Learning Background and Shadow Appearance with 3-D Vehicle Models [*]

Matthew J. Leotta   Joseph L. Mundy
Division of Engineering, Brown University
Providence, RI, USA
{mleotta,mundy}@lems.brown.edu

Abstract

This paper presents a novel algorithm for simultaneous background appearance modeling and coarse-scale vehicle recognition in traffic surveillance applications. 3-d mesh models representing a small set of vehicle classes are used to the hypothesize image segmentations into background, shadow, and vehicle regions. The algorithm optimizes vehicle class and motion parameters to best agree with a Hidden Markov Model for the image appearance. The best hypothesis, combined with image data, is used to adapt the parameters of the appearance model. Experiments on real video show that an appearance model trained in this way performs almost as well as one trained using manually segmented images.

1 Introduction

Background modeling techniques are common in video surveillance to detect moving foreground objects. Often, this problem is confounded by the appearance of shadows cast by the moving foreground objects. While it is common to add an image-based model for shadow, this model does not exploit knowledge of shadow formation in the 3-d world. A correct 3-d model can accurately predict the image location of foreground and shadow pixels. However, in vehicle surveillance, the class of each observed vehicle is *a priori* unknown. Thus, the recognition and segmentation problems are intertwined.

This paper addresses both detection and recognition simultaneously. The focus is on training and adapting an image appearance model using hypothesized image segmentations. These hypotheses are generated from 3-d mesh models of vehicles, an illumination model, and an image projection model. The algorithm optimizes the vehicle class and motion parameters to best agree with the image data and appearance model. This appearance model is a Hidden Markov Model (HMM) at each pixel. The emission densities are adaptive Gaussian Mixture Models (GMM), which help make the appearance model robust to errors in the segmentation hypothesis. Figure 1 gives a diagrammatic overview of the process. The main contribution of this paper is application of 3-d models to predict the location of vehicle shadow in video in order to train an appearance model.

[*] This material is based upon work supported by BAE Systems under Contract No. 04062-6896, which is supported by United States Army Aviation and Missile Command under Contract No. W31P4Q-04-C-R270. The authors would like to thank Matt Antone and Marcus Graly at BAE Systems for supplying data.

BMVC 2006 doi:10.5244/C.20.67

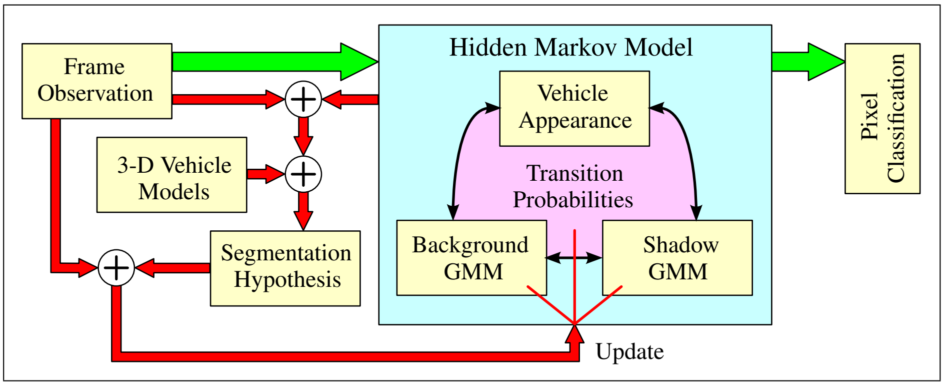

Figure 1: Flowchart overview of the algorithm. The wide green path (top) shows data flow for pixel classification (detection). The narrow red paths (bottom left) show data flow for model parameter adaptation. The ⊕ symbols indicate the combination of data.

The remainder of this paper is organized as follows. Section 2 provides some background and related work. Section 3 describes the appearance model and the algorithm for estimating and updating the model parameters. Section 4 describes the 3-d vehicle models, the illumination model, and the algorithm for vehicle recognition and alignment. Section 5 shows experimentally that the appearance model trained by hypothesized segmentation performs comparably with that trained by manual segmentation. Section 6 summarizes and alludes to future work.

2 Related Work

Several techniques for background modeling appear in the literature. Perhaps the most commonly applied method is based on the work of Stauffer and Grimson using adaptive Gaussian Mixture Models [13]. The authors model the color at each pixel by a weighted sum of Gaussian distributions. Each new observation updates the model, and foreground/background segmentation is determined in real-time. While this algorithm is powerful, it is slow to converge. Other authors have attempted to address this deficiency [4, 10, 6]. Different update equations are used during initialization in [4]. A unified set of equations with a time varying gain are used in [10]. In [6], a different time varying gain is attributed to each component to improve convergence at all stages of the algorithm.

The shadows cast from moving foreground objects onto the background are often undesirably detected as foreground. This problem has been addressed in several ways (See [11] for a survey and comparison of approaches). Several authors apply a color transformation model [8, 4, 15]. They assume that shadow appears in a range of known color transformations of the background. Martel-Brisson and Zaccarin [7] use such a color model to migrate certain mixture components into a second GMM for shadow.

An alternate approach to background modeling is the Hidden Markov Model. In this approach, each pixel is modeled by a state machine with emission and transition probabilities (see [2] for a tutorial on HMMs). Stenger *et al.* [14] use a topology free HMM in which states are split to best model the background. This approach does not explicitly model shadow. A more relevant approach is the three-state HMM of Rittscher *et al.* [12] to model background, foreground, and shadow. The emission distributions are Gaussian for background and shadow, and uniform for foreground. They use the Baum-Welch algorithm to learn the model parameters. The authors note that, without additional constraints, this method fails to distinguish shadow from dark vehicles and results in poor parameter estimation. Wang *et al.* [15] extend this approach by creating an adaptive HMM that is trained with Baum-Welch and then maintained with an incremental version of the EM algorithm. They also apply a Markov Random Field to enforce spatial continuity.

Related to the 3-d aspect of this paper is [5]. This approach fits parameterized 3-d polyhedral models to image edges for tracking and recognition. It also includes an illumination model for the shadow casting. Predicted shadow edges are also matched to edges in the image. The illumination direction is set manually. By contrast, the approach in this paper: aligns fixed 3-d models to an appearance model for background, shadow, and vehicle; uses the alignment to update the appearance model; and automatically sets the illumination direction.

3 Appearance Model

The appearance model used at each pixel is a HMM in RGB color space. The HMM has a fixed topology of three states representing background, shadow, and vehicle. This model differs slightly from previous work in that the emission probability distribution is a GMM instead of a single Gaussian distribution. However, the focus of this work is the novel approach by which the model parameters are learned.

3.1 Hidden Markov Model

In the HMM, each pixel has a hidden state variable $s_t \in \{1,2,3\}$ (for background, shadow, and vehicle) and a transition matrix $\mathbf{A}$ such that a_{ij} is $\mathrm{P}(s_t = j|s_{t-1} = i)$. The emission probability distributions, denoted $\mathrm{P}(\mathbf{o}_t|s_t)$ for observation $\mathbf{o}_t$ at time t, vary for each state.

The distribution over all possible vehicle colors and intensities is difficult to estimate. Thus, it is modeled as a uniform distribution

$$\mathrm{P}(\mathbf{o}_t|s_t = 3) = \left(\frac{1}{R_{\max} - R_{\min}}\right)\left(\frac{1}{G_{\max} - G_{\min}}\right)\left(\frac{1}{B_{\max} - B_{\min}}\right) \qquad (1)$$

where $[R_{\min}, R_{\max}]$ is the range of possible values in the red channel, and likewise for the green and blue channels.

The emission distributions for the background and shadow states are each modeled with an adaptive GMM as defined in Section 3.2. The equation for these emission probabilities is (4). While the equations for each state are the same, the parameters are learned using different sets of observations (See Section 3.3).

For pixel classification, the Forward algorithm is used to propagate normalized state probabilities, $\gamma_{j,t}$, forward in time given the data.

$$\gamma_{j,t} \equiv \mathrm{P}(s_t = j|\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_t) = \frac{\mathrm{P}(\mathbf{o}_t|s_t = j)\sum_{i=1}^{3} a_{ij}\gamma_{i,t-1}}{\sum_{k=1}^{3}\mathrm{P}(\mathbf{o}_t|s_t = k)\sum_{i=1}^{3} a_{ik}\gamma_{i,t-1}} \qquad (2)$$

The pixel classifier selects the most probable state at each time t given the history of observations, (*i.e.* $s_t = \arg\max_j \gamma_{j,t}$).

3.2 Adaptive Gaussian Mixture Model

The adaptive GMM models a multimodal background while robustly accounting for foreground observations. To generalize the application of the adaptive GMM, and to avoid confusion with the HMM states, the terms *background* and *foreground* will be replaced with *inliers* and *outliers*. For the background GMM, background observations are inliers and both vehicle and shadow are outliers. Similarly in the context of the shadow GMM, the inliers are the shadow observations and the outliers are the background and vehicle.

The mixture is a weighted sum ($\sum_k \omega_k N(\mu_k, \Sigma_k)$) of normal distributions N, with mean μ_k, covariance Σ_k, and mixing parameters ω_k. For efficiency the color channels are assumed independent, resulting in a diagonal covariance matrix. The components are sorted by $\omega^3/\det(\Sigma)$ to bring the inlier components to the top of the list. At each time t, the top B_t components are assumed to model inliers. B_t is determined by setting a threshold T on the minimum portion of the data that is considered to be inliers:

$$B_t = \arg\min_b \left(\sum_{k=1}^{b} \omega_{k,t} > T \right) \tag{3}$$

where 0.7 is used for T. The probability that the mixture generates an observation $\mathbf{o}_t$ at time t given that $\mathbf{o}_t$ is an inlier is:

$$\mathrm{P}(\mathbf{o}_t | s_t = \text{inlier}) = \frac{\sum_{k=1}^{B_t} \omega_{k,t} N(\mathbf{o}_t, \mu_{k,t}, \Sigma_{k,t})}{\sum_{k=1}^{B_t} \omega_{k,t}} \tag{4}$$

The equations for updating the mixture are similar those proposed in [6]. Let Θ_t be the set of all mixture parameters at time t. Using the "winner-take-all" strategy, let

$$\mathrm{P}(k|\mathbf{o}_t, \Theta_t) = \begin{cases} 1 & \text{if } k = \arg\min_i \left((\mathbf{o}_t - \mu_{i,t})^T \Sigma_{i,t}^{-1} (\mathbf{o}_t - \mu_{i,t}) < D \right) \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Thus the unique matching component k is the top component such that the observation is within a Mahalanobis distance D of the mean. D is taken to be 2.5. The "winner-take-all" strategy requires only one component to be updated at each time and is used for efficiency. Given the match, the learning rates are adjusted as follows:

$$\alpha_t = \max\left(\frac{1}{n_t}, \frac{1}{L} \right) \qquad \rho_{k,t} = \mathrm{P}(k|\mathbf{o}_t, \Theta_t) \left(\frac{1-\alpha_t}{\eta_{k,t}} + \alpha_t \right) \tag{6}$$

where n_t is the number of total observations, $\eta_{k,t}$ is the number of observations for each component (incremented when k is a match), α_t is the mixture learning rate, $\rho_{k,t}$ are the component learning rates, and L is the window size. α_t is lower bounded by $1/L$ to switch to exponential forgetting after L frames. L is taken to be 300. $\rho_{k,t}$ decays to α_t as more observations are made of component k (see [6] for details). The mixture parameters are updated according to:

$$\omega_{k,t} = (1-\alpha_t)\omega_{k,t-1} + \alpha_t \mathrm{P}(k|\mathbf{o}_t, \Theta_t) \tag{7}$$

$$\mu_{k,t} = (1-\rho_{k,t})\mu_{k,t-1} + \rho_{k,t}\mathbf{o}_t \tag{8}$$

$$\Sigma_{k,t} = (1-\rho_{k,t}) \left(\Sigma_{k,t-1} + \rho_{k,t}(\mathbf{o}_t - \mu_{k,t-1}) \circ (\mathbf{o}_t - \mu_{k,t-1}) \right) \tag{9}$$

where $\circ$ is the entrywise vector product with the result arranged as a diagonal matrix. Note that (9) differs from other author's equations [13, 4, 10, 6, 7] by a factor of $1-\rho_{k,t}$ in the data term. This formulation correctly makes (9) a recursive estimator of $\Sigma_{k,t}$ (if $\alpha_t = 0$). In the limit, when $\eta_{k,t}$ is large, this equation approaches that used by most other authors.

3.3 Learning Model Parameters

Baum-Welch, an expectation-maximization algorithm, is often used to learn HMM parameters from a fixed set of unlabeled data. However, if the training data are labeled, the complexity of an iterative approach is unnecessary. In the labeled case, only the expectation step remains, which permits a recursive formulation. The recursive estimate of a_{ij}, with exponential forgetting, is

$$a_{ij,t} = (1-\alpha_t)a_{ij,t-1} + \alpha_t \delta_t(i,j) \qquad \delta_t(i,j) = \begin{cases} a_{ij,t-1} & \text{if } s_{t-1} \neq i \\ 1 & \text{if } s_{t-1} = i \text{ and } s_t = j \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where α_t is defined in (6) and $a_{ij,0} = 1/3$. Section 3.2 shows how the emission GMM distributions are adapted given the data allocated for each state. Before any samples have been observed, these distributions are modeled with a uniform distribution.

One source of labeled data is manual segmentation of the video frames. Since manual labor is expensive only a small interval of video can be annotated. Such training data results in near optimal model parameters initially, but lack of adaptation causes the error rate to increase over time. The appearance of the background and shadow change drastically throughout the day. The position and size of the shadow also change relative to the vehicle. These changes quickly render manually labeled data obsolete.

In the proposed approach, 3-d models are used to hypothesize the labels (see Section 4). The resulting segmentation is automatic, but may misclassify some pixels because of alignment inaccuracy and shape differences between the vehicle and 3-d model. However, the segmentation is good enough to train the HMM. More pixels are classified correctly than not, so misclassified pixels become the outliers in the adaptive GMMs.

The shadow and vehicle appear infrequently enough be outliers in the background GMM without the need for segmentation. Thus, the background GMM is trained on all the data. This generalization allows the algorithm to bootstrap using the background GMM alone (initially transition and the shadow distributions are uniform). Since cast shadows appear less frequently, the shadow GMM requires segmentation to adapt.

4 Image Segmentation from 3-D Models

Vehicle 3-d mesh models are used to predict image segmentations for appearance model training. This requires the vehicle models, an illumination model, and an image projection model. The vehicle models are triangular polygon meshes each with 268 triangles and 136 vertices (see Figure 2). Three models represent the coarse shapes of common passenger vehicles: sedan, pickup truck, and minivan/SUV. They were manually constructed to scale using highly detailed 3-d CAD models as guides. Enough detail is provided in these meshes to accurately represent the vehicle shape without overburdening the algorithm with excessively large mesh models.

4.1 Illumination and Projection

The image projection model is the standard perspective pinhole camera. It is also possible to include a lens distortion model, though that is ignored here for clarity. Let $\mathbf{P}$ be the 3×4 camera matrix that maps homogeneous 3-d world points $\mathbf{X} = [\ x \quad y \quad z \quad 1\ ]^T$ into homogeneous 2D image points $\mathbf{x} = [\ \lambda u \quad \lambda v \quad \lambda\ ]^T$ as $\mathbf{x} = \mathbf{PX}$ (refer to [3]). The full

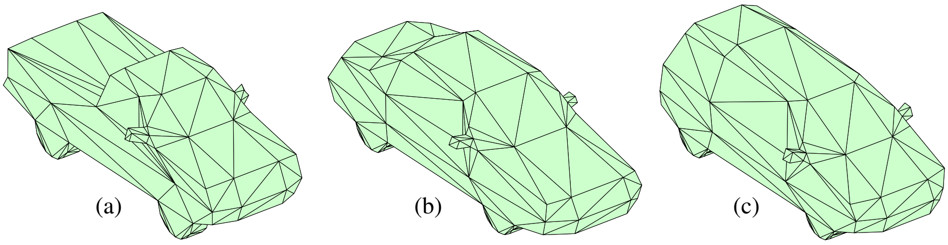

Figure 2: 3-d vehicle models: (a) pickup truck, (b) sedan, (c) minivan/SUV

metric calibration of the camera is performed to compute $\mathbf{P}$ in advance. The location and orientation of the road is also assumed known in the world coordinate system of the camera. These assumptions are reasonable for fixed camera vehicle surveillance.

The illumination model maps all points in the world to their shadow positions on the ground plane. Let $\mathbf{M_s}$ be a 4×4 homogeneous transformation matrix that maps a world point $\mathbf{X}$ to a shadow point $\mathbf{X_s} = \mathbf{M_s X}$. In combination with the projection model, the shadow of a point is imaged as $\mathbf{x_s} = \mathbf{P X_s} = \mathbf{P M_s X}$. A shadow camera matrix, $\mathbf{P_s} = \mathbf{P M_s}$, defines the mapping of a world point directly to the image of it's shadow. The matrix $\mathbf{M_s}$ for shadows cast by the sun is a function of time, date, latitude, and longitude (See Appendix A). It is also reasonable to assume these parameters are known.

The models above can produce an image segmentation, given a vehicle model positioned in the world, in the following way: Initially, label all pixels as background. Project the vertices of the mesh into the image with $\mathbf{P_s}$. Raster scan the front-facing triangles and label these pixels as shadow. Repeat the projection and raster scanning using $\mathbf{P}$ and label these pixels as vehicle.

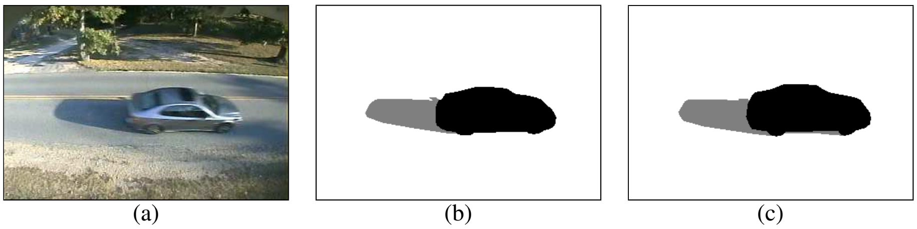

Figure 3: Pixel segmentations: (a) input frame, (b) manual segmentation, (c) segmentation hypothesis from 3-d model alignment

4.2 Model Alignment and Recognition

The goal of model alignment is to estimate the model parameters θ which maximize the posterior distribution $P(\theta | O)$ over the set of vehicle observations O. In this work, θ consists of an initial position on the ground plane, a velocity along the orientation of the road, and a choice of 3-d vehicle model. It is assumed that vehicles travel parallel to the road direction with constant velocity while in the view of the camera.

For simplicity it is assumed that only a single vehicle appears at a time. The subset of frames, $\mathscr{T}$, containing each vehicle is identified with a threshold on the number of outlier pixels detected by the background GMM. Define $\mathscr{X}$ to be the set of all pixel locations.

Let $\mathbf{o}_{t,x}$ denote an observation and $s_{t,x}$ denote a segmentation label of pixel location $x \in \mathscr{X}$ at time $t \in \mathscr{T}$. Finally, designate the set of observations $O = \{\mathbf{o}_{t,x} : \forall t \in \mathscr{T}, x \in \mathscr{X}\}$ and the set of labels $S = \{s_{t,x} : \forall t \in \mathscr{T}, x \in \mathscr{X}\}$.

From Bayes rule $\mathrm{P}(\theta|O) \propto \mathrm{P}(O|\theta)\mathrm{P}(\theta)$ and $\mathrm{P}(\theta)$ is assumed to be a uniform distribution (though it would be trivial to include a more general prior distribution on the vehicle models and trajectories). Furthermore, O is assumed to be conditionally independent of θ given S (*i.e.* $\theta \rightarrow S \rightarrow O$ forms a Markov chain), and S is a deterministic function of θ, so $\mathrm{P}(O|\theta) = \mathrm{P}(O|S(\theta))$. Finally, pixel observations are assumed to be conditionally independent in space and time given the segmentation, so $\mathrm{P}(O|S(\theta)) = \prod_{x \in X, t \in T} \mathrm{P}(o_{t,x}|s_{t,x}(\theta))$. Putting this all together, the maximum likelihood parameter estimate is

$$\hat{\theta} = \arg\max_{\theta} \prod_{x \in X, t \in T} \mathrm{P}\left(o_{t,x}|s_{t,x}(\theta)\right) = \arg\min_{\theta} \sum_{x \in X, t \in T} -log\left(\mathrm{P}\left(o_{t,x}|s_{t,x}(\theta)\right)\right) \qquad (11)$$

Finding $\hat{\theta}$ is a minimization problem. The error surface is not smooth due to discrete pixel labeling, which causes gradient-based minimization techniques to fail. The Nelder-Mead simplex method [9] performs better under these circumstances and is used here. While this method only guarantees a local minimum, using good initial conditions usually leads to solution near the optimum. The average vehicle speed and the known position of the road provide good initial motion parameters. Since each vehicle model may have different optimal motion parameters, optimization is performed for each vehicle keeping the vehicle model fixed. The vehicle class and corresponding motion parameters that maximize the probability in (11) are selected. Figure 3 shows a resulting segmentation.

Note that $-log(\mathrm{P}(o_{t,x}|s_{t,x}(\theta)))$ can be precomputed for each t and x and for each state $s_{t,x}$ independently of θ using the emission distributions for each state. The parameters θ simply determine which of the three precomputed values of $-log(\mathrm{P}(o_{t,x}|s_{t,x}(\theta)))$ to include in the summation for each pixel in each frame.

5 Experimental Results

To test the performance of this algorithm, 4000 frames of 352 × 240 resolution surveillance video containing 14 vehicles in 560 of the frames were manually segmented. The manual segmentation is taken to be ground truth. Four variations of the appearance model, described below, were tested by training on the first 3000 frames and testing on the last 1000 frames.

HMM-3D The complete algorithm as described in this paper
HMM-GT The appearance model described in Section 3 trained on ground truth
GMM-3D Emission GMMs only, uniform transition probs., trained with 3-d models
GMM-GT Emission GMMs only, uniform transition probs., trained with ground truth

A comparison of the HMM-GT and HMM-3D experiments indicates nearly identical pixel classification results. Therefore, training the appearance model with hypothesized labels is a reasonable substitute for training with manually determined ground truth labels. The GMM-3D and GMM-GT experiments set the transitions probabilities in the HMM to 1/3 to demonstrate the advantage of the temporal constraint. In each experiment, the background GMM was allowed to adapt in the testing phase, but all other model parameters were held fixed. The confusion matrix for each model is given in Table 1. Sample segmentations from the same experiments are shown in Figure 4.

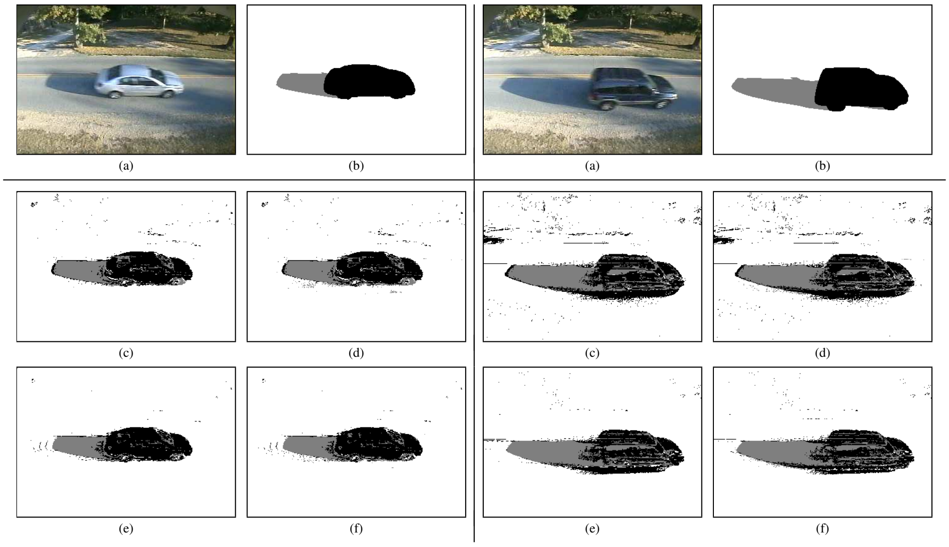

Figure 4: Segmentation results for two different frames: (a) input frame, (b) manual segmentation, (c) segmentation using GMM-GT (d) segmentation using GMM-3D (e) segmentation using HMM-GT (f) segmentation using HMM-3D

			Predicted by GMM-GT			**Predicted by GMM-3D**		
		total	back.	shadow	vehicle	back.	shadow	vehicle
Actual	back.	83621226	99.55%	0.02%	0.43%	99.50%	0.08%	0.42%
	shadow	299690	1.10%	85.58%	13.32%	0.98%	86.14%	12.88%
	vehicle	559084	6.40%	18.00%	75.60%	5.98%	19.01%	75.01%

			Predicted by HMM-GT			**Predicted by HMM-3D**		
		total	back.	shadow	vehicle	back.	shadow	vehicle
Actual	back.	83621226	99.85%	0.02%	0.13%	99.84%	0.04%	0.12%
	shadow	299690	1.70%	86.64%	11.66%	1.50%	87.13%	11.37%
	vehicle	559084	6.30%	14.89%	78.81%	5.32%	15.95%	78.73%

Table 1: Normalized confusion matrices for each appearance model tested on the same data with $1000 \times 352 \times 240$ pixels. The background class is abbreviated as "back.".

To show the need for continuous model adaption, a second experiment was run on 500 frames of data occurring about 23 minutes after the previous data. In this experiment, HMM-3D was adapted continuously using 3-d model alignment, and HMM-GT was trained with the same 3000 frames of ground truth segmentations. To be fair, since segmentation is not required to adapt the background GMM, the background GMM was also adapted continuously for HMM-GT. However, the shadow GMM and transition probabilities were held fixed in HMM-GT after the training period. The resulting confusion matrices are given in Table 2 and sample segmentations are shown in Figure 5.

Only 2% of shadow pixels are misclassified as vehicle using HMM-3D compared to 57% using unadapted HMM-GT. The error rate of shadow misclassified as background is similar in both models and primarily due to background already shadowed by trees. Discrimination between shadow and vehicle is ultimately more important for vehicle recognition. Almost all of the shadow is misclassified when the appearance model is not updated.

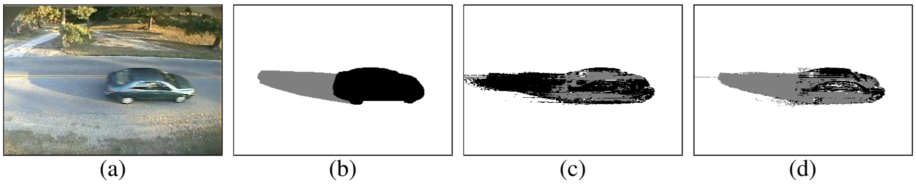

Figure 5: Segmentation results at later time: (a) input frame, (b) manual segmentation, (c) segmentation using HMM-GT (not updated), (d) segmentation using HMM-3D (updated)

			Predicted by HMM-3D			Predicted by HMM-GT		
		total	back.	shadow	vehicle	back.	shadow	vehicle
Actual	back.	41389799	99.90%	0.07%	0.03%	99.80%	0.01%	0.19%
	shadow	385305	35.53%	62.35%	2.12%	38.19%	4.74%	57.07%
	vehicle	464896	15.87%	31.37%	52.76%	13.04%	18.49%	68.47%

Table 2: Normalized confusion matrices for HMM-3D and HMM-GT tested 23 minutes after training HMM-GT (500 × 352 × 240 pixels).

Clearly, adaptation of all model parameters is a critical factor in the success of this HMM classifier.

6 Summary

The results have shown that 3-d vehicle models can be used to train and adapt an appearance model for background, shadow, and vehicle while performing simplistic vehicle recognition. A HMM (with GMM emission distributions) trained in this way performs nearly as well as one trained on manually segmented images. However, the appearance model used has room for improvement.

Future work might investigate more advanced appearance models with spatial continuity and better time continuity constraints. Possibly also extending this approach to handle multiple vehicles simultaneously, and investigating vehicle recognition rates in detail with a larger set of vehicle models or a single parametric model. Finally, this algorithm is well-suited to take advantage of specialized 3-d graphics hardware, providing a likely route to explore on the path to real-time implementation.

References

[1] D. Carruthers, C. Uloth, and G. G. Roy. An evaluation of formulae for solar declination and the equation of time. Research Report RR17, School of Architecture, University of Western Australia, January 1990.

[2] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*, chapter 3.10, pages 128–138. John Wiley and Sons, Inc., second edition, 2001.

[3] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, first edition, 2000.

[4] P. KaewTraKulPong and R. Bowden. An improved adaptive background mixture model for real-time tracking with shadow detection. In *2nd European workshop on Advanced Video Based Surveillance Systems*, September 2001.

[5] D. Koller. Moving object recognition and classification based on recursive shape parameter estimation. In *Israel Conference on Artificial Intelligence, Computer Vision*, pages 27–28, December 1993.

[6] D.-S. Lee. Effective gaussian mixture learning for video background subtraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:827–832, May 2005.

[7] N. Martel-Brisson and A. Zaccarin. Moving cast shadow detection from a gaussian mixture shadow model. In *Int. Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 643–648. IEEE Computer Society, 2005.

[8] I. Mikic, P. Cosman, G. Kogut, and M. Trivedi. Moving shadow and object detection in traffic scenes. In *15th International Conference on Pattern Recognition*, volume 1, pages 321–324, September 2000.

[9] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.

[10] P. Power and J. Schoonees. Understanding background mixture models for foreground segmentation. In *Imaging and Vision Computing New Zealand.*

[11] A. Prati, I. Mikic, M. Trivedi, and R. Cucchiara. Detecting moving shadows: Algorithms and evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):918–923, July 2003.

[12] J. Rittscher, J. Kato, S. Joga, and A. Blake. A probabilistic background model for tracking. In *ECCV '00: Proceedings of the 6th European Conference on Computer Vision-Part II*, pages 336–350, London, UK, 2000. Springer-Verlag.

[13] C. Stauffer and W. Grimson. Adaptive background mixture models for real-time tracking. In *Int. Conf. Computer Vision and Pattern Recognition*, volume 2, pages 246–252, 1999.

[14] B. Stenger, V. Ramesh, N. Paragios, F. Coetzee, and J. Buhmann. Topology free hidden markov models: Application to background modeling. In *IEEE International Conference on Computer Vision (ICCV)*, volume 1, pages 294–301, Vancouver, Canada, 2001.

[15] D. Wang, T. Feng, H.-Y. Shum, and S. Ma. A novel probability model for background maintenance and subtraction. In *The 15th International Conference on Vision Interface*, 2002.

A Shadow Casting Equations

Equations 12 and 13 show the formulae for solar declination δ_s and the equation of time ET as given by Carruthers *et al.* [1]. Solar declination is the altitude angle between the sun position and the equatorial plane. The equation of time gives the difference between clock time and solar time in seconds caused by the earth's elliptical orbit and tilted axis. Here $t = \frac{2\pi J}{366}$, $t' = t + 4.8718$, and $J \in [1, 366]$ is the day of the year.

$$\begin{aligned}\delta_s = {} & 0.322003 - 22.9711\cos(t) - 0.357898\cos(2t) - 0.14398\cos(3t) \qquad (12)\\ & + 3.946380\sin(t) + 0.019334\sin(2t) + 0.059280\sin(3t)\\ ET = {} & 5.0323 - 430.847\cos(t') + 12.5024\cos(2t') + 18.25\cos(3t') \qquad (13)\\ & - 100.976\sin(t') + 595.275\sin(2t') + 3.6858\sin(3t') - 12.47\sin(4t')\end{aligned}$$

The hour angle ξ, the angle between the sun position and solar noon, is computed in Equation 14. The altitude γ_s and azimuth α_s of the sun are solved for using Equations 15 and 16. In these equations, θ and ϕ are longitude and latitude in radians respectively and UTC is the coordinated universal time of day in hours.

$$\xi = \theta + \frac{\pi}{12}\left(\frac{ET}{3600} + UTC\right) - \pi \qquad (14)$$

$$\sin(\gamma_s) = \sin(\delta_s)\sin(\phi) + \cos(\delta_s)\cos(\phi)\cos(\xi) \qquad (15)$$

$$\cos(\alpha_s) = \frac{\sin(\delta_s)\cos(\phi) - \cos(\delta_s)\sin(\phi)\cos(\xi)}{\cos(\delta_s)} \qquad (16)$$

The matrix $\mathbf{M_s}$ in Equation 17 is a 3-d affine homography that maps a homogeneous world point $\mathbf{X} = [\begin{array}{cccc} x & y & z & 1\end{array}]^T$ to a shadow point $\mathbf{X_s} = \mathbf{M_s X}$ on the ground plane $z = 0$. This applies in a world coordinate system where positive x is north, positive y is west, and positive z is up.

$$\mathbf{M_s} = \begin{bmatrix} 1 & 0 & -x_s & 0\\ 0 & 1 & -y_s & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1\end{bmatrix} \quad x_s = \frac{\cos(-\alpha_s)}{\tan(\gamma_s)} \quad y_s = \frac{\sin(-\alpha_s)}{\tan(\gamma_s)} \qquad (17)$$