Evaluating Color Representations for On-line Road Detection

Jose M. Alvarez
NICTA
7 London Circuit, 2601
Canberra, Australia
jose.alvarez@nicta.com.au

Theo Gevers
University of Amsterdam
Science Park 904, 1098 XH
Amsterdam, The Netherlands
th.gevers@uva.nl

Antonio M. Lopez
Computer Vision Center
Edifici O, Campus UAB, 08193
Bellaterra, Spain
antonio@cvc.uab.es

Abstract

Detecting traversable road areas ahead a moving vehicle is a key process for modern autonomous driving systems. Most existing algorithms use color to classify pixels as road or background. These algorithms reduce the effect of lighting variations and weather conditions by exploiting the discriminant/invariant properties of different color representations. However, up to date, no comparison between these representations have been conducted. Therefore, in this paper, we perform an evaluation of existing color representations for road detection. More specifically, we focus on color planes derived from RGB data and their most common combinations. The evaluation is done on a set of 7000 road images acquired using an on-board camera in different real-driving situations.

1. Introduction

The main goal of vision-based road detection is detecting traversable road areas ahead of an ego-vehicle using an on-board camera. Road detection is a key component in modern autonomous driving solving specific tasks such as road following, car collision avoidance and lane keeping [1]. Road detection using a monocular color camera is challenging since algorithms must deal with continuously changing background, the presence of different objects like vehicles and pedestrian, different road types (urban, highways, country side) and varying ambient illumination and weather conditions. Moreover, these algorithms have the additional constraint of real-time execution.

A common road detection approach consists of analyzing road homogeneity to group pixels into road and background areas by training a classifier based on road and non-road examples. However, the large diversity of non-road areas and the lack of annotated datasets has motivated the development of on-line detection algorithms [2, 3, 4, 5]. The core of these algorithms is a single class classifier trained on a small set of positive examples collected from the bottom part of the image being analyzed. Therefore, these algorithms do not require examples of the background class and usually rely on the assumption that the vehicle acquiring the images is on the road. In addition, these algorithms tend to represent pixel values using simple (fast) features such as color [2, 6] or texture [4, 7] to reach the real-time constrain required in most of the applications. Color offers many advantages over texture since texture varies with the distance to the camera due to the perspective effect. Color provides powerful information about the road despite the shape of the road or the perspective effect commonly present in road images. However, using color cues is a challenging task due to the varying photometric conditions in the acquisition process (*e.g.*, illumination variations or different weather conditions). Different color planes exhibiting different invariant properties have been used to reduce the influence of these photometric variations. Color spaces derived from RGB data that have proved to be, to a certain extent, robust to lighting variations are HSV [2, 8, 9], normalized RGB [3], CIE-Lab [10] or their combination [11, 12]. More recently, color constancy has also been used in [5] to minimize the influence of lighting variations at the cost of being camera dependent and requiring specific calibration. Algorithms embed these color representations in complex systems that include inference methods (CRF), post-processing steps and constraints such as temporal coherence [3, 13] or road shape restrictions [2]. Therefore, is difficult to compare and, more importantly, obscures the analysis of different color representations to deal with illumination changes within the road detection context.

In this paper, we focus on evaluating different color representations for on-line road detection. More precisely, in Sect. 2 we devise a simple two stages algorithm consisting of a color conversion and a single class classifier performing in still images. The output of the classifier is then thresholded to provide a binary mask defining road areas. In Sect. 3, a large number of experiments is conducted to evaluate all possible combinations between nineteen color representations and five different classifiers using a dataset

of 7000 manually annotated images acquired using a camera mounted on a vehicle driving in real-world scenarios. Finally, in Sect. 4, conclusions are drawn.

2. On-line Road Detection Algorithm

In this section, the road detection algorithm used for the evaluation. The algorithm performs frame by frame and consists of two stages: color conversion and pixel classification (Fig. 1). RGB pixel values are converted to the selected color representation and used as input to the classification stage. This classifier considers only target samples to learn the model used for inference. The model is learned based on road samples collected from the bottom part of the image. Thus, the algorithm is based on the assumption that the bottom region of the image belongs to the road class. This area usually corresponds to a distance of about four meters ahead the camera and is a reasonably assumption when the car is on the road. The output of the classifier is a road likelihood $\mathcal{L}$ showing the potential of each pixel of belonging to the road class. This likelihood ranges from 0 to 1 where the higher the likelihood the higher the probability of being a road pixel. State of the art algorithms build upon this road likelihood to obtain the traversable road area incorporating post-processing steps such as connected components [5], temporal coherence [3, 13], shape restrictions [2] or even conditional random fields leading to robustified algorithms. Therefore, for fair comparison, we use a simple threshold to assign pixel labels: if $\mathcal{L}_i > \tau$, the i-th pixel is labeled with a road label. Otherwise, a background label is assigned.

In the next subsections, we introduce the color representations and single-class classifiers used in the evaluation.

2.1. Color Conversion

The first stage is the color conversion process aiming at characterizing RGB pixel values using color features. Algorithms have exploited the different invariant/sensitive properties of existing color spaces to reduce the influence of lighting variations in outdoor scenes. In this paper, we analyze the performance of five different color spaces: RGB and four others spaces derived from direct transformations (Table 1): normalized RGB, opponent color space, HSV and CIE-Lab. Thus, these color spaces are device independent and do not require specific calibration. Each of these color planes have different properties as summarized in Table 2. For instance, using R, G, or B to represent pixels provides high discriminative power but low invariance to shadows and lighting variations. On the other hand, using hue or saturation to represent pixels provides higher invariance but lower discriminative power. Interestingly, three of the color spaces consider separate the luminance and the chrominance into different signals. For instance, in the HSV color space, the V channel provides discrim-
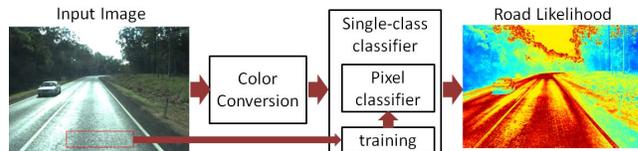

Figure 1. Road detection algorithm used in the experiments.

inative power while H and S components provide different types of invariance. Similarly, the opponent color space comprises the luminance component and two channels providing color information. Thus, these color representations are uncorrelated and provide diversified pixel information that can be efficiently exploited by the classifier.

2.2. Single-Class Classifier

In this paper, we consider two non-parametric and three parametric classifiers. The former are the nearest neighbor and a model-based classifier and the latter rely on learning Gaussians from training data.

Nearest neighbor. This is the first and most basic non-parametric classifier. The road likelihood is the minimum distance between each pixel and the training data. In this paper, we consider the minimum squared Euclidean distance over the training set to obtain the road likelihood for each pixel. Note that this is slightly different from other implementations using circular distances over specific color planes.

Model-based. This is a non-parametric classifier that uses a likelihood measure to approximate the conditional probability of having a road pixel given a pixel value. This probability distribution is estimated for each image using the training samples. In particular, we use the normalized histogram of the training samples. Therefore, the road likelihood as $\mathcal{L}_i = p(x_i)$, where $p(\cdot)$ is the normalized histogram. The higher the likelihood value, the higher the potential of being a road pixel.

Single Gaussian (G). This classifier models road training samples using an unique Gaussian distribution. Therefore, in this case, the road likelihood for the i-th pixel is obtained as $\mathcal{L}_i = G(x, \mu_r, \sigma_r)$, where x_i is the pixel value and μ_r, σ_r are the parameters of the Gaussian distribution learned using the training samples. In practice, to avoid numerical instabilities, we do not estimate the density. Instead, we use the Mahalanobis distance as follows: $\mathcal{L}_i = (x_i - \mu_r)^T \Sigma_r^{-1} (x_i - \mu_r)$, where Σ_r is the covariance matrix estimated using the training set.

Robustified Gaussian. The previous classifier is sensitive to outliers and noise in the training samples since the parameters are estimated without robust statistics. In our case, these outliers are long tails in the distribution mainly due to lighting conditions or different road textures. Therefore, the second parametric classifier is based on a single Gaussian where the parameters are learned using robust

Table 1

$nrngnb$	Opponent Color Space
$nr = \frac{R}{R+G+B}$ $ng = \frac{G}{R+G+B}$ $nb = \frac{B}{R+G+B}$	$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$

HSV

$$\begin{pmatrix} V \\ V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

$$H = \arctan \frac{V_2}{V_1} \qquad S = \sqrt{V_1^2 + V_2^2}$$

CIE Lab

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.490 & 0.310 & 0.200 \\ 0.177 & 0.812 & 0.011 \\ 0.000 & 0.010 & 0.990 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

$$L = 116(\tfrac{Y}{Y_0})^{\frac{1}{3}} - 16$$
$$a = 500\left[(\tfrac{X}{X_0})^{\frac{1}{3}} - (\tfrac{Y}{Y_0})^{\frac{1}{3}}\right] \quad b = 200\left[(\tfrac{Y}{Y_0})^{\frac{1}{3}} - (\tfrac{Z}{Z_0})^{\frac{1}{3}}\right]$$

X_0, Y_0 and Z_0 are the coordinates of a reference white point.

Table 1. Derivation of the different color spaces used to characterize road pixels.

Table 2

	R	G	B	nr	ng	O_1	O_2	L,V	ab	H	S
Global Illumination changes	-	-	-	-	-	+	+	-	+	+	+
Discriminative power	+	+	-	-	+	-	-	-	+	-	-
Distance to the camera	+	+	+	+	+	+	+	+	+	+	+
Strong shadows	-	+	-	-	-	+	+	-	+	+	+
Highlights	-	-	-	-	-	+	+	-	+	+	+
Road type and shape	+	+	+	+	+	+	+	+	+	+	+

Table 2. Properties of different color representations from a theoretical point of view. Robustness to each property is indicated with '+' and weakness with '-'.

statistics. More precisely, training samples are weighted according to their proximity to the real mean value. Distant samples are down weighted to obtain a more robust estimate. Then, the road likelihood is obtained as in the previous one.

Mixture of Gaussians (MoG). Previous classifier has the inconvenient of modeling a single Gaussian distribution. This can be a constraint when dealing with shadows and lighting variations. Therefore, the last classifier models the set of training samples using a mixture of N Gaussians and thus, creates a more flexible description of the road class. In this case, the road likelihood is given by $\mathcal{L}_i = \sum_{n=1}^{N} P_n exp^{-(x-\mu_n)^T \Sigma_n^{-1}(x_i-\mu_n)}$, where μ_n and Σ_n are the parameters of the different Gaussians involved and P_n is the weight assigned to the n-th Gaussian. In this paper, we optimize these parameters using the EM algorithm and we also evaluate different values of N.

3. Experiments

In this section, we present our results on the evaluation of color planes for road detection. For a comprehensive evaluation, experiments are conducted on color plane individually (13 color planes) and the most common combinations of color planes such as HSV, $nrng$, HS, O_1O_2 and RGB combined with ten different classifiers: five non-parametric and five parametric ones. The non-parametric set consists in the nearest neighbor classifier and four instances of the model-based one. Two of these instances use directly the training samples from the bottom part of the road to build the normalized-histogram with 64 and 100 bins, and the

other two instances extend the training set with noisy samples. Extending the training set with synthetic samples is a common practice in the machine learning community to improve the robustness of the algorithms [14]. In particular, we duplicate the samples and adding zero mean and $30/256$ standard deviation noise to half of it. Then, two different histogram configurations are considered: 100 and 64 bins. Using different number of bins to build the histogram enables the stability analysis of variations of this parameter. The parametric set consists of the Gaussian and robust Gaussian classifier and three instances of MoG classifier: $N = 2$, $N = 4$ and optimizing N based on the training set.

The classifiers are trained for each image using only road samples collected from a rectangular area (200×65 pixels) at the bottom part of the image yielding to a training set consisting of 13000 road pixels. Note that this area is suited for right driving situations and it is not extremely large. Therefore, the variability within the training set is not always significant and may not represent all the road areas in the image.

3.1. Road Dataset

The dataset consists of twenty road sequences comprising thousands of images acquired at different days, different daytime (daybreak, morning, noon and afternoon), different weather conditions (sunny, cloudy, rainy) and for three different scenarios (urban like, highways and secondary structured roads). Each of these scenarios exhibits different challenges: highways contain tunnels with low light conditions, crowed scenes or extreme lighting variations; secondary roads contain lighting variations, shadows and the presence of other vehicles in the scene; finally, urban scenes contain crossings, intersections and the presence of multiple objects (*i.e.*, other vehicles and pedestrians). Ground-truth is generated by a single experienced user providing manual segmentations of a subset of 7000 randomly selected images (Fig. 2). To our knowledge, this is the largest set of manually annotated road images today.
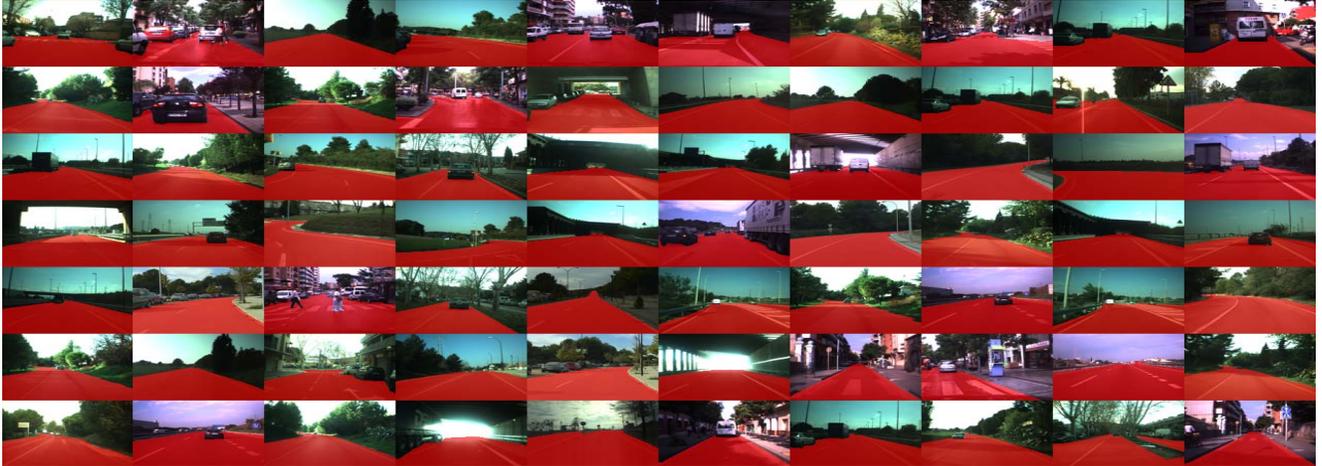
Figure 2. Example of annotated images in the dataset. Best viewed in color.

3.2. Evaluation Measures

Quantitative evaluations are provided using average ROC curves [15] on the pixel-wise comparison between ground-truth and results obtained binarizing the road likelihood $\mathcal{L}$ (Sect. 2) with different threshold values. ROC curves represent the trade-off between true positive rate TPR and false positive rate FPR [16]. These two measures provide different insight into the performance of the algorithm. True positive rate ($TPR = \frac{TP}{TP+FN}$) or sensitivity refers to the ability of the algorithm to detect road pixels. A low sensitivity corresponds to under-segmented results. False positive rate ($FPR = \frac{FP}{FP+TN}$) or fall-out refers to the ability of the algorithm to detect background pixels. Hence, a high fall-out corresponds to over-segmented results. However, in road images, a low fall-out does not ensure a high discriminative power since the number of false positives that can appear within the road areas is negligible compared to the number of background pixels. Hence, small fall-out variations may correspond to significant variations in the final road detection result. Finally, for performance comparison, we consider the area under the curve (AUC $\in [0..1]$), the higher AUC the higher the accuracy, and the equal error rate (EER) defined as the intersection between the curve and the line where error rates are equal *i.e.*, $(1 - TPR) = FPR$.

3.3. Results

The summary of AUC values resulting from combining the nineteen different color representations and the ten instances of single class classifiers is listed in Table 3. As expected, the worst performance corresponds to the nearest neighbor and the classifier using a single Gaussian. This is mainly due to the incapacity of these classifiers to capture all the information provided in the training samples in the form of skewed distributions and the presence of long tails

in these distributions. Further, we can see the stability of the model-based classifier with respect to the number of bins used to build the histogram. The relative low performance of this model-based classifier tends to improve by extending the training set using noisy samples. This is probably due to the lack of training samples representing the target (road) class. Therefore, adding noisy samples improve the variety of the training data. Note also how the performance of this model-based classifier drops by considering multiple color planes. The reason is the lack of ability to capture the joint distribution of these color planes. These results could be improved by considering the likelihood provided by each color plane independently. The use of multiple Gaussian distributions to characterize the training data yields to the best performance except when the number of Gaussians is optimized over the training data. In particular, forcing the classifier to learn up to four Gaussians leads to a significant improvement in the performance for most color planes and their combinations. More importantly, in this case, the improvement is even higher if the Gaussians are learned over multiple color planes. Interestingly, the best accuracy corresponds to color planes providing complete luminance and chrominance information (jointly as RGB or separately as HSV or Lab). Thus, their combination leads to diversified color ensembles. That is, ensembles where the components exhibit different invariant/sensitivity properties. On the other hand, combining the opponent color planes carrying color information only does not improve the performance over single channels. Representative ROC curves are shown in Fig. 3. As shown, results share a common pattern of accuracy except those exhibiting the highest performance. From these results we can conclude that learning Gaussian distributions using training samples from color spaces yields to the best accuracy. Moreover, there is no significant difference between those color spaces having luminance and chrominance information separated.

				Individual Color Representations												
				R	G	B	nr	ng	O_1	O_2	L	a	b	H	S	V
Non-param.		Nearest neighbor		67.9	70.6	70.1	72.3	69.6	71.5	72.9	72.1	72.4	71.6	70.0	70.0	71.1
	model-based	64 bins	S	76.5	77.0	78.1	78.7	75.2	74.0	74.9	76.7	77.9	76.8	80.9	77.3	76.7
			S&N	79.0	79.9	81.5	78.8	75.2	73.9	75.2	81.0	77.9	77.2	81.1	77.9	81.1
		100 bins	S	76.2	76.6	77.7	79.4	77.3	75.6	76.0	76.4	78.4	77.3	80.8	77.7	76.3
			S&N	79.1	79.9	81.5	79.4	77.3	75.5	76.4	81.0	78.4	77.7	81.0	78.3	81.1
Parametric		Gaussian		68.6	71.6	69.7	79.2	78.0	75.5	76.5	70.6	77.1	76.5	78.3	73.8	71.1
		Robust Gaussian		68.6	71.6	69.7	79.2	78.1	75.3	76.6	70.6	77.1	76.6	78.2	73.9	71.1
		MoG	2	79.3	80.0	81.7	80.0	78.2	78.1	79.2	80.2	79.6	78.6	81.2	78.6	80.4
			4	79.3	80.1	81.8	80.0	78.2	78.3	79.3	80.2	79.9	78.7	81.3	78.9	80.5
			Opt.	69.5	61.5	61.0	78.8	77.6	59.9	72.8	66.0	76.3	75.4	76.2	70.6	63.6

(a)

				Combination of Color Representations					
				RGB	$nrng$	O_1O_2	Lab	HSV	HS
Non-param.		Nearest neighbor		80.6	79.3	76.1	82.2	80.6	76.4
	model-based	64 bins	S	58.0	54.2	68.2	54.9	53.8	53.2
			S&N	70.6	55.4	68.3	53.4	62.8	63.0
		100 bins	S	55.8	54.0	68.9	53.9	53.2	53.6
			S&N	61.4	55.7	59.5	53.7	57.6	62.3
Parametric		Gaussian		86.2	87.2	80.3	86.9	86.2	83.0
		Robust Gaussian		81.6	87.0	79.6	84.5	73.0	82.2
		MoG	2	88.4	88.6	84.6	88.6	**88.8**	86.1
			4	88.3	88.6	84.8	88.3	88.5	86.3
			Opt.	61.5	81.5	60.9	65.0	64.0	75.1

(b)

Table 3. Quantative results for our experiments: a) Combinations of classifiers and single color planes. b) Combinations of classifiers and multiple color planes. The bold value indicates the best performance: a mixture of 2 Gaussians in the HSV color space.

4. Conclusions

In this paper, we introduced a comprehensive evaluation combining nineteen color representations with five different single class-classifiers for road detection. Experiments were conducted on a new large set of road image sequences comprising 7000 manually annotated images. From the results we can conclude that combining multiple color representations using a parametric classifier outperforms the accuracy of single color representations. Moreover, in this dataset, learning a mixture of two Gaussians in color spaces providing complete luminance and chrominance information yields best accuracy.

5. Acknowledgements

References

[1] A. Lookingbill, J. Rogers, D. Lieb, J. Curry, and S. Thrun, "Reverse optical flow for self-supervised adaptive autonomous robot navigation," *IJCV*, vol. 74, no. 3, pp. 287–302, 2007. 1

[2] M.A. Sotelo, F.J. Rodriguez, and L. Magdalena, "Virtuous: vision-based road transp. for unmanned operation on urban-like scenarios," *IEEE Trans. Intel. Transp. Systems (ITS)*, vol. 5, no. 2, pp. 69 – 83, June 2004. 1, 2

[3] C. Tan, T. Hong, T. Chang, and M. Shneier, "Color model-based real-time learning for road following," in *ITSC'06*, 2006, pp. 939–944. 1, 2

[4] H. Kong, J. Y. Audibert, and J. Ponce, "General road detection from a single image," *IEEE Trans. on Image Processing (TIP)*, vol. 19, no. 8, pp. 2211 –2220, 2010. 1

[5] J. M. Alvarez and A.M. Lopez, "Road detection based on illuminant invariance," *IEEE Trans. on Intel. Transp. Systems (ITS)*, vol. 12, no. 1, pp. 184 –193, 2011. 1, 2

[6] Y. He, H. Wang, and B. Zhang, "Color–based road detection in urban traffic scenes," *IEEE Trans. Intel. Transp. Systems (ITS)*, vol. 5, no. 24, pp. 309 – 318, 2004. 1

[7] P. Lombardi, M. Zanin, and S. Messelodi, "Switching models for vision-based on–board road detection," in *ITSC'05*, 2005, pp. 67 – 72. 1

[8] X. Lin and S. Chen, "Color image segmentation using modified hsi system for roadfollowing," in *ICRA'91*, 1991, pp. 1998–2003 vol.3. 1

[9] C. Rotaru, T. Graf, and J. Zhang, "Color image segmentation in HSI space for automotive applications," *Journal of Real-Time Image Processing*, pp. 1164–1173, 2008. 1

[10] A. Ess, T. Mueller, H. Grabner, , and L. van Gool, "Segmentation-based urban traffic scene understanding," in *BMVC'09*, Sep. 2009. 1
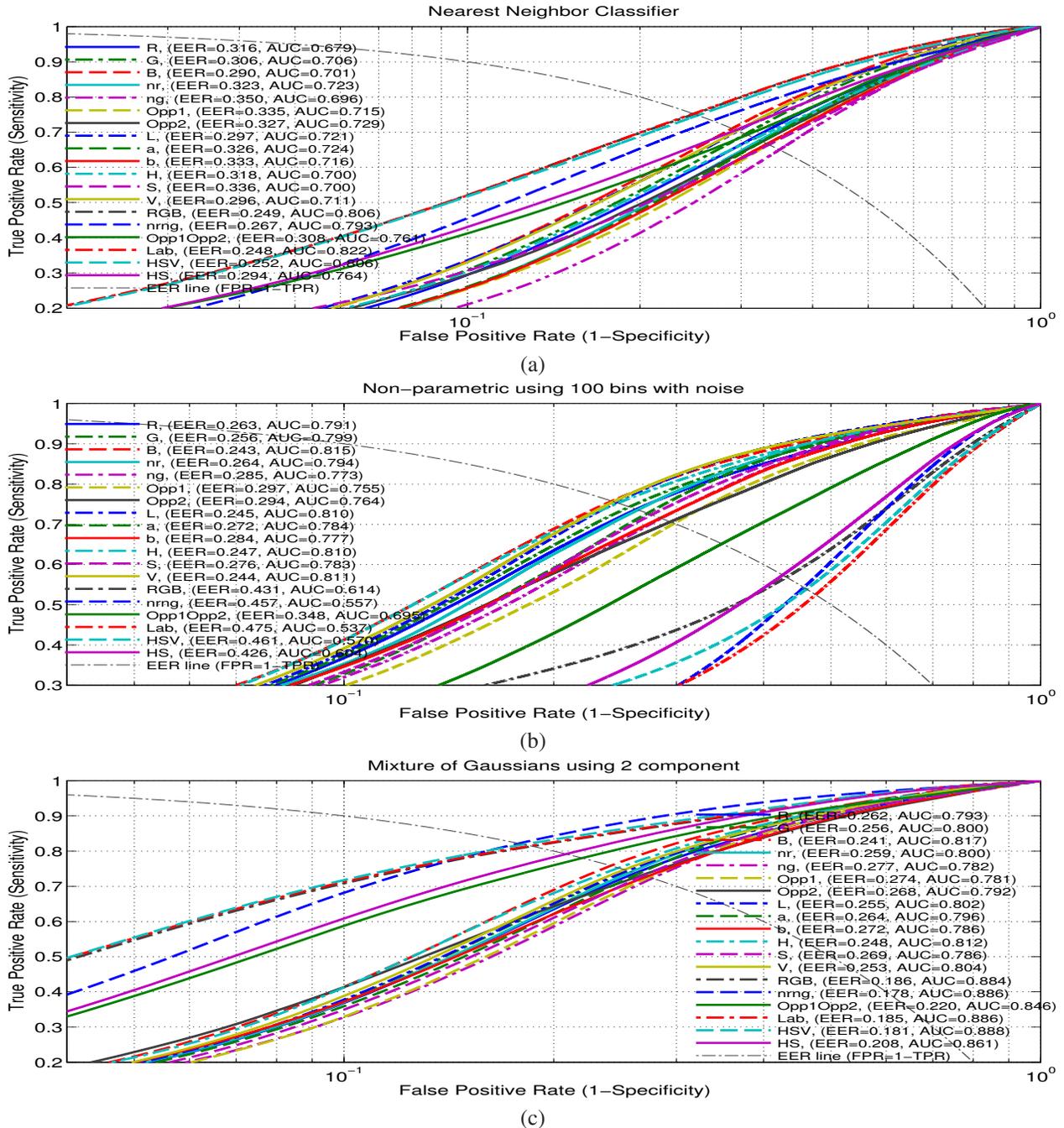
Figure 3. ROC curves for nearest-neighbor (a), model-based classifier using the extended training set and a histogram of 100 bins (b) and mixture of 2 Gaussians classifiers using single color planes and their combination. Best viewed in color.

[11] O. Ramstrom and H. Christensen, "A method for following unmarked roads," in *IV'05*, 2005. 1

[12] J. M. Alvarez, Theo Gevers, Yann LeCun, and A. M. Lopez, "Road scene segmentation from a single image," in *ECCV'12*, 2012, vol. 7578, pp. 376–389. 1

[13] T. Michalke, R. Kastner, M. Herbert, J. Fritsch, and C. Goerick, "Adaptive multi-cue fusion for robust detection of unmarked inner-city streets," in *IV'09*, 2009, pp. 1 –8. 1, 2

[14] J. M. Alvarez, Yann LeCun, Theo Gevers, and A. M. Lopez, "Semantic road segmentation via multi-scale ensembles of learned features," in *ECCVW 2012*, 2012, vol. 7584, pp. 586–595. 3

[15] T. Fawcett, "An introduction to roc analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861 – 874, 2006. 4

[16] J. M. Alvarez and A. M. Lopez, "Novel index for objective evaluation of road detection algorithms," in *ITSC'08*, Nov. 2008, pp. 815–820. 4