

Coloring Local Feature Extraction

Joost van de Weijer and Cordelia Schmid

GRAVIR-INRIA, 655 Avenue de l'Europe, Montbonnot 38330, France
{Joost.van-de-Weijer, Cordelia.Schmid}@inrialpes.fr

Abstract. Although color is commonly experienced as an indispensable quality in describing the world around us, state-of-the art *local* feature-based representations are mostly based on shape description, and ignore color information. The description of color is hampered by the large amount of variations which causes the measured color values to vary significantly. In this paper we aim to extend the description of **local** features with color information. To accomplish a wide applicability of the color descriptor, it should be robust to : 1. photometric changes commonly encountered in the real world, 2. varying image quality, from high quality images to snap-shot photo quality and compressed internet images. Based on these requirements we derive a set of color descriptors. The set of proposed descriptors are compared by extensive testing on multiple applications areas, namely, matching, retrieval and classification, and on a wide variety of image qualities. The results show that color descriptors remain reliable under photometric and geometrical changes, and with decreasing image quality. For all experiments a combination of color and shape outperforms a pure shape-based approach.

1 Introduction

There exists broad agreement that local features are an efficient tool for object representation due to their robustness with respect to occlusion and geometrical transformations [1]. A typical application based on local features starts with the detection phase, in which features are localized. If desired the patches are transformed to be invariant with respect to orientation, scale, and affine transformations (see Fig. 1). Invariant representations are subsequently extracted by a descriptor. The descriptor should robustly represent both the shape and the color of the features. A considerable amount of research has been dedicated to robust local shape descriptors. An extensive study by Mikolajczyk and Schmid [2] reported the SIFT descriptor [3] to perform best. The description of local color has received relatively little attention, and as a result most local features-based methods [3],[4],[5] use only luminance and ignore color information. The aim of this article is to enrich local feature-based methods with color information.

A lot of work has been dedicated to global color features for color object recognition. Ballard and Swain [6] described objects by their color histograms. Moreover, to obtain invariance with respect to lighting geometry the use of normalized *rgb* histograms was advocated. This method remained however variant with respect to illuminant changes. To tackle this problem Funt and Finlayson

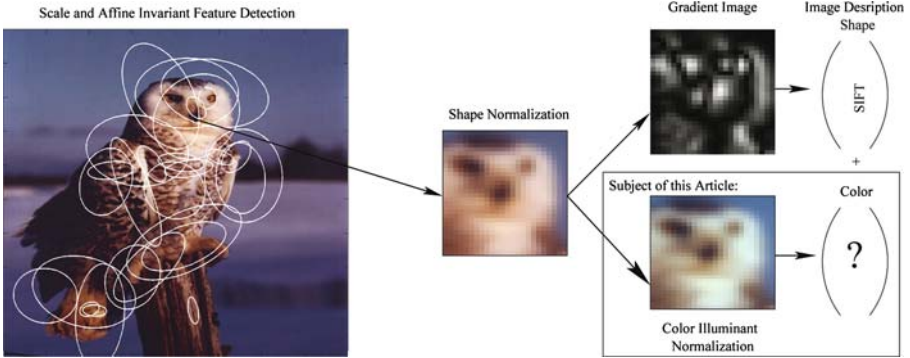


Fig. 1. Overview of a local feature-based method divided in a feature detection and a feature description phase. The aim of this article is to enrich the local feature description with color information.

[7] proposed an illuminant invariant indexing method, which was however variant with respect to lighting geometry. Finlayson et al. [8] combined the theories of [6] and [7] and proposed a indexing method which is both invariant to shading and illuminant changes. All methods remained however variant with respect to specularities. Gevers and Smeulders [9] propose invariants for specularity, in combination with illuminant and lighting geometry. The work was later extended to the derivative structure of images in [10], [11], leading to e.g. photometric invariant edge and corner detection. Furthermore, Gevers and Stokman [12] observed that instabilities, caused by the non-linear transformation to compute the photometric invariants, hamper practical use of photometric invariance theory. Based on an error analysis robust photometric invariants are proposed.

We extend local feature descriptors with color information, by concatenating a color descriptor, \mathbf{K} , to the shape descriptor, \mathbf{S} , according to

$$\mathbf{B} = \left(\hat{\mathbf{F}}, \lambda \hat{\mathbf{K}} \right) \tag{1}$$

where \mathbf{B} is the combined color and shape descriptor and λ is a weighting parameter, and $\hat{\cdot}$ indicates that the vector is normalized. For the shape description we rely on the SIFT descriptor [3]. Since the color descriptor is to be used in combination with a shape descriptor it does not need to contain any spatial information, which leads us to use local histograms. From the analysis of the color literature, discussed above, we deduce the following criteria to which these local color histograms should adhere:

1. photometric robustness: the descriptor should be robust to photometric variations such as shadow, shading, specularities and changes of the light source.
2. geometric robustness: the descriptor should be invariant with respect to geometrical changes, such as viewpoint, zoom, and object orientation variations.
3. photometric stability: the descriptor should adequately handle the instabilities introduced by photometric invariant transformations.

4. generality: the color descriptor should be applicable to a wide variety of applications such as matching, retrieval and classification. Furthermore, it should be robust to variations in image quality, from high quality images to compressed snapshot quality images.

After discussing a physical reflectance model in section 2, we design several color feature descriptors in accordance with the four criteria, in section 3. In section 4 experimental results are given and section 5 contains concluding remarks.

Related to the research proposed in this paper is the work of Mindru et al. [13]. They propose a combined color and shape description of the local neighborhood based on color moments, which are invariant to illuminant color. Since, we aim for a color description, which will be used in combination with the SIFT shape description, we have not pursued this path. Furthermore, in [14] local moments-based descriptors were found to be relatively unstable. Matas et al. [15] describe a method which, based on the modes in the local histogram, computes invariant signatures. The method uses fixed scales, and seems hard to use as an extension to a scale invariant feature detector, where the number of modes for a single feature is often higher than the two or three discussed in [15].

2 Color Preliminaries

In this section the color theory needed for the design of the color descriptors is summarized. We assume that the scene consists of inhomogeneous materials (including e.g. papers and plastics) and we exclude homogeneous materials such as metals. Furthermore, we model the light source locally, i.e. for the extend of a single feature, as a single light source, $e(\lambda)$, where λ is the wavelength. For multiple light sources we assume that the combination can be approximated as a single light source for the local feature. In this case, the measured values, $C \in \{R, G, B\}$, of the camera with spectral sensitivities f^C , are modelled [16] by integrating over the visible spectrum ω ,

$$C(\mathbf{x}) = m^b(\mathbf{x}) \int_{\omega} b(\lambda, \mathbf{x}) e(\lambda) f^C(\lambda) d\lambda + m^i(\mathbf{x}) \int_{\omega} i(\lambda) e(\lambda) f^C(\lambda) d\lambda. \quad (2)$$

The reflection of the light consist of two parts: 1. a body reflection part, which describes the light which is reflected after interaction with the surface albedo b , and 2. the interface reflection which describes the part of the light that is immediately reflected at the surface, causing specularities. We assume neutral interface reflection, meaning that the Fresnel reflectance i is independent of λ . Accordingly, we will omit i in further equations. The geometric dependence of the reflectance is described by the terms m^b and m^i which depend on the viewing angle, light source direction and surface orientation. \mathbf{x} denotes the spatial coordinates, and bold face is used to indicate vectors.

Ambient or diffuse light, i.e. light coming from all directions, is not modelled by Eq. 2 [16]. Diffuse light occurs in outdoor scenes where there is next to the dominant illuminant, i.e. the sun, diffuse light coming from the sky. Similarly,

it occurs in indoor situations where diffuse light is caused by reflectances from walls and ceilings. Shafer [16] models the diffuse light, a , by a third term

$$C(\mathbf{x}) = m^b(\mathbf{x}) \int_{\omega} b(\lambda, \mathbf{x}) e(\lambda) f^C(\lambda) d\lambda + m^i(\mathbf{x}) \int_{\omega} e(\lambda) f^C(\lambda) d\lambda + \int_{\omega} a(\lambda) f^C(\lambda) d\lambda \quad (3)$$

The camera sensitivities, f^C , can be approximated as delta functions, thereby simplifying the reflection function to

$$C(\mathbf{x}) = m^b(\mathbf{x}) b^C(\mathbf{x}) e^C + m^i(\mathbf{x}) e^C + a^C. \quad (4)$$

This function together with its derivative,

$$C_{\mathbf{x}}(\mathbf{x}) = m_x^b(\mathbf{x}) b^C(\mathbf{x}) e^C + m^b(\mathbf{x}) b_x^C(\mathbf{x}) e^C + m_x^i(\mathbf{x}) e^C, \quad (5)$$

will be used in the following sections to derive photometric invariants. Throughout the paper we will use a subscript to indicate spatial differentiation, and we use boldface to indicate vectors over the three channels, e.g. $\mathbf{C} = \{R, G, B\}$.

3 Color Feature Description

In this section we derive a set of color descriptors in accordance with the requirements put forward in section 1. Robustness with respect to photometric variation is discussed in section 3.1 and 3.2. The second criterion, geometrical robustness is handled in section 3.3. Photometric stability issues raised by criterion 3 are handled in section 3.4.

3.1 Photometric Robustness: Color Constancy

In section 2 we derived how the measured sensor values depend on both the color of the illuminant interacting with the object, and the color of the diffuse illuminant. In this section two simple algorithms [17], [18] are described for color illuminant normalization (see method overview in Fig .1).

We first consider the case for which there is no diffuse illuminant present ($a^C = 0$). The relation between two images of the same scene, \mathbf{C}^1 and \mathbf{C}^2 , taken under different illuminants, is modelled by a scalar multiplication, since

$$C^2(\mathbf{x}) = (m^b(\mathbf{x}) b^C(\mathbf{x}) + m^i(\mathbf{x})) e_2^C = \lambda^C C^1(\mathbf{x}) \quad (6)$$

where $\lambda^C = e_2^C / e_1^C$. The colors in the two scenes are hence related by a diagonal matrix $\mathbf{C}_2 = \mathbf{\Lambda} \mathbf{C}_1$. This diagonal matrix relation is well-known and is a consequence of assuming delta functions for the sensitivities of the cameras. Although the delta function assumption seems rather blunt it describes reality surprisingly well [19]. From Eq. 6 it is easily proven that, in the absence of diffuse light, invariance with respect to the illuminant can be obtained, by a normalization of each color channel:

$$C^{*}(\mathbf{x}) = \frac{C(\mathbf{x})}{C(\mathbf{x})} \quad (7)$$

where the bar indicates a spatial average: $\bar{a} = \int_S a dx / \int_S dx$, and S is the surface of the patch. The Grey-World hypothesis [17], assuming average reflectance in the world to be grey, leads to a similar correction for the illuminant.

Let us now consider the case where there is, next to a dominant illuminant, diffuse light present. For this case Eq. 6 does not hold, instead the images are related via $\mathbf{C}_2 = \mathbf{\Lambda} \mathbf{C}_1 + \mathbf{a}$. However, a relation similar to Eq. 7 exists between the derivatives of these images, since from Eq. 5 it follows that

$$C_x^2(\mathbf{x}) = \lambda^C C_x^1(\mathbf{x}). \tag{8}$$

Invariance with respect to the dominant illuminant, $\mathbf{\Lambda}$, can hence be obtained by normalization of the color channels with their average derivative

$$C^*(\mathbf{x}) = \frac{C(\mathbf{x})}{|C_x(\mathbf{x})|}. \tag{9}$$

The absolute is taken to avoid division by zero. Since $C(\mathbf{x})$ is dependent on the diffuse illuminant, the resulting image, C^* , is also. However, its derivatives, C_x^* , are no longer dependent on either the dominant illuminant or the diffuse illuminant. The recently proposed Grey-Edge hypothesis [18], assuming average reflectance of differences in the world to be grey, leads to a similar correction for the illuminant.

3.2 Photometric Robustness: Color Invariance

This section will discuss the photometric invariants on which the color descriptors will be based. With color invariance we refer here to scene incidental variations such as shadows, shading and specularities. A brief overview of photometric invariants known from literature is given here (for more details see e.g. [9], [20]).

Zero-order invariants. Let us first consider the case of a matte surface ($m^i = 0$) and no diffuse lighting ($a^C = 0$). For this case normalized *rgb* can be considered invariant with respect to lighting geometry and viewpoint, m^b . Since,

$$r = \frac{R}{R + G + B} = \frac{m^b b^R e^R}{m^b (b^R e^R + b^G e^G + b^B e^B)}. \tag{10}$$

Similar equations hold for normalized *g* and *b*.

Furthermore, in the case of a white illuminant ($e^R = e^G = e^B = e$) and specular reflectance ($m^i \neq 0$), opponent colors [9] can be proven to be invariant with respect to specularities, m^i . Since,

$$\begin{aligned} O1 &= \frac{1}{\sqrt{2}} (R - G) = \frac{1}{\sqrt{2}} (m^b e (b^R - b^G) + m^i e - m^i e) \\ O2 &= \frac{1}{\sqrt{6}} (R + G - 2B) = \frac{1}{\sqrt{6}} (m^b e (b^R + b^G - 2b^B) + 2m^i e - 2m^i e), \end{aligned} \tag{11}$$

are invariant for m^i . The opponent colors are still variant for lighting geometry variations. Invariance with respect to both the lighting geometry and specularities is obtained by hue,

$$hue = \arctan\left(\frac{O1}{O2}\right) = \arctan\left(\frac{\sqrt{3}(b^R - b^G)}{(b^R + b^G - 2b^B)}\right) \quad (12)$$

First-order invariants. We continue by describing two photometric invariant derivatives [20]. Again consider a matte surface ($m^i = 0$). For this case, changes caused by lighting geometry variation ($m_x^b \neq 0$) are equal to

$$C_{\mathbf{x}} = \frac{m_x^b}{m^b} C, \quad (13)$$

meaning that all lighting geometry changes of $C_{\mathbf{x}}$ occur in the direction of the object color, C . Changes in the two direction perpendicular to object color are hence invariant with respect to geometry variations. These directions are equal to the angular derivatives after a spherical coordinate transformation,

$$ang1_{\mathbf{x}} = \frac{G_{\mathbf{x}}R - RG_{\mathbf{x}}}{\sqrt{R^2 + G^2}}, \quad ang2_{\mathbf{x}} = \frac{R_{\mathbf{x}}RB + G_{\mathbf{x}}GB - B_{\mathbf{x}}R^2 - B_{\mathbf{x}}G^2}{\sqrt{(R^2 + G^2)(R^2 + G^2 + B^2)}}. \quad (14)$$

If we subsequently consider specular reflection, the derivatives of the opponent colors,

$$O1_{\mathbf{x}} = \frac{1}{\sqrt{2}}(R_{\mathbf{x}} - G_{\mathbf{x}}), \quad O2_{\mathbf{x}} = \frac{1}{\sqrt{6}}(R_{\mathbf{x}} + G_{\mathbf{x}} - 2B_{\mathbf{x}}) \quad (15)$$

can be proven to be invariant with respect to specular variations, similarly as in Eq. 11. If the opponent derivative is computed after applying the illuminant normalization of Eq. 9 the opponent derivative is the only invariant insensitive to a diffuse illuminant.

A combined illuminant and geometric invariant. In [8] a method, called comprehensive color image normalization (CCIN), is proposed as a global image feature. We will here apply it as a local image feature. The method proposes an iterative use of Eq. 10 and Eq. 7, and hence is invariant for both lighting geometry and illuminant color.

3.3 Geometric Robustness: Color Angles

The third criterion requires geometrical robustness with respect to changes caused by viewpoint, zoom, and object orientation. Invariance with respect to these transformation is already partially obtained by affine invariant feature detection, however special care should be taken when working with derivative based invariants. This problem is usually overlooked in derivative-based invariance literature [7], [10], [20]. We will clarify the problem by investigating the influence of edge-sharpness for the opponent derivative. In Fig. 2a an edge is depicted. White is increasingly added to the blue patch along the y-axis (mimicking a transformation similar to specularities). The opponent derivative, $O1_x$,

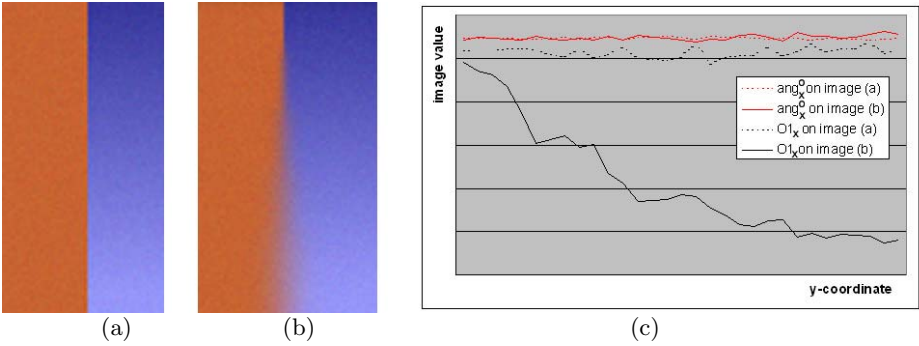


Fig. 2. (a) a red-blue edge. For the blue patch white light is added along the y-axis. (b) same as (a) but with varying smoothing along the y-axis. (c) the filter responses on the center vertical line of the images (a) and (b) of the opponent angle and the opponent derivative along the x-axis.

is invariant with respect to this phenomenon as can be seen by the constant response of the dashed black line in Fig. 2c. If we look at the opponent derivative response on the same image, but now with decreasing spatial smoothing along the y-axis, the response changes drastically. This behavior is undesired, since edge-sharpness changes occur a lot due to geometrical variation, or acquisition parameters such as zoom and focus. To overcome this problem we propose two new invariants, called color angles.

Assume that an edge can locally be modelled as a smoothed step edge

$$C(x) = \alpha^C \mathbf{u}(x) \otimes G^\sigma(x), \tag{16}$$

where α^C indicates the amplitude of the step edge u for the different channels C . Its derivative is equal to

$$C_x(x) = \alpha^C \frac{\partial}{\partial x} (\mathbf{u}(x) \otimes G^\sigma(x)) = \alpha^C \left(\left(\frac{\partial}{\partial x} \mathbf{u}(x) \right) \otimes G^\sigma(x) \right) = \alpha^C G^\sigma(x) \tag{17}$$

where we used that the derivative of a step edge is equal to the delta function $\left(\frac{\partial}{\partial x} \mathbf{u}(x) \right) = \delta(x)$. It is now straightforward to prove that the angles between the color channels are invariant to this smoothing and are only dependent on α^C , since

$$\phi = \arctan \left(\frac{R_x}{G_x} \right) = \arctan \left(\frac{\alpha^R}{\alpha^G} \right). \tag{18}$$

We can now add the geometrical invariance to the photometrical invariant derivatives derived in section 3.2. This leads to the opponent angle, $ang_{\mathbf{x}}^O$, and the spherical angle, $ang_{\mathbf{x}}^S$, with

$$ang_{\mathbf{x}}^O = \arctan \left(\frac{O1_{\mathbf{x}}}{O2_{\mathbf{x}}} \right) , \quad ang_{\mathbf{x}}^S = \arctan \left(\frac{ang1_{\mathbf{x}}}{ang2_{\mathbf{x}}} \right). \tag{19}$$

Table 1. Overview of the physical events to which the photometric invariants are insensitive. Prior illuminant normalization is assumed by either Eq. 7 or Eq. 9. Invariance with respect to the diffuse lighting for the opponent angle is obtained with Eq. 9.

	lighting geometry	specularities	illuminant variations	diffuse lighting
<i>rgb</i>	×	-	×	-
<i>hue</i>	×	×	×	-
<i>spher.ang.</i>	×	-	×	-
<i>opp.ang.</i>	-	×	×	×
<i>CCIN</i>	×	-	×	-

In Fig. 2c the results for $ang_{\mathbf{x}}^O$ are given in red. The smoothing does not influence the response, thereby demonstrating the robustness with respect to geometrical changes of the opponent angle. Note that in [21] color angles are also mentioned. They refer however to angles of color distributions, while in this paper we study the distribution of color angles of color derivatives. In Table 1 an overview of the invariants is given. The results assume that the invariants are applied in combination with the illuminant normalization method provided in section 3.1.

3.4 Photometric Stability: Robust Local Histograms

We are now in the final stage of the construction of the descriptor. In this section we describe how the derived invariants are transformed into a robust local histogram. Photometric invariants are known to have inherent instabilities [12], which we do not want to significantly influence the final histogram. Here, we propose to adjust the weight of a color value in the histogram according to its certainty. We discuss the computation of the weight for every invariant discussed in section 3.2. For *CCIN*-method we apply the parameters as indicated in [8].

rgb-histogram. We partition the *rgb* plane in triangles of equal size. To cope with the instability for low intensities we follow [22] and consider points below a threshold intensity as being grey.

hue-histogram. The *hue* is known to be unstable around the grey axis. We follow [12] and apply an error analysis to the hue:

$$(\partial hue)^2 = \left(\frac{\partial hue}{\partial O1} \partial O1 \right)^2 + \left(\frac{\partial hue}{\partial O2} \partial O2 \right)^2 = \frac{1}{O1^2 + O2^2} = \frac{1}{sat^2}, \quad (20)$$

where *sat* is the saturation (a similar results was derived for derivative of the hue in [20]). The certainty of the *hue* is hence inversely proportional to the saturation, which is what we expect. The smaller the saturation the more uncertain the *hue* estimation. We will use this relation to robustify the histogram construction, by weighting each sample by its saturation.

Opponent and spherical angle. Similarly as for the hue we apply an error analysis to the color angle equations of Eq. 19, which yield the following results

$$\partial ang_{\mathbf{x}}^O = \frac{1}{\sqrt{O1_x^2 + O2_x^2}} \quad , \quad \partial ang_{\mathbf{x}}^S = \frac{1}{\sqrt{ang1_x^2 + ang2_x^2}}. \quad (21)$$

Hence, we will use $\partial ang_{\mathbf{x}}^O$ as a weight for the opponent angle, and $\partial ang_{\mathbf{x}}^S$ as a weight for the spherical angle when converting them to a local color histogram. It is interesting to note that Lowe [3] intuitively arrives at the same conclusion. The orientation histogram of the SIFT descriptor is weighted with the gradient strength, which is exactly the result which would follow from an error analysis of the orientation parameter, $\theta = \arctan(f_y/f_x)$.

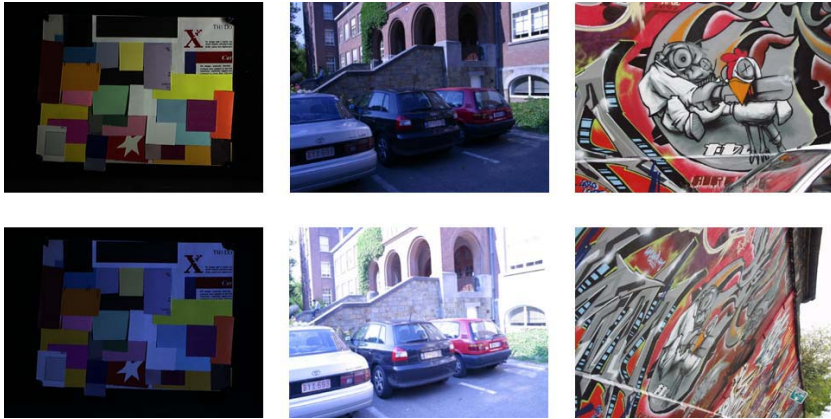
4 Experiments

The experiments test the color and the combined color and shape descriptor on the criteria put forward in section 1: 1. photometric robustness, 2. geometric robustness 3. photometric stability 4. generality. Although the first three criteria are tested by all experiments, the emphasis shifts: experiment 1 focusses on photometric robustness, experiment 2 demands geometrical robustness, and experiment 3 requires geometrical and photometrical robustness, and photometric stability to cope with the low quality internet images. The fourth criteria, generality, is illustrated by testing the descriptors for multiple tasks.

Experimental setup. For all the experiments we use the schema as given in Fig. 1. We use an affine invariant Harris-Laplace detector [2]. In the shape normalization step the images are reduced to 20 by 20 neighborhoods. The SIFT is computed from this shape normalized patch. For the color descriptors first color normalization is applied. Next the color descriptors, being the histograms of *rgb* (Eq. 10), *hue* (Eq. 12), *opponent angle* (Eq. 19), and *spherical angle* (Eq. 19), *CCIN* [8], are computed, with the weights as proposed in section 3.4. Furthermore, $\lambda = .6$ (see Eq.1) was experimentally found to give good results, and the descriptor lengths are 128 bins for SIFT, for the one-dimensional descriptors *hue*, *opponent angle* and *spherical angle* the histogram is divided in 37 bins, for the two-dimensional descriptor *rgb* 121 bins are used.

4.1 Matching: Free Illumination - Controlled Geometry

To test the color descriptors with respect to photometric variations, the descriptors are compared on a matching task on three sequences (see Fig. 3). The first sequence tests the robustness with respect to color illuminant changes. It consists of a Mondrian composition captured under 11 different illuminants (the images are from the Simon Frasier data set [23]). The second sequence [2], of six images, is an outdoor scene, taken with varying exposure times. This not only provokes an intensity change but also changes the amount of diffuse light captured by the camera (as modelled in Eq. 3). The third sequence [2] contains six images and tests the descriptors with respect to geometry variations.



matching	COLOR												
	rgb			hue			opp. ang.			sph.ang.			CCIN
color const.	-	0	1	-	0	1	-	0	1	-	0	1	×
Mondrian	6,9	46,7	45,1	3,5	58,0	45,0	12,8	55,4	56,3	29,0	<u>62,0</u>	56,4	40,0
cars	2,4	2,2	2,3	15,0	13,1	6,0	9,1	12,6	<u>20,2</u>	11,0	12,5	11,8	7,7

matching	shape	color					shape & color				
	SIFT	rgb	hue	opp.ang.	sph.ang.	CCIN	rgb	hue	opp.ang.	sph.ang.	CCIN
Mondrian	100	46,7	58,0	56,3	62,0	40,0	97,5	100,3	101,0	<u>101,8</u>	87,2
cars	100	2,2	13,1	20,2	12,5	7,7	91,6	80,5	<u>103,6</u>	94,3	87,5
graffiti	100	28,8	48,1	53,9	53,6	28,8	107,2	119,5	118,2	<u>121,4</u>	103,6

Fig. 3. Top: example images, Mondrian, cars, and graffiti. Middle: relative matching scores for various color normalization methods:- = no normalization, 0 = zero-order normalization, and 1 = first-order normalization. Bottom: relative matching scores for shape, color, and shape & color.

For all sequences we compute the matching score, defined in [2] as: the ratio between the number of correct matches and the smaller number of detected regions in the pair of images. A match is the nearest neighbor in the descriptor space (using Euclidean distance). For both color and color & shape we give the matching score relative to the matching score obtained by a unique shape descriptor. Values smaller than 100 indicate a performance worse than the shape descriptor and above 100 a performance better than the shape descriptor.

We start by selecting for each of the invariants the most appropriate color illuminant normalization method. This is done by comparing the matching scores for the two sequences with illuminance changes. In Fig. 3 the matching scores of descriptors for: 1. no illuminant. 2. zero-order illuminant normalization (Eq. 7) and 3. first-order (Eq. 9) illuminant normalization are given. The necessity of color illuminant normalization becomes clear from the results on the Mondrian sequence, where a significant gain in performance is achieved. The results on the car sequence show the importance of invariance with respect to diffuse light. Due to varying exposure times the amount of diffuse light entering the camera varies.

Hence for this sequence, the opponent derivative based descriptor outperforms the others significantly, since it is the only one which is invariant with respect to diffuse light changes. *Based on these results we will, in the remainder of the experiments, apply zero-order illuminance invariance to the descriptors rgb, hue, and spherical angle, and first order illuminance invariance to the opponent angle. The CCIN explicitly uses a zero order illuminance normalization.*

In Fig. 3 bottom the matching scores are summarized. Only for the graffiti sequence a substantial gain is obtained by adding color to the descriptor. Furthermore, for the car sequence, the descriptors which are not robust to diffuse lighting fail, and the combined performance of shape and color drops below a solely shape approach. For these sequences, where the assumption of an affine transformation between the images is not broken, the shape description performs outstanding, and relatively small gains are obtained by adding color.



retrieval	shape	color					shape & color				
	SIFT	rgb	hue	opp. ang.	sph.ang.	CCIN	rgb	hue	opp. ang.	sph.ang.	CCIN
objects	21	24	<u>36</u>	20	29	27	27	34	30	33	30

Fig. 4. Two instantiations of five example objects from the data set, with recall scores for shape, color and color & shape

4.2 Retrieval: Controlled Illumination - Free Geometry

Robustness with respect to geometrical variations in a relatively stable photometric environment is tested with a retrieval task on a data set containing 55 objects¹, see Fig. 4. Each object is captured 5 times under varying viewpoint, and object orientations. Because retrieval based on the entire image is close to perfect on this database, we assess the usefulness to retrieval of single features. Each single features was asked to retrieve the four most probable images (there are four relevant images for each query). To measure the descriptor performance we compute the recall, defined as the number of relevant images retrieved to the total number of relevant images in the database.

In Figure 4 the average recall over all local features is given, e.g. for shape a single feature alone obtains a recall of 21 percent. On this data set a unique color description outperforms the shape description. This has two causes. Firstly, the

¹ The authors lost the origin of this data set and would appreciate any suggestions.

objects have relatively saturated colors and photometric conditions are rather stable. Secondly, the shape variations are not constrained to affine transformation as in section 4.1, thereby complicating the shape matching task. Furthermore, the domination of specular reflectances on both objects and background in combination with the spatial variation of the lighting geometry, results in the best performance for the *hue* color descriptor, which is the only descriptor robust to both lighting geometry and specularities.

4.3 Classification: Free Illumination - Free Geometry

This experiment tests the descriptors on an image classification task. Based on the descriptors in the image, a decision is made whether the image is a member of the class or not. The multi-class classification is performed on two data sets. A bird data set [4] containing 6 classes of bird species, with 100 instantiations each (see Fig. 5). The classes are divided in 50 training and 50 testing images. For the second data set we collected images from 7 soccer teams, containing 40 images per class, divided into 25 training and 15 testing images per class². Although, players of other teams were allowed to appear in the images, no players being a member of the other classes in the database were allowed. Both databases consist of low-quality internet images. We use a bag-of-keypoints scheme [24]. The descriptors are clustered by a K-means algorithm which forms a set of visual words. Subsequently, each image is represented by a frequency histogram of the visual words. Based on these histograms, one-against-all classifiers are trained with a linear SVM. A test image is subsequently classified with all classifiers, and is appointed to the class for which it obtained the highest score.

In Fig. 5 the multi-class classification results for the birds data set are given. In all cases the combination of color and shape performs best. Only small gains are obtained with the zero-order invariants, *rgb*, *hue*, and *CCIN*. The derivative-based methods, the *opponent* and *spherical* angle, give considerably better results. For the soccer team data set the necessity of color information is especially apparent. Although, two of the teams have no color, and three of the teams have red as their main color, the color description performs considerably better than the solely shape-based description. The combination of color and shape further improves the results, and the 43% correct classification of shape is increased to a 73% for the combined descriptor. For these highly saturated colors, the results for the different color descriptors do not differ greatly.

The difference in performance for both the birds and the soccer teams can be explained by the different properties of the data sets. The soccer players are not only colorful, they have also undergone considerable non-affine shape transformations, reducing the performance of the SIFT descriptor. On the other hand, the non-saturated colors of the bird data set complicate the task to collect reliable color information, whereas the birds shape is relatively stable compared to the large variability encountered in the soccer team data set.

² The data set is available on <http://lear.inrialpes.fr/people/vandeweijer/data>



Fig. 5. Examples and multi-class classification results for the birds and soccer database

5 Conclusions

In this paper, we have taken a principled approach to extend the SIFT shape descriptor with a color descriptor. Based on four criteria, namely photometric robustness, geometric robustness, photometric stability and generality, we derive a set of photometric invariant color histograms, which are used as a color descriptor. We propose a solution to dependance of derivative-based invariants to the edge-sharpness. The descriptors are tested on a matching, a retrieval, and a classification task. For the colorful objects a pure color-based approach outperforms a shape-based approach. And for all reported data the combination of shape and color outperforms a pure shape-based approach, with gains going up as much as 70 percent. Depending on the data set different color descriptors obtain the best results. However, in general we would advice to use the robust *hue* descriptor for scenes with saturated colors, such as the object data set and the soccer teams. For scenes with less saturated colors, such as the bird data set, and especially in the presence of diffuse lighting, as for the cars sequence, we would advice the color descriptor based on the *opponent angle*.

Acknowledgments

This work is supported by the Marie Curie Intra-European Fellowship Program of the Commission of the European Union.

References

1. Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **19** (1997) 530–534
2. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., van Gool, L.: A comparison of affine region detectors. *International Journal of Computer Vision* **65** (2005) 43–72
3. Lowe, D.: Distinctive image features from scale-invariant keypoints. *International Journal Computer Vision* **60** (2004) 91–110
4. Lazebnik, S., Schmid, C.: A maximum entropy framework for part-based texture and object recognition. In: *Proc. IEEE Int'l Conf. Comp. Vision.* (2005) 832–838
5. Fergus, R., Perona, P., Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Volume 2. (2003) 264–271
6. Ballard, D.H.: Generalizing the Hough transform to detect arbitrary shapes. *Pattern Recognition* **12** (1981) 111–122
7. Funt, B., Finlayson, G.: Color constant color indexing. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **17** (1995) 522–529
8. Finlayson, G.D., Schiele, B., Crowley, J.L.: Comprehensive colour image normalization. In: *ECCV '98: Proceedings of the 5th European Conference on Computer Vision-Volume I*, Springer-Verlag (1998) 475–490
9. Gevers, T., Smeulders, A.: Color based object recognition. *Pattern Recognition* **32** (1999) 453–464
10. Geusebroek, J., van den Boomgaard, R., Smeulders, A., Geerts, H.: Color invariance. *IEEE Trans. Pattern Analysis Machine Intell.* **23** (2001) 1338–1350
11. van de Weijer, J., Gevers, T., Geusebroek, J.: Edge and corner detection by photometric quasi-invariants. *IEEE Trans. Pattern Analysis and Machine Intelligence* **27** (2005) 625–630
12. Gevers, T., Stokman, H.: Robust histogram construction from color invariants for object recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence* **26** (2004) 113–118
13. Mindru, F., Tuytelaars, T., Gool, L.V., Moons, T.: Moment invariants for recognition under changing viewpoint and illumination. *Computer Vision Image Understanding* **94** (2004) 3–27
14. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. *IEEE Trans. on pattern analysis and machine intelligence* **27** (2005) 1615–1630
15. Matas, J., Koubaroulis, D., Kittler, J.: Colour image retrieval and object recognition using the multimodal neighbourhood signature. In: *Proc. of the European Conference on Computer Vision, London, UK, Springer-Verlag* (2000) 48–64
16. Shafer, S.: Using color to separate reflection components. *COLOR research and application* **10** (1985) 210–218
17. Buchsbaum, G.: A spatial processor model for object colour perception. *Journal of the Franklin Institute* **310** (1980)
18. van de Weijer, J., Gevers, T.: Color constancy based on the grey-edge hypothesis. In: *ICIP'05, Genoa, Italy* (2005)
19. Finlayson, G., Drew, M., Funt, B.: Color constancy: Generalized diagonal transforms suffice. *Journal of the Optical Society of America A.* **11** (1994) 3011–3022
20. van de Weijer, J., Gevers, T., Smeulders, A.: Robust photometric invariant features from the color tensor. *IEEE Trans. Image Processing* **15** (2006) 118–127

21. Finlayson, G.D., Chatterjee, S.S., Funt, B.V.: Color angular indexing. In: ECCV 96: Proceedings of the 4th European Conference on Computer Vision-Volume II, London, UK, Springer-Verlag (1996) 16–27
22. Ohta, Y., Kanade, T., Sakai, T.: Color information for region segmentation. *Computer Graphics and Image Processing* **13** (1980) 222–241
23. Barnard, K., Martin, L., Funt, B., Coath, A.: A data set for colour research. *Color Research and Application* **27** (2002) 147–151
24. Willamowski, J., Arregui, D., Csurka, G., Dance, C.R., Fan, L.: Categorizing nine visual classes using local appearance descriptors. In: IWLAVS. (2004)