Robust Histogram Construction from Color Invariants for Object Recognition

Theo Gevers, *Member*, *IEEE*, and Harro Stokman

Abstract—An effective object recognition scheme is to represent and match images on the basis of histograms derived from photometric color invariants. A drawback, however, is that certain color invariant values become very unstable in the presence of sensor noise. To suppress the effect of noise for unstable color invariant values, in this paper, histograms are computed by variable kernel density estimators. To apply variable kernel density estimation in a principled way, models are proposed for the propagation of sensor noise through color invariant variables. As a result, the associated uncertainty is obtained for each color invariant value. The associated uncertainty is used to derive the parameterization of the variable kernel for the purpose of robust histogram construction. It is empirically verified that the proposed density estimator compares favorably to traditional histogram schemes for the purpose of object recognition.

Index Terms—Object recognition, color invariants, noise robustness, histogram construction, noise propagation, kernel density estimation, matching.

1 Introduction

COLOR is a powerful information cue for object recognition. To provide object recognition robust against confounding imaging conditions (e.g., illumination, shading, highlights, and interreflections), color histograms are usually computed from photometric color invariants [1], [2], [3], [4], [5]. For example, illumination-independent color ratios have been proposed by Funt and Finlayson [1] and Nayar and Bolle [6]. Further, for the dichromatic reflection model, Gevers and Smeulders [2] showed that normalized color rgb ($c_1c_2c_3$) 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 ($l_1l_2l_3$) is insensitive to highlights under the restriction of white illumination or a white-balanced camera.

Although color invariance is essential for robust color-based object recognition, the corresponding color transformations used to compute these color invariants bring with them several serious drawbacks, since these transformations are singular at some sensor values and unstable at many others. In fact, color ratios, rgb, and the $c_1c_2c_3$ color space become unstable near the black point while hue H and $l_1l_2l_3$ are (very) unstable near the achromatic axis [2]. As a consequence, a small perturbation of sensor values will cause a large jump in the transformed values. Traditionally, the effect of noise blow-up at unstable color invariant values is simply ignored or suppressed by ad hoc color thresholding. For instance, during histogram construction, all sensor values along the achromatic axis could be discarded by eliminating all sensor values having a saturation and intensity value smaller than 5 percent of the total range. Inevitably, more elaborated computational methods are required to construct robust histograms from color invariants.

Therefore, in this paper, a more principled method is proposed to suppress the effect of noise during histogram construction from color invariants. To achieve this, variable kernel density estimation is employed to construct color invariant histograms. To apply variable kernel density estimation in a proper way, computational methods are presented for the propagation of sensor noise through color invariant transformations. As a result, the associated uncertainty is known for each color invariant value. The associated uncertainty is used to derive the optimal parameterization of the variable kernel used during histogram construction.

This paper is organized as follows: In Section 2, related work is reviewed. A theoretical model is presented in Section 3 to derive the uncertainty for each color invariant value. In Section 4, the optimal kernel parameterization is proposed and incorporated in the variable kernel density estimator. Then, in Section 5, experiments are conducted in which the computed uncertainty in the transformed color values is empirically compared to the actual uncertainty. Finally, in the context of color-based object recognition, the variable kernel density estimator is compared to traditional histogram matching.

2 Previous Work

It is known from Kender [7] that normalized color rgb is undefined at the black point ($R = G = B = 0$), and that hue H is undefined at the entire achromatic axis ($R = G = B$). Moreover, rgb and H become very unstable at these singularities, where a small perturbation in the RGB-values (e.g., due to sensor noise) will cause a large jump in the transformed values. Traditionally, this effect of noise blow-up is suppressed by ad hoc thresholding of the transformed values. Ohta [8] considers only RGB-values if the intensity is larger than 30 (on a range of 256 values), and rejects hue values if the saturation times the intensity is less than nine. Healey [9] rejects rgb-values when the RGB-values fall within the sphere of radius 4σ centered at the origin of the RGB space.

A more elaborated approach is given by Burns and Berns [10], analyzing the error propagation through the CIE $L^*a^*b^*$ color space. The goal was to estimate the influence of the mean, variance, and covariance of the CIE $L^*a^*b^*$ colors under the influence of noise. Further, Shafarenko et al. [11] use an adaptive filter for noise reduction in the CIE $L^*u^*v^*$ space prior to 3D color histogram construction. In fact, the filter width is steered based on the covariance matrix of the noise distribution in the CIE $L^*u^*v^*$ space.

As opposed to previous work, the aim of this paper is to use kernel density estimation (see, e.g., [12], [13]) to construct robust color invariant histograms for the purpose of object recognition. Kernel density estimators form an alternative to histograms as density estimators. The novelty of this paper is to propose a mathematical framework to variable kernel density estimation based on noise propagation through color invariants. In this way, the associated uncertainty is computed for each color invariant value which is used to steer the kernel sizes. Although the method is suited for different color invariants (e.g., color ratios, $c_1c_2c_3$, $l_1l_2l_3$, etc.) or color transformations in general, in this paper, we focus on normalized color rgb, opponent color o_1o_2, and hue θ as these color spaces are widely in use in computer vision tasks.

3 Noise Propagation through Color Invariants

In Section 3.1, we review on the photometric color invariant, properties of the color models. Then, in Section 3.2, models are proposed for the propagation of sensor noise through these color invariant models.

3.1 Photometric Color Invariance

The reflection from inhomogeneous dielectric materials under white, spectrally smooth illumination or a white-balanced camera is given by [2], [14]

• *The authors are with Intelligent Sensory Information Systems, Department of Computer Science, Faculty of Science, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands.*
E-mail: gevers@science.uva.nl.

Manuscript received 3 Apr. 2001; revised 25 June 2002; accepted 17 Oct. 2002.
Recommended for acceptance by S. Dickinson.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 113919.

$$\omega_k = G_B(\vec{n},\vec{s})E\int_\lambda B(\lambda)F_k(\lambda)d\lambda + G_S(\vec{n},\vec{s},\vec{v})ESF, \quad (1)$$

for $k \in \{R,G,B\}$ giving the red, green, and blue sensor response of an infinitesimal matte surface patch under the assumption of a white or spectrally smooth light source. Spectral sensitivities are given by $F_R(\lambda)$, $F_G(\lambda)$, and $F_B(\lambda)$, respectively, where λ denotes the wavelength. We assume that the integration white conditions holds, i.e., $\int_\lambda F_R(\lambda)d\lambda = \int_\lambda F_G(\lambda)d\lambda = \int_\lambda F_B(\lambda)d\lambda = F$. $B(\lambda)$ is the surface albedo. Further, E denotes the white light source and S is the Fresnel reflectance. These are constant over the wavelengths assuming white or spectrally smooth illumination (i.e., approximately equal/smooth energy density for all wavelengths within the visible spectrum) and the neutral interface reflection (NIR) model (i.e., $S(\lambda)$ has a constant value independent of the wavelength). Consequently, we have $E(\lambda) = E$ and $S(\lambda) = S$. Further, $\vec{n}$ is the surface patch normal, $\vec{s}$ is the direction of the illumination source, and $\vec{v}$ is the direction of the viewer. Finally, geometric terms G_B and G_S denote the geometric dependencies on the body and surface reflection component.

Based on the measured RGB-values, the normalized color rg is computed by:

$$r = R/(R+G+B) \quad (2)$$
$$g = G/(R+G+B). \quad (3)$$

rg is a color invariant for matte surfaces by substituting the body reflection term of (1) in (2) [2]:

$$r(\omega_R,\omega_G,\omega_B) = \frac{\int_\lambda B(\lambda)F_R(\lambda)d\lambda}{\int_\lambda B(\lambda)F_R(\lambda)d\lambda + \int_\lambda B(\lambda)F_G(\lambda)d\lambda + \int_\lambda B(\lambda)F_B(\lambda)d\lambda}, \quad (4)$$

factoring out dependencies on illumination and object geometry and, hence, only dependent on the sensors and the surface albedo.

Further, we focus on the opponent color space defined by:

$$o_1(R,G,B) = (R-G)/2 \quad (5)$$
$$o_2(R,G,B) = \frac{2B-R-G}{4}. \quad (6)$$

The opponent color space is well-known and has its fundamentals in human perception. The opponent color space o_1o_2 is independent of highlights (assuming the NIR model) as follows from substituting (1) in (5) and (6):

$$o_1(\omega_R,\omega_G,\omega_B) = (G_B(\vec{n},\vec{s})E\int_\lambda B(\lambda)F_R(\lambda)d\lambda - G_B(\vec{n},\vec{s})E\int_\lambda B(\lambda)F_G(\lambda)d\lambda)/2. \quad (7)$$

Equal argument holds for o_2. Note that o_1o_2 is still dependent on $G_B(\vec{n},\vec{s})$ and E and, consequently, being sensitive to object geometry and shading.

The hue θ is computed as

$$\theta = \arctan\left(\frac{\sqrt{3}(G-B)}{(R-G)+(R-B)}\right), \quad (8)$$

also insensitive to illumination, object geometry, and highlights by substituting the reflection term of (1) in (8) [2]:

$$\theta(\omega_R,\omega_G,\omega_B) = \arctan\left(\frac{\sqrt{3}(\int_\lambda B(\lambda)F_G(\lambda)d\lambda - \int_\lambda B(\lambda)F_B(\lambda)d\lambda)}{(2\int_\lambda B(\lambda)F_R(\lambda)d\lambda - \int_\lambda B(\lambda)F_G(\lambda)d\lambda - \int_\lambda B(\lambda)F_G(\lambda)d\lambda)}\right). \quad (9)$$

3.2 Noise Propagation

Additive Gaussian noise is widely used to model thermal noise and is the limiting behavior of photon counting noise and film grain noise. Therefore, in this paper, we assume that sensor noise is normally distributed.

Then, for an indirect measurement, the true value of a measurand u is related to its N arguments, denoted by u_j, as follows:

$$u = q(u_1,u_2,\cdots,u_N). \quad (10)$$

Assume that the estimate $\hat{u}$ of the measurand u can be obtained by substitution of $\hat{u}_j$ for u_j. Then, when $\hat{u}_1,\cdots,\hat{u}_N$ are measured with corresponding standard deviations $\sigma_{\hat{u}_1},\cdots,\sigma_{\hat{u}_N}$, we obtain [15]

$$\hat{u} = q(\hat{u}_1,\cdots,\hat{u}_N). \quad (11)$$

It is known that the approximation of a given function can be written in the form of Taylor series. For $N=2$ (to simplify calculation), the Taylor series with respect to noise is given by

$$q(\hat{u}_1,\hat{u}_2) = q(u_1,u_2) + \left(\frac{\partial}{\partial u_1}\mathcal{E}_1 + \frac{\partial}{\partial u_2}\mathcal{E}_2\right)q(u_1,u_2) + \cdots + \frac{1}{m!}\left(\frac{\partial}{\partial u_1}\mathcal{E}_1 + \frac{\partial}{\partial u_2}\mathcal{E}_2\right)^m q(u_1,u_2) + R_{m+1}, \quad (12)$$

where $\hat{u}_1 = u_1 + \mathcal{E}_1$, $\hat{u}_2 = u_2 + \mathcal{E}_2$ ($\mathcal{E}_1$ and $\mathcal{E}_2$ are the errors of $\hat{u}_1$ and $\hat{u}_2$), and R_{m+1} is the remainder term. Further, $\partial q/\partial\hat{u}_j$ is the partial derivative of q with respect to $\hat{u}_j$.

As the general form of the error of an indirect measurement is

$$E = \hat{u} - u = q(\hat{u}_1,\hat{u}_2) - q(u_1,u_2), \quad (13)$$

we obtain in terms of the Taylor series the following:

$$E = \left(\frac{\partial}{\partial u_1}\mathcal{E}_1 + \frac{\partial}{\partial u_2}\mathcal{E}_2\right)q(u_1,u_2) + \cdots + \frac{1}{m!}\left(\frac{\partial}{\partial u_1}\mathcal{E}_1 + \frac{\partial}{\partial u_2}\mathcal{E}_2\right)^m q(u_1,u_2) + R_{m+1}. \quad (14)$$

In general, only the first linear term is used to compute the error

$$E = \frac{\partial q}{\partial u_1}\mathcal{E}_1 + \frac{\partial q}{\partial u_2}\mathcal{E}_2. \quad (15)$$

Then, for N arguments, it follows that if the uncertainties in $\hat{u}_1,\cdots,\hat{u}_N$ are independent, random, and relatively small, the predicted uncertainty in q is given by [15]

$$\sigma_q = \sqrt{\sum_{j=1}^{N}\left(\frac{\partial q}{\partial\hat{u}_i}\sigma_{\hat{u}_i}\right)^2}, \quad (16)$$

the so-called squares-root sum method. Although (16) is deduced for random errors, it is used as an universal formula for various kinds of errors.

Substitution of (2) and (3) in (16) gives the uncertainty for the normalized coordinates

$$\sigma_r = \sqrt{\frac{R^2(\sigma_B^2+\sigma_G^2) + (G+B)^2\sigma_R^2}{(R+G+B)^4}}$$
$$\sigma_g = \sqrt{\frac{G^2(\sigma_B^2+\sigma_R^2) + (R+B)^2\sigma_G^2}{(R+G+B)^4}}. \quad (17)$$

Assuming normally distributed random quantities, the standard way to calculate the standard deviations σ_R, σ_G, and σ_B is to compute the mean and variance estimates derived from homogeneously colored surface patches in an image under controlled imaging conditions. From the analytical study of (17), it can be

derived that normalized color becomes unstable around the black point $R = G = B = 0$.

The uncertainties of o_1 and o_2 are given by

$$\begin{aligned}\sigma_{o_1} &= \frac{1}{2}\sqrt{\sigma_G^2 + \sigma_R^2} \\ \sigma_{o_2} &= \frac{1}{4}\sqrt{4\sigma_B^2 + \sigma_G^2 + \sigma_R^2},\end{aligned} \tag{18}$$

which are the same (stable) at all RGB points.

Substitution of (8) in (16) gives the uncertainty for the hue

$$\sigma_\theta = \sqrt{\frac{3}{4}\frac{(B-G)^2\sigma_R^2 + (B-R)^2\sigma_G^2 + (R-G)^2\sigma_B^2}{\left(R^2 + G^2 + B^2 - RG - RB - GB\right)^2}}, \tag{19}$$

which is unstable at low intensity and saturation (i.e., the gray axis $R = G = B$).

In conclusion, it can be analytically derived that normalized color is unstable at low intensity. Hue is unstable at low intensity and saturation. Opponent color is relatively stable at all RGB values.

4 Histogram Construction by Variable Kernel Density Estimation

A density function f gives a description of the distribution of the measured data. A well-known density estimator is the histogram. The (one-dimensional) histogram is defined as

$$\hat{f}(x) = \frac{1}{nh}(\text{number of } X_i \text{ in the same bin as } x), \tag{20}$$

where n is the number of pixels with value X_i in the image, h is the bin width, and x is the range of the data. Two choices have to be made when constructing a histogram. First, the bin-width parameter needs to be chosen. Second, the position of the bin edges needs to be established. Both choices affect the resulting estimation.

Alternatively, the kernel density estimator is insensitive to the placement of the bin edges

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right). \tag{21}$$

Here, kernel K is a function satisfying $\int K(x)dx = 1$. In the *variable* kernel density estimator, the single h is replaced by n values $\alpha(X_i), i = 1, \cdots, n$. This estimator is of the form

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\alpha(X_i)} K\left(\frac{x - X_i}{\alpha(X_i)}\right). \tag{22}$$

The kernel centered on X_i has associated with it its own scale parameter $\alpha(X_i)$, thus allowing different degrees of smoothing. To use variable kernel density estimators for color images, we let the scale parameter be a function of the RGB-values and the color space transform. We are now left with the problem of determining the scale and shape of the kernel.

Assuming normally distributed noise, the distribution is approximated well by the Gauss distribution [15]

$$K(x) = \frac{1}{\sqrt{2\pi}}\exp^{\frac{-x^2}{2}}. \tag{23}$$

Then, the variable kernel method estimating the univariate, directional hue density is as follows:

$$\hat{f}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \sigma_{\theta_i}^{-1} K\left(\frac{(\theta - \theta_i) \bmod (\pi)}{\sigma_{\theta_i}}\right), \tag{24}$$

where σ_θ is derived according to (19), with corresponding value θ.

The variable kernel method for the bivariate normalized rg kernel is given by:

$$\hat{f}(r,g) = \frac{1}{n}\sum_{i=1}^{n} \sigma_{r_i}^{-1} K\left(\frac{r - r_i}{\sigma_{r_i}}\right)\sigma_{g_i}^{-1} K\left(\frac{g - g_i}{\sigma_{g_i}}\right), \tag{25}$$

where σ_r, σ_g are derived according to (17). Similarly, the variable kernel method for the bivariate normalized o_1o_2 kernel is given by:

$$\hat{f}(o_1,o_2) = \frac{1}{n}\sum_{i=1}^{n} \sigma_{o_{1_i}}^{-1} K\left(\frac{o_1 - o_{1_i}}{\sigma_{o_{1_i}}}\right)\sigma_{o_{2_i}}^{-1} K\left(\frac{o_2 - o_{2_i}}{\sigma_{o_{2_i}}}\right), \tag{26}$$

where σ_{o_1} and σ_{o_2} are derived according to (18).

In conclusion, to reduce the effect of sensor noise during density estimation, we use variable kernels where the normal distribution defines the shape of the kernel. Further, kernel sizes are steered by the amount of uncertainty of the color invariant values.

5 Experiments

In this section, the performance of the proposed variable kernel density estimator will be evaluated. First, the accuracy of noise propagation is empirically verified. Then, the kernel density estimation is experimentally compared to traditional histogram schemes in the context of color based object recognition. For the experiments, the hue range is defined from 0° to 360° over a 1° interval. The normalized color range is defined from 0 to 255 units over 1 unit intervals. The images are obtained using a Sony XC-003P color camera and Matrox Corona framegrabber.

5.1 Propagation of Uncertainties

The aim of this experiment is to empirically verify the validity of the proposed model of noise propagation through color invariant formulae. For nonlinear functions, such as rg and θ, the estimation could be slightly biased, since only the first term of the Taylor series are taken to approximate the uncertainty, see (15) in Section 3.2. Further, although the Sony XC-003P color camera have narrow-band filters, they may still overlap partially, introducing correlation between tristimulus value measurements. To test this, the predicted uncertainty σ_θ of the hue space is computed first for each color pixel according to (19). Then, the measured (actual) uncertainty is computed as the standard deviation of hue values recorded 10 times for nine different colors, i.e., the experiment is conducted on nine different homogeneously colored sheets of paper material. Sheet number 1 has a bright red color, number 2 is red colored, 3 is yellow, 4 is light green, 5 is green, 6 is cyan, 7 is darkblue, 8 is blue, and 9 is purple. The predicted and actual uncertainties are shown in Fig. 1. From this experiment, we obtain that the difference between the predicted (computed by (19)) and measured hue values is $0.07^\circ \pm 0.03^\circ$, which is well below 1 percent of the hue range. Further, the experiment is repeated for normalized colors where the uncertainty is computed according to (17). The difference obtained between the predicted and measured normalized red values was 0.4 ± 0.2 and, for the green values, we obtained 0.2 ± 0.1. This is again well below 1 percent of the normalized color range. The experiment shows that the predicted uncertainties compare favorably to the measured (actual) uncertainties.

5.2 Color-Based Object Recognition

In this section, we consider object recognition on the basis of color invariant histograms. Therefore, in Section 5.2.1, the data set and performance measure are discussed. Then, in Sections 5.2.2 and 5.2.3, we compare traditional histogram-based object construction with our variable kernel density estimation in the context of object recognition.

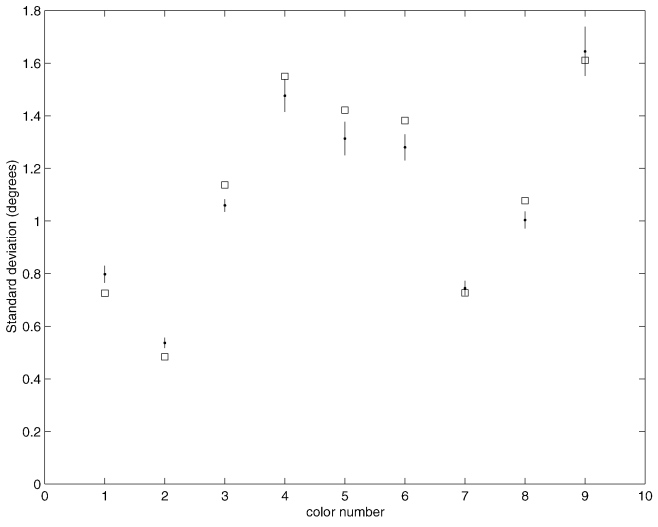

Fig. 1. Analysis of the theoretically predicted uncertainty versus the actual uncertainty of the hue. The squares correspond to actual standard deviations of the hue, and the error bars correspond to the predicted standard deviations.

5.2.1 Data Set and Performance Measure

For comparison reasons, the same data set used by [2], [16], has been taken to conduct the experiments. In Fig. 2, various images from the image database are shown. These images are recorded by the SONY XC-003P CCD color camera and the Matrox Magic Color frame grabber. Two light sources of average daylight color are used to illuminate the objects in the scene. The database consists of $N_1 = 500$ target images taken from colored objects, tools, toys, food cans, art artifacts, etc. Objects were recorded in isolation (one per image), i.e., 500 images are recorded from 500 different objects. The size of the images are 256×256 with 8 bits per color. The images show a considerable amount of shadows, shading, and highlights. A second, independent set (the query set) of $N_2 = 70$ query or test recordings was made of randomly chosen objects already in the database. These objects were recorded again one per image with a new, arbitrary position and orientation with respect to the camera, some recorded upside down, some rotated, and some at different distances.

Then, for each image, traditional histograms (c.f. (20)) and histograms based on variable density estimation are constructed on the basis of the rg and θ-space. For the histograms, we have determined the appropriate bin size for our application empirically by varying the number of bins on the axes over $q \in \{2, 4, 8, 16, 32, 64, 128, 256\}$. The results show (not presented here) that the number of bins was of little influence on the retrieval accuracy when the number of bins ranges from $q = 32$ and up. Therefore, the color histogram bin size for each axis used during histogram formation is $q = 32$.

For a measure of match quality, let rank r^{Q_i} denote the position of the correct match for test image Q_i, $i = 1, \ldots, N_2$, in the ordered list of N_1 match values. The rank r^{Q_i} ranges from $r = 1$ from a perfect match to $r = N_1$ for the worst possible match.

Then, for one experiment, the average ranking percentile is defined by:

$$\overline{r} = \left(\frac{1}{N_2} \sum_{i=1}^{N_2} \frac{N_1 - r^{Q_i}}{N_1 - 1} \right) 100\%. \tag{27}$$

In the remaining sections, we study the performance of the variable kernel estimator with respect to noise for the 70 test images and 500 target images. For comparison reasons in the literature, matching is based on histogram intersection [3].

5.2.2 Robustness Against Noise: Simulated Data

The effect of noise is produced by adding independent zero-mean additive Gaussian noise with $\sigma \in \{2, 4, 8, 16, 32, 64\}$ to the query images. In Fig. 3, two images are shown generating together 10 images by adding noise with $\sigma \in \{8, 16, 32, 64, 128\}$.

We concentrate on the quality of the recognition rate with respect to different noise levels. To compare traditional histogram matching with histogram matching based on the proposed kernel density estimator, we have constructed four different histograms. First, no thresholding has been performed. This histogram construction scheme does not cope with unstable color invariant values. Hence, all color invariant values are equally weighted in the histogram, as used by [3]. The color histogram without thresholding is denoted by $\mathcal{H}_{\theta_1}$ based on the hue θ color model and $\mathcal{H}_{rg_1}$ for the rg color model. Second, we have discarded rg and θ values when the intensity was below 5 percent of the total range as proposed by [2], [8]. For this histogram construction scheme, we denote $\mathcal{H}_{\theta_2}$ based on θ and $\mathcal{H}_{rg_2}$ derived from rg. Third, rg and hue

Fig. 2. Various images which are included in the image database of 500 images. The images are representative for the images in the database. Objects were recorded in isolation (one per image).

Fig. 3. Two images generating together 10 images by adding noise with $\sigma \in \{8, 16, 32, 64, 128\}$.

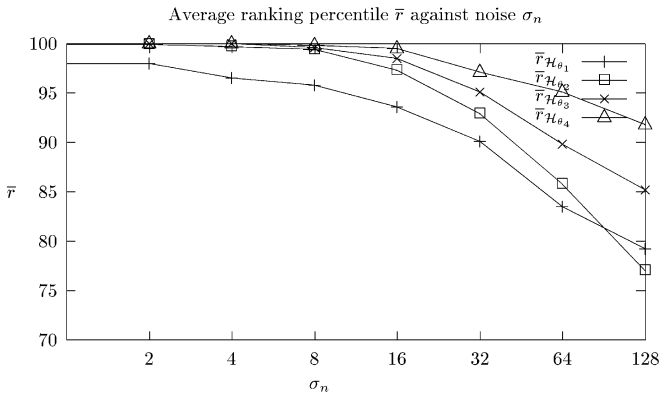

Fig. 4. The discriminative power of the matching process differentiated for the various histogram construction schemes based on θ with respect to noise. The average percentile $\overline{r}$ for histogram $\mathcal{H}_{\theta_1}$, $\mathcal{H}_{\theta_2}$, $\mathcal{H}_{\theta_3}$, and $\mathcal{H}_{\theta_4}$ is given by $\overline{r}_{\mathcal{H}_{\theta_1}}$, $\overline{r}_{\mathcal{H}_{\theta_2}}$, $\overline{r}_{\mathcal{H}_{\theta_3}}$, and $\overline{r}_{\mathcal{H}_{\theta_4}}$, respectively.

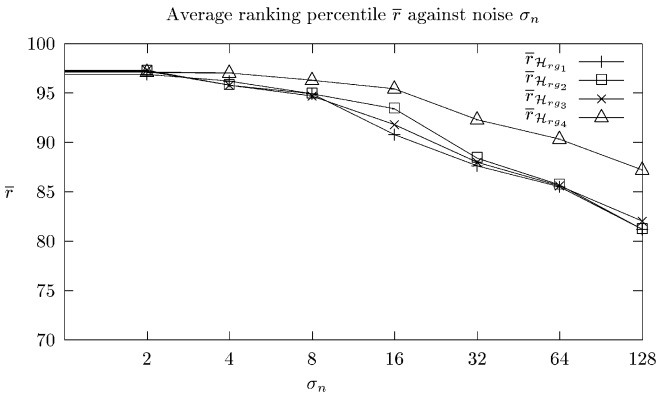

Fig. 5. The discriminative power of the matching process differentiated for the various histogram construction schemes based on rg with respect to noise. The average percentile $\overline{r}$ for histogram $\mathcal{H}_{rg_1}$, $\mathcal{H}_{rg_2}$, $\mathcal{H}_{rg_3}$, and $\mathcal{H}_{rg_4}$ is given by $\overline{r}_{\mathcal{H}_{rg_1}}$, $\overline{r}_{\mathcal{H}_{rg_2}}$, $\overline{r}_{\mathcal{H}_{rg_3}}$, and $\overline{r}_{\mathcal{H}_{rg_4}}$, respectively.

values were discarded during histogram construction when the intensity and saturation were within the range of 4σ centered at the origin of the RGB space, used by [9], yielding $\mathcal{H}_{\theta_3}$ and $\mathcal{H}_{rg_3}$. Finally, the histogram based on the proposed variable kernel density estimator is given by $\mathcal{H}_{\theta_4}$ and $\mathcal{H}_{rg_4}$.

A drawback of constructing histograms based on kernel density estimation, compared to traditional histogram construction schemes, is that the method is more computational expensive. The time to compute a traditional histogram for an image of 256×256 pixels is on average 0.2 seconds on a Ultra 10 Sparc station. The time required to construct a histogram based on kernel density estimation is on average 2.2 seconds (a factor of eleven) on a Ultra 10 Sparc station. Note that histograms, for images in the data set, are constructed offline. Online execution time includes the construction of the histogram for the query image, and the matching based on the histogram intersection. The storage complexity, based on the number of bins ($\theta = 32$ and $rg = 32$ x 32), is the same for the different histogram construction schemes. The influence of noise differentiated by the various histogram construction schemes, shown in Fig. 4, based on the hue color model, reveals that kernel density estimation outperforms the ad hoc thresholding schemes. In fact, the histogram intersection based on kernel density estimation gives good results up to considerable amounts of noise ($\sigma = 64$). Further, the thresholded histogram construction schemes always give higher recognition accuracy than no thresholding at all. Further, on the basis of the rg color model, the impact of noise differentiated by the various histogram construction schemes is shown in Fig. 5. Again, the kernel density estimator provides higher recognition accuracy than the ad hoc thresholding schemes. However, the thresholding schemes give similar recognition accuracy than no thresholding at all. This is due to the fact that normalized color becomes unstable around the black point. Hence, thresholding on saturation does not affect the instability of rg. In contrast, $\mathcal{H}_{rg_3}$ also eliminates low intensity values. Is seems that eliminating dark regions does not affect the recognition rate significantly. Finally, from the experimental results of Figs. 4 and 5, it can be concluded that, globally, rg-based object recognition gives slightly worse recognition accuracy than recognition based on θ.

5.2.3 Robustness Against Noise: Realistic Data

To measure the sensitivity of different histogram construction schemes with respect to varying SNR, 10 objects were randomly chosen from the image dataset. Then, each object has been recorded again under a global change in illumination intensity (i.e., dimming the light source) generating images with $\mathrm{SNR} \in \{24, 12, 6, 3\}$, see Fig. 6. These low-intensity images can be seen as images of snap shot quality, a good representation of views from everyday life as it appears in home video, the news, and consumer digital photography in general. Matching based on the tradition histogram construction scheme, computed for rg, is denoted by $\mathcal{H}_{rg_T}$, and for θ, we obtain $\mathcal{H}_{\theta_T}$. For fair comparison, thresholding has been applied on the images (not on the query image) and consequently rg-values and θ are discarded when the intensity was below 5 percent of the total range. The kernel density estimation, based on rg, is denoted by $\mathcal{H}_{rg_K}$, and for θ, we have $\mathcal{H}_{\theta_K}$. The discriminative power of the histogram matching process based on rg and θ differentiated for the different histogram construction methods plotted against the amount of SNR is shown in Fig. 7.

For $24 <$ SNR < 48, the results show a rapid decrease in the performance of the traditional method as opposed the kernel density estimation. For these SNR's, the kernel density estimation outperforms the traditional histogram construction scheme. For SNR < 12, the performance of both methods decrease in the same way, where the performance of the kernel density estimation remains slightly higher than the traditional histogram matching. This is due to quantization errors for very low intensity pixels which disturb the underlying Gaussian noise model. In fact, quantization errors are caused by reducing the image intensity and consequently limiting the range of RGB color values from which the color invariants are computed. To this end, only a reduced number of different color invariant values can be generated for

Fig. 6. Two objects under varying illumination intensity generating each four images with $\mathrm{SNR} \in \{24, 12, 6, 3\}$.

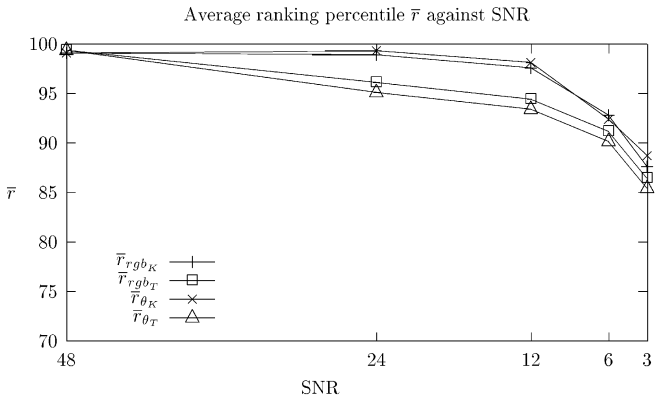

Fig. 7. The discriminative power of the matching process, differentiated for the traditional histogram and kernel density estimation scheme, based on rg and θ with respect to SNR.

which the assumption of a Gaussian noise model is not valid anymore.

In conclusion, the kernel density estimator outperforms the traditional histogram method up to considerable amounts of noise (SNR $= 12$). However, for very low-intensity images (SNR < 12), due to quantization errors, the kernel density estimation behaves the same as traditional histogram methods.

6 Conclusion

In this paper, variable kernel density estimation is used to construct robust color invariant histograms. The variable kernel density estimation is derived from a theoretical framework for noise propagation through color invariants. In this way, the associated uncertainty is computed for each color invariant value, which is used to steer the kernel sizes. From the theoretical and experimental result, we conclude that kernel density estimator overcome the problem of ad hoc thresholding at unstable color invariants. Further, our method is less sensitive to Gaussian noise than traditional histogram construction schemes. A drawback of the variable kernel method compared to traditional histogram construction is that the method is computationally more expensive.

Acknowledgments

The authors are grateful to Arnold Smeulders and the anonymous reviewers for their valuable comments.

References

[1] B.V. Funt and G.D. Finlayson, "Color Constant Color Indexing," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 5, pp. 522-529, May 1995.
[2] T. Gevers and A.W.M. Smeulders, "Color Based Object Recognition," *Pattern Recognition,* vol. 32, no. 3, pp. 453-464, 1999.
[3] M.J. Swain and D.H. Ballard, "Color Indexing," *Int'l J. Computer Vision,* vol. 7, no. 1, pp. 11-32, 1991.
[4] D. Berwick and S.-W. Lee, "A Chromaticity Space for Specularity-, Illumination color- and Illumination Pose-Invariant 3-D Object Recognition," *Proc. Int'l Conf. Computer Vision,* pp. 165-170, 1998.
[5] J.L. Crowley, G.D. Finlayson, and B. Schiele, "Comprehensive Colour Image Normalization," *Proc. European Conf. Computer Vision,* pp. 475-490, 1998.
[6] S.K. Nayar and R.M. Bolle, "Reflectance Based Object Recognition," *Int'l J. Computer Vision,* vol. 17, no. 3, pp. 219-240, 1996.
[7] J.R. Kender, "Saturation, Hue and Normalized Color: Calculation, Digitation Effects, and Use," technical report, Carnegie-Mellon Univ., 1976.
[8] Y. Ohta, T. Kanade, and T. Sakai, "Color Information for Region Segmentation," *Computer Graphics and Image Processing,* vol. 13, pp. 222-241, 1980.
[9] G. Healey, "Segmenting Images Using Normalized Color," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 22, no. 1, pp. 64-73, 1992.
[10] P.D. Burns and R.S. Berns, "Error Propagation Analysis in Color Measurement and Imaging," *Color Research and Applications,* vol. 22, no. 4, Aug. 1997.
[11] L. Shafarenko, M. Petrou, and J. Kittler, "Histogram-Based Segmentation in a Perceptually Uniform Color Space," *IEEE Trans. Image Processing,* vol. 7, no. 9, pp. 1354-1358, 1998.
[12] B.W. Silverman, *Density Estimation.* London: Chapman and Hall, 1986.
[13] M.P. Wand and M.C. Jones, *Kernel Smoothing.* London: Chapman and Hall, 1995.
[14] S.A. Shafer, "Using Color to Separate Reflection Components," *Color Research and Applications,* vol. 10, no. 4, pp. 210-218, 1985.
[15] J.R. Taylor, *An Introduction to Error Analysis.* Mill Valley, Calif.: Univ. Science Books, 1982.
[16] N. Sebe, M.S. Lew, and D.P. Huijsmans, "Toward Improved Ranking Metrics," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 10, pp. 1132-1143, Apr. 2000.