DETECTION OF DRIVABLE CORRIDORS FOR OFF-ROAD AUTONOMOUS NAVIGATION

Ara V. Nefian

Application Research Lab
Intel Corp., Santa Clara, CA
ara.nefian@intel.com

Gary R. Bradski

Application Research Lab
Intel Corp., Santa Clara, CA
gary.bradski@intel.com

ABSTRACT

This paper describes a hierarchical Bayesian network used for segmenting desert images and detecting off road drivable corridors for autonomous navigation. Unlike the embedded hidden Markov model the Bayesian network presented in this paper can successfully account for natural dependencies between neighboring pixels in both image dimensions making it more suitable for a larger class of images. The method described here was developed within the Stanford racing team that won the DARPA Grand Challenge 2005 after driving over 130 miles autonomously in the Nevada desert.

***Index Terms*—** Mobile robot motion-planning, image segmentation, hidden Markov models

1. INTRODUCTION

Intelligent vehicles able to navigate autonomously in various environments have numerous applications ranging from indoor robotics to unmanned commercial and military vehicles and interplanetary exploration. Remarkable progress towards achieving autonomous navigation in off road corridors was achieved recently during the DARPA Grand Challenge Race [1] won by the Stanford racing team [2]. The work presented in this paper grows out of the race team work where the authors were part of the team that developed a generative road segmentation model [3]. We had wanted a way to give the model road boundary and horizon semantics but were not able to complete the work before the code freeze deadline. This paper completes this intended work. In order to robustly and accurately determine road and off road corridors, the overall navigation system (Figure 1) incorporates information from several sensors including GPS, laser and cameras. Laser sensors can very accurately determine the road at short range but power considerations limit our laser range to 20 meters. The role of the vision system is to sample the near visual scene where the lasers identify drivable terrain and use that region as a "seed" to segment drivable areas in the rest of the image. A Gaussian mixture model over the color space [3] was used to do this scene segmentation. In this approach the parameters of the model are trained from the indicated nearby "seed" region and the remaining pixels in the image are assigned to drivable surface based on their color space distance to the learned model. Alternatively the image can be segmented using a two class model which better discriminates between the road and non-road regions [4]. The above segmentation method worked for winning the race, but drivable areas could show up in non-sensible side areas, on hillsides etc. In this paper we introduce a Bayesian image model that incorporates a set of geometrical and smoothness constraints that yield more global "semantics" for navigation: road/corridor, off-road and sky.

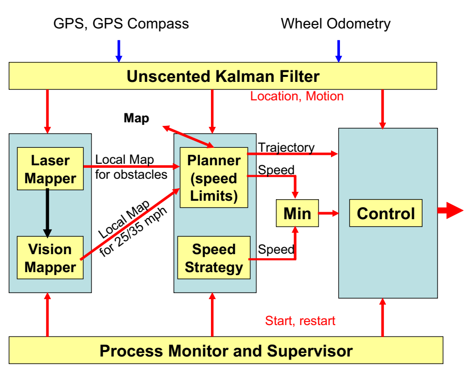
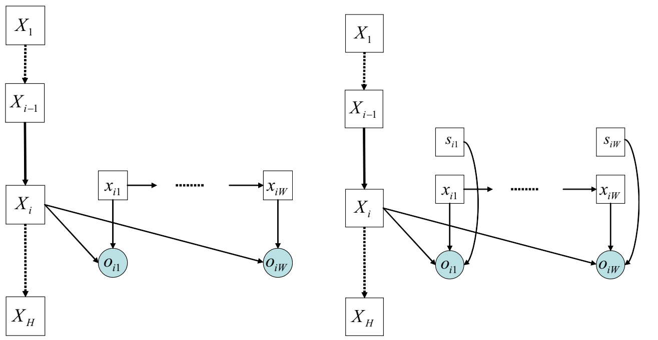


Fig. 1. The overall software system or the robot. This paper focuses on the vision modules that feed into the vision mapper.

2. THE IMAGE MODEL

This section describes a hierarchical Bayesian network used to model images captured in the desert. The image model is used to cluster image regions such as the road, sky or sides of the road. These regions have a consistent position relative to the other regions in the same image while displaying a large variability in size and shape. For example the sky region is always above the land region (road or sides of the road). Similarly there is a consistent position of the road region relative to the sides of the road. On the other hand the shape of the region and its size depends on the horizontal orientation of the road, the inclination of the road (sky will appear a larger region while looking down hill) or the landscape around the road (high mountains, or hills will shape differently the sky region than plateaus). One of the most successful statistical models used to describe a specific category of images with similar properties is the embedded hidden Markov model (EHMM). This model was used in character recognition [5] and face recognition [6]. The success of this model for faces relies on its ability to describe the relative position of the significant facial features such as eyes nose or mouth while allowing for a larger flexibility than template based approaches. The EHMM (Figure 2 a) is a Bayesian network [7] that describes an array of observations $o_{ij}, i = 1, \dots H, j = 1, \dots W$ with W columns and H rows. This network approximates the fully connected two-dimensional hidden Markov model(HMM) for which the inference is an NP complete problem [5]. The EHMM can be seen as hidden Markov model in which each hidden node X_i, $i = 1, \dots, H$ associated with a row in an image is a parent of another HMM describing the sequence of observations in that row. The discrete hidden nodes $x_{ij}, j = 1 \dots W$ describe the observed nodes o_{ij} along each column of the i^{th} row.



(a) (b)

Fig. 2. An EHMM (a) and an extended EHMM with shadow node (b).

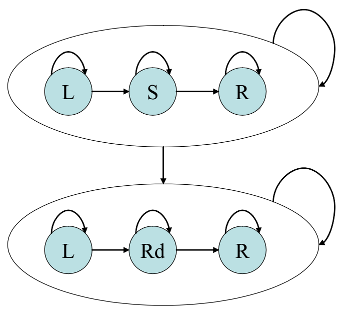


Fig. 3. The state diagram of the image model. The transparent ovals represent the values of the nodes X_i and the shaded circles represent the values of the nodes x_{ij}.

Figure 3 illustrates the state diagram for an EHMM. In our experiments, the nodes X_i have binary values such as "sky row" or "land row" denoting that the global assignment of the i^{th} row. The x_{ij} hidden nodes have three values each describing the region ("left side-L", "road-Rd", "right side-R" for "land row" or "left side-L", "sky-S", "right side-R" for "sky row") to which the observation o_{ij} is assigned. The constraints imposed by the state diagram segment the image in compact regions and preserve the relative positions of a region relative to the other regions in the image.

The flexibility of the EHMM in modelling images can be enhanced by introducing an additional set of binary hidden nodes s_{ij} associated with each observed node to determine if the pixel is observed in shadow or direct light (Figure 2 b).

Both the above models describe dependencies between consecutive rows and consecutive columns in the same row but do not model directly dependencies between a hidden node x_{ij} and the hidden node associated with the pixel in the same column of the previous row $x_{i-1,j}$. To overcome this problem we introduced the image model illustrated in Figure 4. In this model each row i is described by a "duration" HMM [8] with parameters d_{ik}. These parameters represent the number of pixels assigned to state k in row i or the "duration" of state k in row i.

$$\sum_k d_{ik} = W$$

The hidden nodes D_i associated with row i describe the set of parameters d_{ik}. Formally, the image model can be described by the following

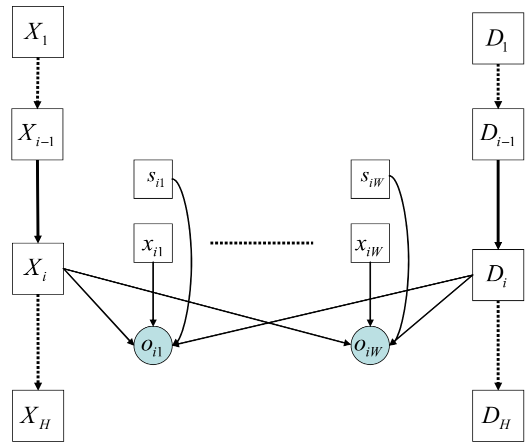


Fig. 4. A Bayesian network for image modelling.

$$P(\mathbf{O}) = \sum_{\mathbf{X},\mathbf{D},\mathbf{x},\mathbf{s}} P(X_1)P(D_1)\prod_i P(D_i|D_{i-1})P(X_i|X_{i-1}) \cdot \\ \cdot P(\mathbf{o}_i|\mathbf{x}_i,\mathbf{s}_i,X_i,D_i)$$

where

$$P(D_i|D_{i-1}) = \prod_k P(d_{ik}|d_{i-1k})$$

and

$$P(d_{i,k}|d_{i-1,k}) = N(d_{i,k}, d_{i-1,k}, \sigma_k^2)$$

is Gaussian density function with mean d_{ik} and variance σ_k. Within each row the observation likelihood is given by

$$P(\mathbf{o}_i|\mathbf{x}_i,\mathbf{s}_i,X_i,D_i) = \prod_{j=1}^{d_{i,1}} P(o_{ij}|x_{ij}=k, s_{ij}=m, X_i) \cdot \\ \cdot \prod_{k=1}^{K-1}\prod_{j=d_{i,k}}^{d_{i,k+1}} P(o_{ij}|x_{ij}=k, s_{ij}=m, X_i)$$

where

$$P(o_{ij}|x_{ij}=k, s_{ij}=m, X_i) = N(o_{ij}, \mu_{km}, \mathbf{C}_{km})$$

is a Gaussian density function with mean μ_{km} and diagonal covariance matrix $\mathbf{C}_{km}$. The conditional probabilities $P(X_i|X_{i-1})$ are determined by the state diagram in Figure 3.

3. LEARNING THE IMAGE MODEL

The parameters of the model described in the previous section can be learned from a large set of images using the expectation-maximization algorithm [7] or can be approximated using a modified version of the segmental K-means algorithm [8]. The segmental K-means shown in Figure 5 initializes the model parameters from an initial segmentation and iteratively improves the model parameters and the segmentation results until convergence or a fixed number of iterations is reached. In order to increase the computation speed, maintain a set of constraints on the segmentation result, and learn robustly the model parameters from one image, the number of the trained model parameters described in the previous section must be reduced. Hence, in this paper the state transition probabilities described in Figure 3, and the variance of the duration parameters σ_k are fixed. Furthermore, the values of the shadow nodes s_{ij} are kept fixed during the learning of the remaining parameters of the model. The values of the shadow nodes, which determine the pixels in shadow, are determined by a simple comparison with a fixed threshold in the color space. In the initial segmentation the hidden nodes X_i and

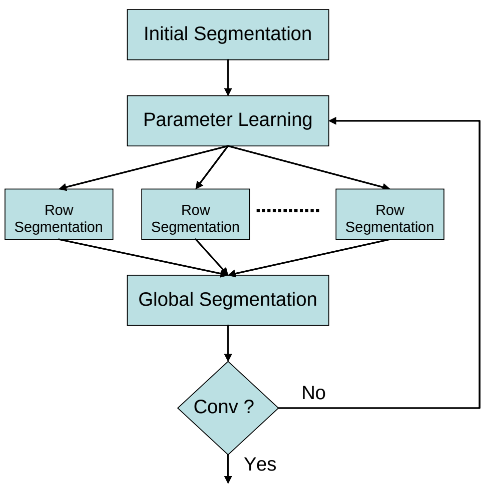


Fig. 5. Image segmentation

x_{ij} can be assigned to a value at random or based on some a priori information. A random initialization can often lead to a poor image segmentation. The performance can be improved by using several non-uniform image segmentation and picking the one with the highest image likelihood, but this leads to a higher computational complexity. An alternative solution is to start with shadow and sky detection since these regions are efficiently segmented using color information. In our system the segmentation of the remaining part of the image is initialized using the information provided by the laser sensors. The short range road area provided by the laser is mapped to a trapezoidal region in the image plane. The region between the sky and the road detected by laser (where the laser information is less reliable) is segmented uniformly in three horizontal regions according to the state diagram.

Following the initialization, the model parameters μ_{km} and $\mathbf{C}_{km}$ are learned from the current image segmentation. Next, the image is segmented by finding the best values for the set of hidden states of the image model that fits the observations extracted from the image. With the values of the shadow nodes assumed known and fixed the segmentation problem is stated as finding

$$\{\mathbf{X}, \mathbf{D}, \mathbf{x}\} = \mathbf{argmax}_{\{\mathbf{X},\mathbf{D},\mathbf{x}\}} \mathbf{P}(\mathbf{O}|\mathbf{X}, \mathbf{D}, \mathbf{x}).$$

The image segmentation starts with the computation in each row of the observation likelihood $P(\mathbf{o}_i|\mathbf{x}_i, X_i, D_i)$ for a set of discrete values of D_i. In our experimental results we selected 100 possible values describing a set of possible row segmentations. The best sequence of values for the nodes X_i and D_i that maximize 1 is then decoded using the Viterbi algorithm [8].

The resulting segmentation can be used to re-learn the model parameters in a process that can run iteratively until convergence or a fixed number of iterations is reached.

4. EXPERIMENTAL RESULTS

The models presented in this paper were tested on several sequences consisting of several hours of video captured in the Nevada desert. Figure 6 and Figure 7 illustrate typical road segmentation results on the same frames of three different video sequences using the embedded HMM with shadow nodes in Figure 2b. and the Bayesian network shown in Figure 4 respectively. Although the EHMM with shadow nodes has a higher run time speed, the use of the Bayesian model in Figure 4 significantly increases the accuracy of the system and operates at 10 frames/second on images of size 320×240 pixels on a 1.6 GHz Pentium M processor. The use of the hidden nodes D_i in Figure 4 introduces a set of smoothness and geometrical constraints that model better the perspective effects in an image and the natural shapes of various regions. This advances the “drivable blobs” segmentation used in the race to now indicate drivable corridors together with horizon boundaries.

5. CONCLUSIONS

This paper describes a statistical model for segmenting of images captured in the desert and reliably detecting of off road drivable corridors. The model introduced in this paper is a hierarchical Bayesian network that describes the data at fine grain along one dimension and a course grain along the second dimension. In addition the model introduces a set of geometrical and smoothness constraints that make it particularly suitable for the type of images discussed in this paper, and improves significantly the image segmentation results obtained using the embedded HMM.

6. REFERENCES

[1] “DARPA Grand Challenge,” http://www.darpa.mil/grandchallenge/.

[2] “Stanford racing team,” http://www.stanfordracing.org/.

[3] G.R. Bradski, A. Kaehler, and V. Pisarevski, “Learning-based computer vision with Intel’s open source computer vision library,” *Intel Technology Journal*, vol. 9, no. 2, May 2006.

[4] B. Davies and R. Lienhart, “Using CART to segment road images,” in *Proceedings of the SPIE Conference on Visual Computing and Image Processing*, January 2006.

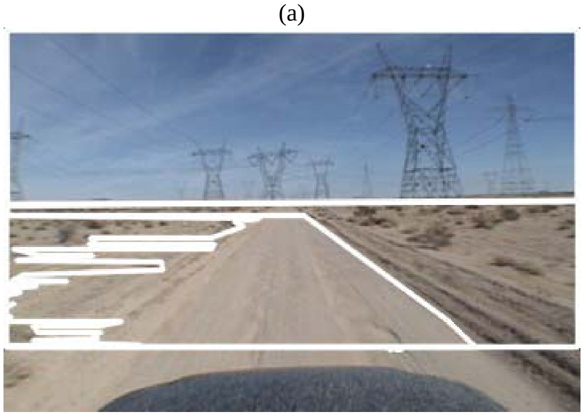
[5] O.E. Agazzi and S. Kuo, “Hidden Markov Models based optical character recognition in presence of deterministic transformations,” *Pattern Recognition*, vol. 26, no. 12, pp. 1813–1826, 1993.

[6] A. V Nefian and M. H. Hayes, “An embedded HMM approach for face detection and recognition,” in *Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing*, March 1999, vol. 6, pp. 3553–3556.

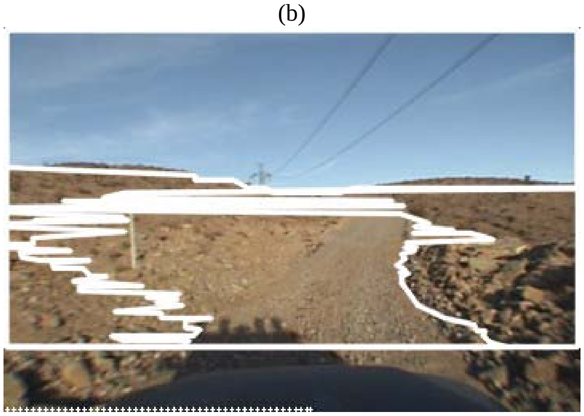
[7] Finn V. Jensen, *An Introduction to Bayesian Networks*, UCL Press Limited, London, UK, 1998.

[8] L. Rabiner and B.H. Huang, *Fundamentals of Speech Recognition*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

(a)



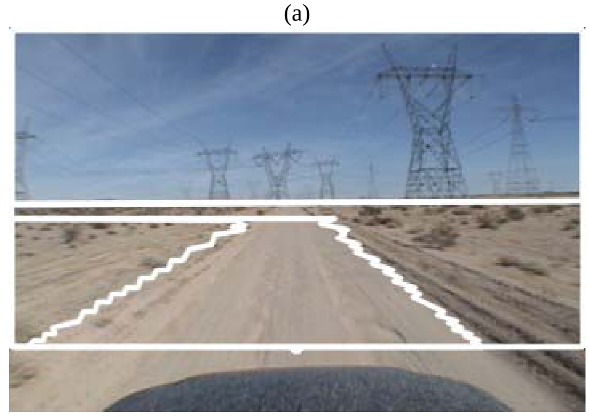
(b)



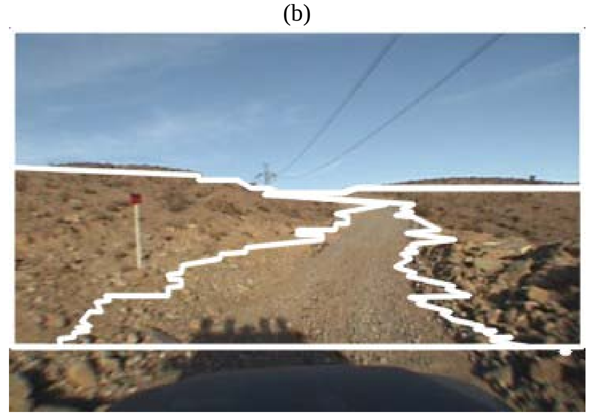
(c)

Fig. 6. Road segmentation results using the EHMM in Figure 2b.

(a)



(b)



(c)

Fig. 7. Road segmentation results using the Bayesian network in Figure 4.