Robust Histogram Construction from Color Invariants for Object Recognition

Gevers, Th.; Stokman, H.M.G.

Published in:
IEEE Transactions on Pattern Analysis and Machine Intelligence

DOI:
10.1109/TPAMI.2004.1261083

Citation for published version (APA):
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 images 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 colorized color reflection model, Gevers and Smeulders [2] showed that normal-
color invariants [1], [2], [3], [4], [5]. For example, illumination-
tations), color histograms are usually computed from photometric conditions (e.g., illumination, shading, highlights, and interreflec-

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 colorized color reflection model, Gevers and Smeulders [2] showed that normal-
color invariants [1], [2], [3], [4], [5]. For example, illumination-
tations), color histograms are usually computed from photometric conditions (e.g., illumination, shading, highlights, and interreflec-

keep the color space from inhomogeneous dielectric materials under white, spectrally smooth illumination or a white-balanced camera.

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]

\[
\omega_k = G_d(\mathbf{n}, \mathbf{s}) E \int \lambda B(\lambda) F_c(\lambda) d\lambda + G_S(\mathbf{n}, \mathbf{s}, \lambda) ESF,
\]

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 \(F_B(\lambda), F_G(\lambda), \text{ and } F_R(\lambda)\), respectively, where \(\lambda\) denotes the wavelength. We assume that the integration white conditions hold, i.e., \(\int \lambda F_B(\lambda) d\lambda = \int \lambda F_G(\lambda) d\lambda = \int \lambda F_R(\lambda) d\lambda = F(\lambda)\) for
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/smeared 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, $n$ is the surface patch normal, $\bar{v}$ is the direction of the viewerm, and finally, geometric terms $G_R$ and $G_G$ denote the geometric dependencies on the body and surface reflection component.

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

$$r = R/(R + G + B)$$

$$g = G/(R + G + B).$$

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

$$r(\omega_B, \omega_G, \omega_B) = \frac{\int_B B(\lambda)F_B(\lambda)d\lambda}{\int_B B(\lambda)F_B(\lambda)d\lambda + \int_F B(\lambda)F_C(\lambda)d\lambda + \int_F B(\lambda)F_B(\lambda)d\lambda}$$

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:

$$o_1(R, G, B) = (R - G)/2$$

$$o_2(R, G, B) = \frac{2B - R - G}{4}.$$  

The opponent color space is well-known and has its fundamentals in human perception. The opponent color space $o_{12}$ is independent of highlights (assuming the NIR model) as follows from substituting (1) in (5) and (6):

$$o_1(\omega_B, \omega_G, \omega_B) = (G_B(\bar{n}, \bar{v})E \int_B B(\lambda)F_R(\lambda)d\lambda - G_B(\bar{n}, \bar{v})E \int_B B(\lambda)F_C(\lambda)d\lambda)/2.$$  

Equal argument holds for $o_2$. Note that $o_{12}$ is still dependent on $G_B(\bar{n}, \bar{v})$ and $E$ and, consequently, being sensitive to object geometry and shading.

The hue $\theta$ is computed as

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

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

$$\theta(\omega_B, \omega_G, \omega_B) = \text{arctan} \left( \frac{\sqrt{3}((\int_B B(\lambda)F_C(\lambda)d\lambda - \int_B B(\lambda)F_B(\lambda)d\lambda)}{2 \int_B B(\lambda)F_R(\lambda)d\lambda - \int_B B(\lambda)F_C(\lambda)d\lambda - \int_B B(\lambda)F_B(\lambda)d\lambda} \right).$$

### 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:

$$u = q(u_1, u_2, \ldots, u_N).$$

Assume that the estimate $\hat{u}$ of the measurand $u$ can be obtained by substitution of $\hat{u}_j$ for $u_j$. Then, when $\hat{u}_1, \ldots, \hat{u}_N$ are measured with corresponding standard deviations $\sigma_{u_1}, \ldots, \sigma_{u_N}$, we obtain [15]

$$\hat{u} = q(\hat{u}_1, \ldots, \hat{u}_N).$$

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

$$q(\hat{u}_1, \hat{u}_2) = q(u_1, u_2) + \left( \frac{\partial}{\partial u_1} + \frac{\partial}{\partial u_2} \right) q(u_1, u_2) + \cdots$$

$$+ \frac{1}{m!} \left( \frac{\partial}{\partial u_1} + \frac{\partial}{\partial u_2} \right)^m q(u_1, u_2) + R_{m+1},$$

where $\hat{u}_1 = u_1 + \epsilon_{11}, \hat{u}_2 = u_2 + \epsilon_{12} (\epsilon_{11}$ and $\epsilon_{12}$ are the errors of $\hat{u}_1$ and $\hat{u}_2$), and $R_{m+1}$ is the remainder term. Further, $\partial q/\partial u_j$ is the partial derivative of $q$ with respect to $u_j$.

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

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

we obtain in terms of the Taylor series the following:

$$E = \left( \frac{\partial}{\partial u_1} + \frac{\partial}{\partial u_2} \right) q(u_1, u_2) + \cdots$$

$$+ \frac{1}{m!} \left( \frac{\partial}{\partial u_1} + \frac{\partial}{\partial u_2} \right)^m q(u_1, u_2) + R_{m+1}.$$  

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

$$E = \frac{\partial q}{\partial u_1} \epsilon_{11} + \frac{\partial q}{\partial u_2} \epsilon_{12}.$$  

Then, for $N$ arguments, it follows that if the uncertainties in $u_1, \ldots, u_N$ are independent, random, and relatively small, the predicted uncertainty in $q$ is given by [15]

$$\sigma_q = \sqrt{\sum_{k=1}^{N} \left( \frac{\partial q}{\partial u_k} \sigma_{u_k} \right)^2},$$

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

$$\sigma_r = \sqrt{\frac{R^2(\sigma_{B1}^2 + \sigma_{C1}^2) + (G + B)^2 \sigma_{B1}^2}{(R + G + B)^4}},$$

$$\sigma_g = \sqrt{\frac{G^2(\sigma_{B1}^2 + \sigma_{C1}^2) + (R + B)^2 \sigma_{B1}^2}{(R + G + B)^4}},$$

Assuming normally distributed random quantities, the standard way to calculate the standard deviations $\sigma_{u_1}, \sigma_{u_2},$ and $\sigma_{u_3}$ 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 $u_1$ and $u_2$ are given by

$$\sigma_{u_1} = \frac{1}{2} \sqrt{\sigma_{u_1}^2 + \sigma_{u_2}^2},$$

$$\sigma_{u_2} = \frac{1}{2} \sqrt{4\sigma_{u_1}^2 + \sigma_{u_2}^2 + \sigma_{u_3}^2},$$

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

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

$$\sigma_\theta = \sqrt{\frac{3(B - G)^2 \sigma_{B1}^2 + (B - R)^2 \sigma_{C1}^2 + (R - G)^2 \sigma_{B1}^2}{4(R^2 + G^2 + B^2 - RG - RB - GB)^2}},$$
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

\[
\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h} \right)
\]

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

\[
\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K \left( \frac{x - X_i}{h} \right)
\]

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 \( \alpha(X_i), i = 1, \ldots, n \). This estimator is of the form

\[
\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\alpha(X_i)} K \left( \frac{x - X_i}{\alpha(X_i)} \right)
\]

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]

\[
K(x) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right)
\]

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

\[
\hat{f}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sigma_{\theta_i}} K \left( \frac{\theta - \theta_i}{\sigma_{\theta_i}} \right)
\]

where \( \sigma_{\theta} \) is derived according to (19), with corresponding value \( \theta \).

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

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

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

\[
\hat{f}(o_1, o_2) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sigma_{o_1i}} K \left( \frac{o_1 - o_{1i}}{\sigma_{o_1i}} \right) \frac{1}{\sigma_{o_2i}} K \left( \frac{o_2 - o_{2i}}{\sigma_{o_2i}} \right)
\]

where \( \sigma_{o_1} \) and \( \sigma_{o_2} \) 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 \( \theta \), the estimation could be slightly biased, since only the first term of the Taylor series could be slightly biased, since only the first term of the Taylor series is 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 \( \sigma_{\theta} \) 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° ± 0.03°, 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
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 $\mathcal{H}_0$ based on the hue $\theta$ color model and $\mathcal{H}_g$ for the rg color model. Second, we have discarded rg and $\theta$ values when the intensity was below 5 percent of the total range as proposed by [2], [8]. For this histogram construction scheme, we denote $\mathcal{H}_0$ based on $\theta$ and $\mathcal{H}_g$, derived from $rg$. Third, $rg$ and hue values were discarded during histogram construction when the intensity and saturation were within the range of $4\sigma$ centered at the origin of the $RGB$ space, used by [9], yielding $\mathcal{H}_0$ and $\mathcal{H}_g$. Finally, the histogram based on the proposed variable kernel density estimator is given by $\mathcal{H}_0$ and $\mathcal{H}_g$.

A drawback of constructing histograms based on kernel density estimation, compared to traditional histogram construction schemes, is that the method is more computationally expensive. The time to compute a traditional histogram for an image of $256 \times 256$ 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.

For a measure of match quality, let rank $r_i$ 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 $r_i$ 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:

$$\tau = \left( \frac{1}{N_2} \sum_{i=1}^{N_2} \frac{N_1 - r_i}{N_1 - 1} \right) \times 100\%.$$  \hspace{1cm} (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.
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, $H_{rg}$ 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 $\theta$.

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 $\text{SNR} \in \{24, 12, 6, 3\}$, see Fig. 6. These low-intensity images can be seen as images of snapshot 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 $H_{rg}$, and for $\theta$, we obtain $H_{\theta}$. For fair comparison, thresholding has been applied on the images (not on the query image) and consequently $rg$-values and $\theta$ are discarded when the intensity was below 5 percent of the total range. The kernel density estimation, based on $rg$, is denoted by $H_{rg}$, and for $\theta$, we have $H_{\theta}$. The discriminative power of the histogram matching process based on $rg$ and $\theta$ differentiated for the different histogram construction methods plotted against the amount of SNR is shown in Fig. 7.

For $24 < \text{SNR} < 48$, 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 $\text{SNR} < 12$, 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 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 ($\text{SNR} = 12$). However, for very low-intensity images ($\text{SNR} < 12$), 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 Smeluders and the anonymous reviewers for their valuable comments.
Fig. 7. The discriminative power of the matching process, differentiated for the traditional histogram and kernel density estimation scheme, based on $r_g$ and $\theta$ with respect to SNR.

REFERENCES


For more information on this or any other computing topic, please visit our Digital Library at http://computer.org/publications/dlib.