Learning Photometric Invariance from Diversified Color Model Ensembles

Jose M. Álvarez
Computer Vision Center and
Computer Science Dpt.
Univ. Autonoma de Barcelona
jalvarez@cvc.uab.es

Theo Gevers
Faculty of Science
University of Amsterdam
th.gevers@uva.nl

Antonio López
Computer Vision Center and
Computer Science Dpt.
Univ. Autonoma de Barcelona
antonio@cvc.uab.es

Abstract

*Color is a powerful visual cue for many computer vision applications such as image segmentation and object recognition. However, most of the existing color models depend on the imaging conditions affecting negatively the performance of the task at hand. Often, a reflection model (*e.g.*, Lambertian or dichromatic reflectance) is used to derive color invariant models. However, those reflection models might be too restricted to model real–world scenes in which different reflectance mechanisms may hold simultaneously.*

Therefore, in this paper, we aim to derive color invariance by learning from color models to obtain diversified color invariant ensembles. First, a photometrical orthogonal and non–redundant color model set is taken on input composed of both color variants and invariants. Then, the proposed method combines and weights these color models to arrive at a diversified color ensemble yielding a proper balance between invariance (repeatability) and discriminative power (distinctiveness). To achieve this, the fusion method uses a multi–view approach to minimize the estimation error. In this way, the method is robust to data uncertainty and produces properly diversified color invariant ensembles.

Experiments are conducted on three different image datasets to validate the method. From the theoretical and experimental results, it is concluded that the method is robust against severe variations in imaging conditions. The method is not restricted to a certain reflection model or parameter tuning. Further, the method outperforms state–of–the–art detection techniques in the field of object, skin and road recognition.

1. Introduction

Color is a powerful visual cue for many applications in computer vision such as image segmentation and object recognition. Most of the existing color models depend on the imaging conditions under which the image is recorded (such as illumination and camera viewpoint). Varying imaging conditions may disturb the measured color model values and hence the task at hand. Although reflection models (*e.g.*, Lambertian or dichromatic reflectance) are used to derive color invariant models [5, 18], these reflection models may be too restricted to model real–world scenes in which different reflectance mechanisms may hold simultaneously.

To avoid the requirement of explicit reflection models, in this paper, a combining strategy is proposed to obtain photometric invariance. In general, combining multiple classifiers (*e.g.*, color descriptors) is a powerful technique to improve the performance of single classifiers [2, 12, 14]. The improvement may even be higher when the method uses a learning step to adapt to the specific classification problem (*e.g.*, Boosting, Bagging and Random forests). To accomplish the learning procedure, systems use training data corresponding to the object to be recognized (*i.e.*, positive examples) and for instance background (*i.e.*, negative examples). However, systems using only positive data within the training step are more desirable since obtaining a comprehensive representation of negatives or unknown universe is often unfeasible. In addition, if negative data is not chosen properly this may lead to lower classification accuracy [28].

Therefore, in this paper, we aim to derive color invariance by learning from positive examples of color models to obtain diversified color invariant ensembles. The training examples should include a broad range of varying imaging conditions under which the object/image is recorded. An orthogonal and non–redundant color model set is taken on input composed of both color variants and invariants. Then, the proposed method combines and weights these color models to arrive at a diversified color ensemble yielding a proper balance between invariance (repeatability) and discriminative power (distinctiveness). To achieve this, the method uses a multi–view approach to minimize the estimation error. In addition, the contribution of each observation is estimated using a Monte Carlo simulation. In this way, the method is robust to data uncertainty and produces prop-

978-1-4244-3991-1/09/$25.00 ©2009 IEEE

erly diversified color invariant ensembles.

The paper is organized as follows. First, in Sect. 2, related work is discussed. In Sect. 3, the multi–view fusion scheme is introduced. Color–based region detection is outlined in Sect. 4. Then, in Sect. 5, experiments are presented and the results are discussed. Finally, conclusions are drawn.

2. Related Work

Combining strategies that consider the differences between their components are promising [14]. In this paper, the measure of disagreement is referred as diversity. Some approaches consider the diversity directly in the process of defining the ensemble. For instance, Melville *et al*. [16] consider the diversity as the disagreement of an ensemble member with the ensemble's prediction to learn the ensemble based on positive and negative data. Jacobs *et al*. [10] proposes a minimum variance estimator where the estimated aggregate has a variance at most as large as the variance of any of the input features. Stokman *et al*. [26] uses the Markowitz diversification criterion [15] in the process of defining the ensemble. The method assumes that each descriptor can be characterized by its mean and variance and computes the best combination to yield maximum mean to variance ratio. The method provides maximal feature discrimination assuming unimodal data distributions. However, in practice, the distribution of the training data is often not unimodal. Hence, the uncertainty in estimating the mean and variance may increase. Further, a quadratic optimization technique is used to estimate the contribution of each component for the final ensemble. This optimizer method tends to select those components with very attractive features and tends to deselect those with the worst features. These are the cases where estimation error is likely to be maximal [22].

3. Combining Method based on Diversified Ensembles

We aim to derive color invariance by learning from color models to obtain diversified color invariant ensembles. To be precise, an object (U) is modeled considering the central value of its data distribution (μ_U) and the small deviations from the central value due to noise in the device (σ_U),

$$U = \mu_U \pm \sigma_U, \tag{1}$$

A weighted linear combination of the observations estimates μ_U and σ_U,

$$\mu_U = w\mu^T, \tag{2}$$

$$\sigma_U = w\Sigma w^T, \tag{3}$$

where $w = [w_1, ..., w_N]$ is the contribution of each feature (observation) to the final combination, Σ is the covariance matrix between the central values of each view and each representation. $\mu = [\mu_1, ..., \mu_N]$ is a vector with the central value of all the views for each observation. N corresponds to the number of components in the ensemble.

To estimate w_i, the algorithm aims to minimize σ_U to obtain a minimum mean value $\mu_{U_{min}}$ as follows,

$$\begin{aligned} &minimize(\sigma_U) \\ &\text{s.t. } \mu_U \geq \mu_{U_{min}}, \end{aligned} \tag{4}$$

Further, the full combination constraint is imposed. That is, the weights must sum up to one.

To estimate the central value of each observation (μ_i), a multi–view framework is proposed. This framework characterizes the information from each observation using two stages. First, the central value of the j–th view, μ_{ij}, is computed and used to build a data distribution. Second, the central value of the resulting distribution is obtained corresponding to the central value of the i-th observation, μ_i. To obtain each central value, the algorithm uses the mode to minimize the influence of skewed distributions. Thus, the estimation error is minimized. The variance–covariance matrix of the data represents the existing relations between observations when the viewing conditions change.

What remains now is the estimation of w_i. Using standard optimization techniques, a set of efficient ensembles called the efficient frontier is obtained [22]. That is, the efficient frontier contains different values of μ_U and their associated weights which minimize the corresponding σ_U. However, to deal with the estimation error and improve the diversity of the ensemble, a resampling technique is proposed using Monte Carlo simulation which yields a resampled frontier [17]. Ensembles lying on the resampled frontier are composed of weights vectors which are the average of the efficient frontiers given a certain μ_U.

The above method can be summarized as follows:

1. Estimate the efficient frontier using the training data and quadratic programming techniques. This frontier is composed of ensembles varying from the minimum–variance to the maximum expected value ensemble. Divide the difference between the minimum and maximum return in m ranks.

2. Estimate the variance–covariance matrix, Σ, and mean vector, μ, of the training data,

$$\mu_i = \frac{1}{K}\sum_{j=1}^{K} \mu_{ij}, \tag{5}$$

$$\Sigma_{il} = cov(\mu_i, \mu_l), \tag{6}$$

   where K is the number of views.

3. Resample, using the inputs created in 2, taking T draws for the input distribution. T, reflects the degree of uncertainty in the training data. Calculate a

new variance–covariance matrix from the sampled series. Estimation error will result in different variance–covariance matrices and mean vector for those in Step 2.

4. Calculate the efficient frontier for the inputs derived in Step 3. Record the optimal ensemble weights for m equally distributed points along the frontier.

5. Repeat Step 3 and Step 4 K times. Calculate the averaged ensemble weights for each expected value point

$$\overline{w}_m^{resampled} = \frac{1}{K}\sum_{i=1}^{K} w_{mi}, \tag{7}$$

where w_{mi} denotes the weight vector for the m–th ensemble along the frontier for the i–th sampling.

6. Evaluate the frontier of averaged ensembles with the variance–covariance matrix from the original training data to obtain the resampled frontier.

Finally, the Sharpe Ratio (*SR*) [4] can be used to select the most efficient ensemble from the frontier. *SR* is a single statistical performance measure of variance–adjusted return defined as,

$$SR_i = \frac{\mu_d}{\sigma_d}, \tag{8}$$

where SR_i is the performance of each ensemble in the resampled frontier, μ_d is the difference between the expected value of the i–th ensemble and the expected value of a reference ensemble. σ_d is the variance of the differential ensemble.

4. Application to Color–based Region Detection

In this section, the combining method is applied to color–based region detection, *i.e.*, the detection of object patches in images recorded under varying imaging conditions using a set of color models composed of both color variant and invariant models. In this way, U is the data distribution of the final combination of color (invariant) planes/models and μ_U and σ_U its central value and variance respectively. μ_i is the central of the i–th color plane estimated using the multi-view procedure. Finally, w_i denotes the contribution of the i–th color plane to the final ensemble.

During **training** (*i.e.*, estimating μ_U, σ_U and w_i), the following steps are performed:

- Select a set of training images containing the object to be detected under different views (*e.g.*, different lighting conditions).
- Select a region of interest for each training image and for each color plane/model estimate μ_{ij} using the mode.
- Estimate the correlation matrix of these central values. This matrix contains information regarding the relative variations of each color plane when the acquisition conditions vary.
- Compute the weights using the Monte Carlo method considering the central value of each color model for all the views and the covariance matrix as input data.

Then, during **classification**, the following steps are performed:

- Convert the image into the color models (the same as during training) and apply the weights obtained in the training phase to combine them. This leads to a grey–level image.
- Estimate the signal to noise ratio, *SNR*, by dividing, at each pixel, the local mean value by the local standard deviation. The *SNR* is estimated using a rectangular region ($M \times N$ pixels) at each pixel.
- Compute the error between the *SNR* based on the training data and the local *SNR* for each pixel. The lower the error, the more similar the colors are.
- Threshold the error image e and assign pixels values close to zero the target color:

$$C(x,y) = \begin{cases} 1 & \text{if } e(x,y) < T \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

The appropriate value of T can be obtained using automatic thresholding techniques such as isodata method [20].

Taxonomy of color spaces	LI change	LI shift	LI c.&s.	LC change	LC c.&s.
RGB	-	-	-	-	-
O_1,O_2	-	+	-	-	-
O_3, Intensity, L	-	-	-	-	-
Saturation (*S*)	-	+	+	-	-
Hue (*H*)	+	+	+	-	-
r,*g*,*a*,*b*	+	-	-	-	-
$\mathfrak{I}$	+	+	+	+	+

Table 1. Invariance of color planes for different types of lighting variations *i.e.*, light intensity (LI) or light color (LC) change and/or shift [18]. Invariance is indicated with '+' and lack of invariance with '-'.

To provide robustness against confounding imaging conditions (*e.g.*, illumination, shading, highlights, and inter-reflections), photometric invariants have been proposed (Table 1) derived from [18]. For instance, for the dichromatic

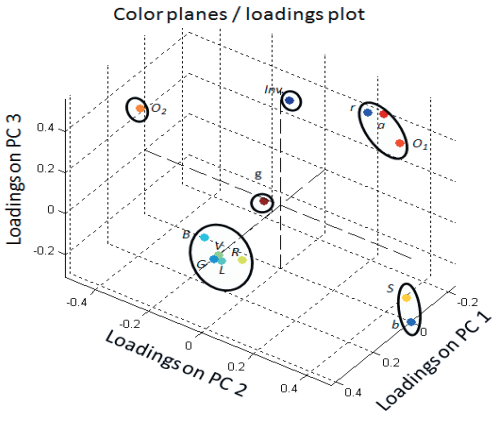


Figure 1. PCA is used to reduce redundancy within the training data. The analysis is done using the loadings plot of each color plane. This example corresponds to the training set from the face database.

reflection model, normalized color *rgb* is to a large extent invariant to a change in camera viewpoint, object pose, and the direction and intensity of the incident light. In addition, the hue color space H is insensitive to highlights under the restriction of white illumination or a white–balanced camera. In addition to the models described by [18], the illumination invariant model $\mathfrak{I}$ proposed in [5] is included.

Considering all the color models in Table 1, a set is obtained of both color variants and invariants to achieve both distinctiveness and repeatability respectively. The next step is to obtain a non–redundant set. This subset is obtained by computing a PCA to reduce the number of color models. Correlation between color models are computed by the loadings of each color model, see Fig. 1. The input data to PCA is the matrix containing the mean values for each view of each color plane (μ_{ij}). The number of principal components depends on the data and the amount of variation. The selection of color planes which represent each cluster (*e.g.*, S or b in Fig. 1) is computed by the Hartigan's test for unimodality [8]. In this way, an orthogonal (variant/invariant) and non–redundant (decorrelated) color model set is obtained which will be used as input of the proposed method as explained and tested in the next section.

5. Experiments

In this section, the proposed fusion algorithm is applied on three different databases: (1) the Amsterdam Library of Object Images (ALOI) [7], (2) Caltech Face database [19] and (3) a road sequence taken by an on–board camera. See Fig. 2. The goal of the first experiment on the ALOI dataset is to detect object regions under varying imaging conditions. The second experiment consists of detecting skin to find faces in the Caltech image dataset. The aim of the last experiment is to detect roads under uncontrolled imaging conditions.

5.1. Error Measures

Contingency Table		Ground Truth	
		Non–Target	Target
Detection Result	Non–Target	TN	FN
	Target	FP	TP

Table 2. The contingency table. Algorithms are evaluated based on the number of pixels correctly and incorrectly classified.

Quantitative evaluations are provided using pixel–wise measures, see Table 2, from which the following error measures are computed: quality, detection accuracy, detection rate and effectiveness, see Table 3. Each of these measures provides a different insight in the performance of a method. Quality takes into account the completeness of the extracted data as well as its correctness. Detection accuracy, also known as precision, is the probability that the result is valid. Detection rate, or recall, is the probability that the ground–truth data is detected. Effectiveness is a single measure that trades–off the detection accuracy versus detection rate. Further, the performance of our method is compared, on each dataset, with existing algorithms. Pair-wise comparisons are computed by the Wilcoxon significance test [29].

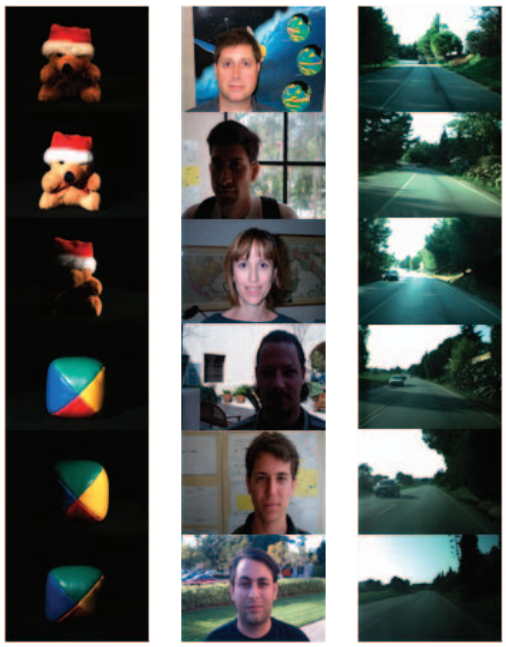
Pixel–wise measure	Definition
Quality ($\hat{g}$)	$\hat{g} = \frac{TP}{TP+FP+FN}$
Detection Accuracy (DA)	$DA = \frac{TP}{TP+FP}$
Detection Rate (DR)	$DR = \frac{TP}{TP+FN}$
Effectiveness (F)	$F = \frac{2DADR}{DA+DR}$

Table 3. Pixel–wise measures used to evaluate the performance of the different algorithms. These measures are defined using the entries of the contingency table (Table 2).

5.2. Man–made Object Region Detection

Objects are taken from the Amsterdam Library of Object Images (ALOI) [7]. The objects are captured under varied viewing angle, illumination angle, and illumination color. These lights were chosen to be representative of the spread of common illuminants. The goal of this experiment is to detect object regions and it is conducted on each image (36 per object) for three different objects. Two of them are shown in Fig. 2. All images have their corresponding ground–truth.

Different regions from each image have been manually selected to train the algorithm. Although these regions comprise the same object under different lighting conditions,



(a) (b) (c)

Figure 2. Example images from the different image datasets. (a) Two objects under different light conditions from the ALOI database. (b) Different faces from the Caltech database. (c) Different frames from the road sequence.

each region contains pixels from an homogeneously illuminated patch. That is, the data distribution of each patch is approximately unimodal. The set of photometric orthogonal and non–redundant color models has been computed using the (PCA) procedure described in Sect. 4. As a result, we have obtained the weights listed in Table 4. High weights are given to the $\mathfrak{I}$, B and g. As $\mathfrak{I}$ and V do not contain chromatic information, the weights of B and g reflect the color of the objects region.

For comparison, two different weighting algorithms have been implemented: the minimum variance [10] and the single–view fusion scheme [26]. These two algorithms have been trained using the same dataset as our method. The summary of the results is reported in Table 5. From the results, it can be observed that considering only the variance of the training data (*i.e.*, the minimum variance method) is not sufficient to provide a proper model. In fact, single–view and multi–view methods not only aim at minimizing the variance but also yield a certain mean value of the ensemble.

Since training pixels are obtained under different imaging conditions, the behavior of the different color planes can not be captured properly. In contrast, the proposed method is able to model this phenomenon due to the relative variations around the central value in each view and hence outperforming the other methods.

	ALOI object 1	Faces	Road
$\mathfrak{I}$	0.595	-0.017	0.929
R	0.048	—	—
G	—	—	—
B	0.328	—	—
r	—	—	0.157
g	0.276	0.022	0.342
O_1	—	—	0.266
O_2	—	0.013	-0.024
L	—	0.176	—
a	—	0.652	-0.356
b	—	0.154	-0.082
S	-0.032	—	-0.452
V	-0.215	—	0.220

Table 4. Set of weights obtained for the experiments. '—' corresponds to an unselected color plane by the PCA procedure.

5.3. Skin Detection

The second experiment consists of detecting skin pixels of faces from The Frontal Face Image Database of Caltech. This image dataset contains 450 face images taken from 27 different persons under different lighting, expressions and backgrounds. All these images have been manually segmented to generate ground–truth. The training set has been obtained by manually selecting 100 different patches from 100 different (randomly chosen) images. Further, the unimodality test is used to discard unappropriate patches. Finally, 58 patches have been used for training representing 1% of facial pixel in the database. Again, the color model set has been computed using the procedure described in Sect. 4. The set of weights obtained are listed in Table 4. The set of weights obtained reveals a dominance in a and b reflecting pale reddish (*i.e.*, skin).

The performance of the proposed method is compared to six different existing skin detection algorithms. Three of them use fixed boundaries in *RGB* [6], *CbCr* [3] and *HS* [24] color spaces. The fourth is a statistical approach using a mixture of Gaussians in *RGB* space. Note that these methods are particularly designed and fine–tuned to detect skin. The other two methods correspond to the (more generic) fusion schemes proposed by [10] and [26]. The same training set has been used to train the fusion schemes as our method. The summary of the results is listed in Table 6. From the results, it can be derived that our method outperforms the others except for the *RGB* based method. This is because, after the analysis of the failures, although the *RGB*–based method fails in the presence of low intensity (due to illumination and shadows) there were only a few instances of this type *i.e.*, only (3%) of images in the image dataset show severe intensity and shadow changes. Moreover, the method has been fine–tuned for skin detection. Further, the results of

	$\hat{g}$	Detection Accuracy	Detection Rate	F
Minimum variance [10]	0.156 ± 0.09	0.269 ± 0.12	0.305 ± 0.21	0.259 ± 0.15
Single–view fusion [26]	0.478 ± 0.11	0.627 ± 0.13	$\mathbf{0.789 \pm 0.06}$	0.694 ± 0.14
Multi–view (without color plane selection)	0.294 ± 0.15	0.419 ± 0.28	0.784 ± 0.24	0.627 ± 0.25
Multi–view (our method)	$\mathbf{0.639 \pm 0.13}$	$\mathbf{0.909 \pm 0.03}$	0.687 ± 0.15	$\mathbf{0.778 \pm 0.11}$

Table 5. Performance of different detection algorithms for the first object from the ALOI database. Bold values indicate the maximum performance.

	$\hat{g}$	Detection Accuracy	Detection Rate	F
***RGB* based method [6, 13]**	$\mathbf{0.640 \pm 0.19}$	0.694 ± 0.20	$\mathbf{0.884 \pm 0.17}$	$\mathbf{0.761 \pm 0.17}$
***CbCr* based method [3]**	0.259 ± 0.18	0.309 ± 0.21	0.548 ± 0.31	0.379 ± 0.23
***HS* based method [24]**	0.443 ± 0.21	0.514 ± 0.21	0.807 ± 0.28	0.585 ± 0.21
***RGB* Statistical skin detection [11]**	0.510 ± 0.23	0.635 ± 0.23	0.723 ± 0.28	0.643 ± 0.22
Minimum variance [10]	0.189 ± 0.03	0.195 ± 0.03	0.190 ± 0.02	0.318 ± 0.05
Single–view fusion [26]	0.314 ± 0.24	0.365 ± 0.26	0.636 ± 0.34	0.430 ± 0.27
Multi–view (without color plane selection)	0.410 ± 0.23	0.703 ± 0.18	0.497 ± 0.20	0.550 ± 0.15
Multi–view (our method)	0.569 ± 0.14	$\mathbf{0.776 \pm 0.20}$	0.696 ± 0.11	0.714 ± 0.13

Table 6. Performance of different detection algorithms on Caltech face database. Bold values indicate the maximum performance.

	RGB based				*CbCr* based				*HS* based				*RGB* Statistical				Single–view				Multi–view[a]			
	$\hat{g}$	*DA*	*DR*	*F*	$\hat{g}$	*DA*	*DR*	*F*	$\hat{g}$	*DA*	*DR*	*F*	$\hat{g}$	*DA*	*DR*	*F*	$\hat{g}$	*DA*	*DR*	*F*	$\hat{g}$	*DA*	*DR*	*F*
Multi–view	-1	**1**	-1	-1	**1**	**1**	**0.9**	**1**	**1**	**1**	-1	**1**	**1**	**1**	-1	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**

Table 7. The Wilcoxon test for the skin detection experiment. A positive value indicates that our method outperforms the others. A negative value indicates when our method does not perform significantly better. Minimum variance method has been excluded for space reasons. Bold values indicate when the proposed method outperforms the others.

[a] without color plane selection

the Wilcoxon test are shown in Table 7. These results show that the proposed method outperforms all methods except for *RGB*–based, and for the *HS* and *RGB* statistical method in terms of detection rate. This is because of the high variability in both skin appearance and lighting variations. This yields a data distribution in each view which is not unimodal except for very small patches of skin.

5.4. Road Detection

The final experiment is conducted on a video sequence of more than 800 images recorded using an on–board camera. The aim is to detect the (not occluded) road in front of a moving vehicle using a color camera. The images present different backgrounds, the occurrence of occluding and cluttered objects (vehicles) and different road appearances under varying illumination changes.

The training set consists of 15 different road patches which have been manually selected from 15 different (randomly) selected images. The selection process avoids successive image indexes. These patches contain different illumination effects of the road (*i.e.*, shadows and highlights). These regions represent less than 0.053% of the total amount of road pixels within the sequence. The selection of the most suitable color planes has been done following the PCA procedure described in Sect. 4. The obtained weights for the ensemble are listed in Table 4 and shows a dominant weight for the invariant color plane corresponding to an achromatic (grey) surface independent of illumination changes (*e.g.*, sun casts and shadows) *i.e.*, roads.

For comparison, the video sequence has been processed using three state–of–the–art methods. The first algorithm is the *HSI* road detection (RD) algorithm proposed in [25] and used in [21]. The *HSI* color space has also been used to process generic outdoor scenes under varying illumination [9, 23]. The second algorithm is the illuminant–invariant algorithm presented in [1]. The third algorithm is based on $2D$ histograms in the *rg* space [27]. Further, the two fusion methods proposed in [10] and [26] are considered. Note that the *HSI* and illuminant–invariant algorithms are based on a frame–by–frame procedure. Further, these algorithms require various parameter settings. For fair comparison, a brute force approach has been applied. In this way, a set of images has been processed and evaluated using all possible values within the range of each parameter. The optimal set

	$\hat{g}$	Detection Accuracy	Detection Rate	F
***HSI* based RD [25]**	0.662 ± 0.11	0.933 ± 0.10	0.715 ± 0.16	0.789 ± 0.10
Invariant RD [1]	0.784 ± 0.13	0.904 ± 0.14	0.863 ± 0.10	0.872 ± 0.09
***rg* model based [27]**	0.257 ± 0.25	0.778 ± 0.28	0.409 ± 0.39	0.352 ± 0.30
Minimum variance [10]	0.137 ± 0.22	0.237 ± 0.30	0.193 ± 0.31	0.187 ± 0.28
Single–view fusion [26]	0.680 ± 0.14	0.936 ± 0.02	0.716 ± 0.15	0.801 ± 0.10
Multi–view (without color plane selection)	0.801 ± 0.36	0.714 ± 0.10	0.826 ± 0.05	0.746 ± 0.07
Multi–view (our method)	$\mathbf{0.882 \pm 0.06}$	$\mathbf{0.982 \pm 0.03}$	$\mathbf{0.869 \pm 0.06}$	$\mathbf{0.926 \pm 0.03}$

Table 8. Performance of different detection algorithms on road database. Bold values indicate the maximum performance.

Figure 3. Results of the proposed algorithm to detect roads.

of parameter values is the one which maximizes the average performance. All algorithms (which need training) have been trained using the same road pixels. The performance of all algorithms is outlined in Table 8. Various detection results of our method are shown in Fig. 3. Further, the results of the Wilcoxon test are shown in Table 9. From the results it can be concluded that the proposed method performs significantly better than the others except for the detection rate on Invariant based method and detection accuracy on the *HSI* method. However, regarding the performance trade–off between detection rate and detection accuracy, the proposed method performs best. This means that the proposed algorithm has a higher trade–off between invariance (detection rate) and discriminative power (detection accuracy).

6. Conclusions

In this paper, photometric invariance has been derived by learning from color models to obtain diversified color invariant ensembles using only positive examples. A combining method is proposed to provide a multi–view approach to minimize the estimation error. In this way, the method is robust to data uncertainty and produces properly diversified color invariant ensembles.

Experiments are conducted to validate the method. From these experiments it is concluded that method is robust against variations in imaging conditions and is not restricted to a certain reflection model. Further, the method performs similar or outperforms state–of–the–art detection

	rg model based				Invariant RD				*HSI* RD				Min. Variance				Single–view				Multi–view[a]			
	$\hat{g}$	*DA*	*DR*	*F*	$\hat{g}$	*DA*	*DR*	*F*	$\hat{g}$	*DA*	*DR*	*F*	$\hat{g}$	*DA*	*DR*	*F*	$\hat{g}$	*DA*	*DR*	*F*	$\hat{g}$	*DA*	*DR*	*F*
Multi–view	**1**	**1**	**1**	**1**	**1**	**1**	-0.8	**1**	**1**	-0.9	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**

Table 9. Wilcoxon test for the road detection experiment. Positive values indicate that the proposed method performs significantly better. Negative values indicate that our method does not outperform the others. Bold values indicate when our method outperforms the others.

[a] without color plane selection

techniques in the field of object, skin and road recognition.

7. Acknowledgements

This work was supported by the Spanish Ministry of Education and Science under project TRA2007-62526/AUT and research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018).

References

[1] J. M. Álvarez, A. M. López, and R. Baldrich. Illuminant-invariant model-based road segmentation. In *Procs. of the 2008 IEEE Intel. Vehicles Symposium (IV'08)*, Eindhoven, The Netherlands. 6, 7

[2] G. Brown, J. Wyatt, R. Harris, and X. Yao. Diversity creation methods: a survey and categorisation, 2005. 1

[3] D. Chai and K. Ngan. Face segmentation using skin-color map in videophone applications. *Circuits and Systems for Video Tech., IEEE Trans. on*, 9(4):551–564, Jun 1999. 5, 6

[4] K. Dowd. *Beyond value at risk : the new science of risk management*. Wiley, 1998. 3

[5] G. Finlayson, S. Hordley, C. Lu, and M. Drew. On the removal of shadows from images. *IEEE Trans. on PAMI*, 28(1), 2006. 1, 4

[6] M. M. Fleck, D. A. Forsyth, and C. Bregler. Finding naked people. In *ECCV (2)*, volume 1065, pages 593–602. Springer, 1996. 5, 6

[7] J.-M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders. The amsterdam library of object images. *Int. J. Comput. Vision (IJCV)*, 61(1):103–112, 2005. 4

[8] J. A. Hartigan and P. M. Hartigan. The dip test of unimodality. *The Annals of Statistics*, 13(1):70–84, 1985. 4

[9] N. Ikonomakis, K. Plataniotis, and A. Venetsanopoulos. Color image segmentation for multimedia applications. *Journal of Intel. Robotics Systems*, 28(1-2), 2000. 6

[10] R. A. Jacobs. Methods for combining experts' probability assessments. *Neural Comp.*, 7(5):867–888, 1995. 2, 5, 6, 7

[11] M. J. Jones and J. M. Rehg. Statistical color models with application to skin detection. *Int. J. Comput. Vision (IJCV)*, 46(1):81–96, 2002. 6

[12] J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE Trans. on PAMI*, 20(3):226–239, March 1998. 1

[13] J. Kovac, P. Peer, and F. Solina. Human skin color clustering for face detection. *EUROCON*, 2003. 6

[14] L. I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004. 1, 2

[15] H. M. Markowitz. *Portfolio selection: efficient diversification of investments.* Wiley, New York, 1959. 2

[16] P. Melville and R. J. Mooney. Creating diversity in ensembles using artificial data. *Information Fusion*, 6(3):1553–1563, May 2005. 2

[17] R. O. Michaud. *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Oxford University Press, USA, June 1998. 2

[18] Koen E. A. van de Sande and Theo Gevers and Cees G. M. Snoek. Evaluation of color descriptors for object and scene recognition. In *CVPR*, pages 453–464, 2008. 1, 3, 4

[19] The Caltech Frontal Face Dataset, by M. Weber. available online at: http://www.vision.caltech.edu/html-files/archive.html. *California Inst. of Tech., USA*. 4

[20] T. Ridler and S. Calvard. Picture thresholding using an iterative selection method. *Systems, Man and Cybernetics, IEEE Trans. on*, 8(8):630–632, Aug. 1978. 3

[21] C. Rotaru, T. Graf, , and J. Zhang. Color image segmentation in hsi space for automotive applications. *Journal of Real-Time Image Processing*, pages 1164–1173, 2008. 6

[22] B. Scherer. *Portfolio Construction and risk Budgeting*, chapter 4. Rosk Books, London, 2002. 2

[23] L. Sigal, S. Sclaroff, and V. Athitsos. Skin color-based video segmentation under time-varying illumination. *IEEE Trans. on PAMI*, 26(7):862–877, 2004. 6

[24] K. Sobottka and I. Pitas. A novel method for automatic face segmentation, facial feature extraction and tracking. *Signal Processing: Image Communication*, 12(3):263–281, June 1998. 5, 6

[25] M. Sotelo, F. Rodriguez, L. Magdalena, L. Bergasa, and L. Boquete. A color vision-based lane tracking system for autonomous driving in unmarked roads. *Auton. Robots*, 16(1), 2004. 6, 7

[26] H. Stokman and T. Gevers. Selection and fusion of color models for image feature detection. *IEEE Trans. on PAMI*, 29(3):371–381, 2007. 2, 5, 6, 7

[27] C. Tan, T. Hong, T. Chang, and M. Shneier. Color model-based real-time learning for road following. *Procs. IEEE ITSC*, pages 939–944, 2006. 6, 7

[28] D. M. J. Tax and R. P. W. Duin. Uniform object generation for optimizing one-class classifiers. *J. Mach. Learn. Res.*, 2:155–173, 2002. 1

[29] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945. 4