighbours on the basis of homogeneity to determine whether they belong to the growing region.

Method	Accuracy (%)
Ours	93.1
Liu and He [12]	88.9
Brostow et al. [7]	89.5
Liu et al. [13]	92.4
Sturgess et al. [16]	95.3

Table 1. Results in terms of pixel-wise percentage accuracy for the CamVid dataset.

Pixels within the previously estimated region I^S are selected as initial seeds. All grown regions are unified into a single region, which is the final output of our method.

4. Experiments and results

The proposed method is evaluated on the CamVid dataset, which includes daytime and dusk sequences, captured from the driver's perspective. The dataset offers additional motion and 3D structure cues; we disregard these cues as we rely on visual information. We split the dataset into the training and testing subsets following [7]. In terms of implementation details, we use $\theta = 100$ superpixels and set $\tau = 0.5$ to control superpixel merging.

Quantitative results, given in Table 4, show that our method achieves high accuracy in road detection. The method from [16] achieves the highest score but makes use of the aforementioned additional cues.

Fig. 2 provides qualitative results of our method for four challenging cases: row (i) shows an unusual road shape; rows (ii) and (iii) illustrate shadow effects; in row (iv) cyclists occlude the road region and illumination varies near the centre of the frame. Our method detects the road with high accuracy.

5. Conclusion and future work

We proposed a new online method for road detection which uses as input video captured by a single video camera. Our method first builds a statistical road model using training data and then detects the road area in new frames using this model. The road model is based on video segmentation with evolving GMMs. After the initial detection of the road area, the result is improved with post-processing, which handles problems caused by shadows, illuminations and unusual road shapes.

The proposed method achieves high accuracy in road detection for the CamVid dataset. There is room for optimisation: future work will investigate the inclusion of additional cues such as depth information and GIS in our framework to improve its performance.

References

[1] R. Adams and L. Bischof. Seeded region growing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*,

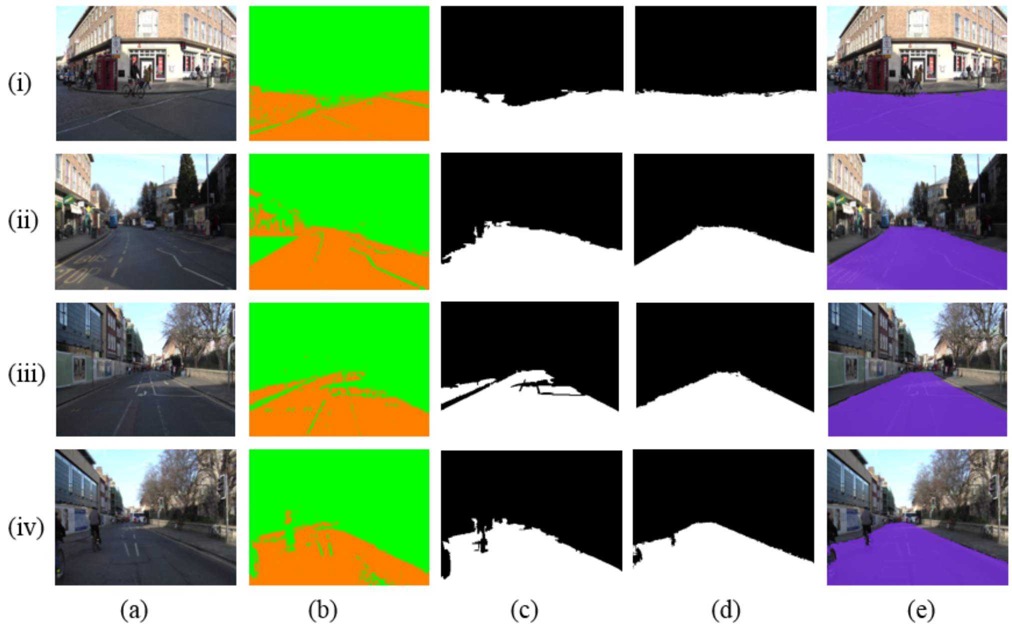

Figure 2. Qualitative evaluation of our method for the CamVid dataset: (a) original; (b) initial road region; (c) superpixel boundary refinement; (d) region growing; (g) superimposition of the output of our method on the original frame.

16(6):641–647, Jun 1994. 3

[2] J. M. Alvarez, T. Gevers, Y. LeCun, and A. M. Lopez. Road scene segmentation from a single image. In *Computer Vision–ECCV 2012*, pages 376–389. Springer, 2012. 1

[3] J. M. Alvarez, T. Gevers, and A. M. López. Vision-based road detection using road models. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 2073–2076. IEEE, 2009. 1

[4] J. M. Álvarez, A. M. López, T. Gevers, and F. Lumbreras. Combining priors, appearance, and context for road detection. *Intelligent Transportation Systems, IEEE Transactions on*, 15(3):1168–1178, 2014. 1

[5] J. M. Alvarez, M. Salzmann, and N. Barnes. Data-driven road detection. In *Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*, pages 1134–1141. IEEE, 2014. 1

[6] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35:99–109, 1943. 2

[7] G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla. Segmentation and recognition using structure from motion point clouds. In *Computer Vision–ECCV 2008*, pages 44–57. Springer, 2008. 1, 3

[8] C. Guo, S. Mita, and D. McAllester. Robust road detection and tracking in challenging scenarios based on markov random fields with unsupervised learning. *Intelligent Transportation Systems, IEEE Transactions on*, 13(3):1338–1354, Sept 2012. 1

[9] J. Han, D. Kim, M. Lee, and M. Sunwoo. Enhanced road boundary and obstacle detection using a downward-looking lidar sensor. *Vehicular Technology, IEEE Transactions on*, 61(3):971–985, March 2012. 1

[10] Z. He, T. Wu, Z. Xiao, and H. He. Robust road detection from a single image using road shape prior. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 2757–2761, Sept 2013. 1

[11] I. Kaloskampis and Y. Hicks. Estimating adaptive coefficients of evolving GMMs for online video segmentation. In *6th IEEE International Symposium on Communications, Control and Signal Processing (ISCCSP)*, pages 513–516, May 2014. 2

[12] B. Liu and X. He. Multiclass semantic video segmentation with object-level active inference. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4286–4294, June 2015. 3

[13] B. Liu, X. He, and S. Gould. Multi-class semantic video segmentation with exemplar-based object reasoning. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 1014–1021, Jan 2015. 3

[14] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa. Entropy rate superpixel segmentation. In *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pages 2097–2104, 2011. 3

[15] Z. Prochazka. Improved algorithm for road region segmentation based on sequential monte-carlo estimation. *Computer Science and Information Technology*, page 171183, 2014. 1

[16] P. Sturgess, K. Alahari, L. Ladicky, and P. H. Torr. Combining appearance and structure from motion features for road scene understanding. In *BMVC*, volume 1, page 6, 2009. 3

[17] Q. Wang, J. Fang, and Y. Yuan. Adaptive road detection via context-aware label transfer. *Neurocomputing*, 158:174 – 183, 2015. 1