

Review Article

Performance Evaluation of Color Descriptors under Illumination Variation

Danyu Sun ^{1,2}, Xixuan Zhao ^{1,2} and Jiangming Kan ^{1,2}

¹School of Technology, Beijing Forestry University, 100083, China

²Key Laboratory of State Forestry Administration on Forestry Equipment and Automation, 100083, China

Correspondence should be addressed to Xixuan Zhao; zhaoxixuan@bjfu.edu.cn and Jiangming Kan; kanjm@bjfu.edu.cn

Received 30 December 2020; Revised 16 August 2021; Accepted 20 August 2021; Published 6 September 2021

Academic Editor: Domenico Caputo

Copyright © 2021 Danyu Sun et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Color descriptors, which involve the extraction of color information that is robust to illumination variation, are indispensable for accessing reliable visual information as illumination variation is inevitable in many practical cases. There has been many color descriptors proposed in literature, but the performance of different color descriptors in different scenes under illumination variation and the influence of the surface characteristics have not been investigated. In this paper, we first systematically introduced the theoretical basis of color descriptors, categorized the existing color descriptors according to the theoretical basis, and then compared the performance of different color descriptors utilized for image recognition and image retrieval tasks on both the indoor and outdoor image datasets. We adopted the recognition rate and normalized average rank as the evaluation criteria to measure the performance of color descriptors. Experiment results show that the color moment invariants (CMI) provide the optimal balance between the performance and dimensions in most tests, and color descriptors derived from physical reflectance models are more suitable for object recognition and image retrieval. We also concluded the best color descriptors for each kind of scene and surface characteristics.

1. Introduction

Color information is an important and efficient cue obtained by optical image sensors. However, in computer vision field, as illumination variation occurs in most of the real-world scenes, the color-based information will change and thus may easily result in errors in the follow-up applications. Consequently, to obtain a satisfactory performance in digital image processing, achieving color constancy in computer vision is of great importance. Color constancy is usually considered the effect whereby the perceived or apparent color of a surface remains constant despite changes in the intensity and spectral composition of the illumination [1]. To eliminate the influence of illuminations, illumination estimation [1–3] and illumination-independent color descriptors are two major approaches of achieving color constancy.

Color descriptors, which can represent the color features of an image, have been proved to be successful in applications such as image registration [4, 5], object recognition [6–8], face recognition [9–13], human detection [14], image

retrieval [15–18], video retrieval [19, 20], and image classification [21, 22]. Hor [23] proposed an image retrieval method based on the combination of local texture information of two different texture descriptors robust to rotation, which is constructed to improve computational efficiency. Khwildi and Ouled Zaid [24] proposed a method to improve the accuracy of high dynamic range image retrieval, whose feature extraction is based on a combination of the HSV histogram and color moment. Jacob et al. [25] proposed an interchannel color texture mode based on deep learning algorithm, which gives the interchannel color texture information of an image. It considers the unique channel information and its relationship with the adjacent pixel information in the opponent space. The experimental analysis using this descriptor is much better than the previous work in the context of image retrieval and face recognition. Kumar et al. [26] focused on the 3D sign language recognition in human action recognition and proposed a new color-coded topographical descriptor, which combines joint distances and angles computed from joint locations. Based on the improved Otsu threshold

algorithm, Wei et al. [27] proposed a new automatic fruit extraction method for a fruit picking robot, which uses a new feature in OHTA color space. In order to improve the picking accuracy of fruit-picking robot in a three-dimensional space, Wu et al. [28] proposed a new improved descriptor combining color features and three-dimensional geometric features.

Color descriptors are various and can be applied in different scenes, but their robustness to illumination variances is different. Some color descriptors are derived by mapping images to a different color space or via a physical reflectance model. In addition, the moment functions such as Hu moment [29] and Zernike moment [30] are also used as a theoretical basis for obtaining color descriptors. A recent work has concentrated on divisive information-theoretic feature clustering (DITC) [31] to deduce color descriptors. In addition, a color-texture descriptor against impulse noise [32] has been proposed, the descriptor is robust to rotating and impulsive noise, and its calculation is simple. Accordingly, it is necessary to analyze the illumination's robustness of existing color descriptors effectively.

In this paper, we evaluated the performance of different color descriptors for image recognition and image retrieval on both the indoor and outdoor image scenes under different illumination conditions. We compared a series of color descriptors, which are influential in the field of pattern recognition and compared them using the same evaluation scenario and the same image data. The evaluation criteria are the recognition rate of object recognition and the normalized average rank of image retrieval.

1.1. Related Work. There are some studies that conducted comprehensive evaluation of the strengths and the weaknesses of different descriptor approaches which help in identifying future research directions [33] in computer vision. Mikolajczyk and Schmid [34] compared descriptors computed on interest regions extracted with scale and affine-invariant detectors for image matching and object recognition; the performance is comprehensively compared for affine transformations, scale changes, rotation, blur, jpeg compression, and illumination changes. It has been proved that descriptors SIFT-style [35] descriptor outperforms other methods in their paper. In the context of image classification, van de Sande et al. [36] evaluated invariance properties and distinctiveness of several color descriptors. The performance is measured by mean average precision using the SVM algorithm [37] and the position in rankings for image benchmark. Burghouts and Geusebroek [38] compared the discriminative power and invariance of gray value invariants to that of local color invariants in the context of image matching, where the discriminative power is measured by determining the recall of the regions that are to be matched and the precision of the matches. Setkov et al. [39] discussed the invariance properties of some existing color descriptors in that application for nonintrusive geometric compensation. The performance is mainly measured by two indicators: detection rate, which is the ratio between the number of detected matches, and the number of feature points in the image and FM precision indicating how match-

ing quality deteriorates in matching images. Bianconi [40] proposed a general and extensible framework to classify color texture, and the existing methods for color texture representation are compared both theoretically and experimentally. They pointed out that a separate color and texture processing can achieve a balance between performance and limited dimensions. In order to investigate the effects of illumination variations on color texture features, Cusana et al. [41] evaluated and compared several color texture descriptors using a new texture database RaWFOOT, they pointed out that traditional texture descriptors have limited robustness to illumination variations, and CNN-based descriptors [42] outperform the handcrafted traditional and object-oriented features in the context of texture classification. Ma et al. [43] focused on the classic methods in the field of image matching in the past two decades, as well as the methods based on deep learning in recent years, which covers feature detection, feature description, feature matching, image registration, stereo matching, point set or point cloud registration, and other related subfields. For the traditional feature description algorithm, they pointed out that extracting more accurate and repetitive features, more significant and distinguishable feature descriptors are the future development trend. In order to remove the mismatches from image features, Ma's team has done a lot of researches; they proposed a locality preserving matching method, which preserves the potential real matching in local neighborhood [44]. They also try to transform feature matching into a spatial clustering problem, which achieves promising performance [45].

However, very little work focus on how color descriptors are affected by different kinds of surface characteristics with illumination variations. In addition, there is little introduction and classification of the theoretical basis for deriving color descriptors.

In this paper, we aim to analyze the robustness of various descriptors to illumination variations and the influence of object surface characteristics on image recognition and retrieval. The comparison for practicability of different color descriptors is carried out in the context of image recognition and retrieval. We believe that this manuscript may serve as a handbook for those who need to select an appropriate method for a specific scenario (indoor, outdoor scenes or different surface characteristics), as well as a basis for development of new color descriptors.

1.2. Overview. This paper is organized as follows: In Section 2, the taxonomy of color descriptors based on a theoretical basis and a brief introduction of selected color descriptors are presented. Section 3 describes the experimental setup as well as our evaluation criterion and image data. Finally, a discussion of results is given.

2. Color Descriptors

2.1. Theoretical Basis. A novel descriptor is possible to be derived from different theories. The common theoretical bases are the selection of color space or the establishment

of new color space, physical reflectance model, mathematical model, and information theory.

2.1.1. Color Space's Selection or Establishment. The color space is actually the process of image color digitalization, different color spaces are just different representations of the same physical quantity, there are transformations between them. Koschan and Abidi [46] divided the most classical color space into three families:

- (i) Color spaces based on physics and technology, which includes the color space based on the trichromatic theory (RGB, CMY (K), etc.), luminance-chrominance color space (YIQ, YUV, etc.), and independent axis color model. The RGB color space is the most common color space, which is mainly used in the TV and computer color display system. It uses red, green, and blue components to represent the image. The CMY (k) color space is often used in printing and publishing. The YUV and YIQ color spaces, which use luminance and chrominance components to represent color image, are commonly used to represent color image in the TV system
- (ii) Uniform color spaces (Lab, Luv, etc.) which are important applications for comparison of similar colors. Lab and Luv color spaces, using the digital method to describe the human visual perception, are color systems based on physiological characteristics
- (iii) Color spaces based on perception (HIS, HSV, opponent, etc.) quantify the human color perception using intensity, hue, and saturation. The HSV color space is created according to the intuitive characteristics of color, which divides the color signal into three attributes: hue, saturation, and brightness. The principle of the HIS color space is similar to HSV

Several color descriptors are derived from the selection or establishment of different color spaces. Li et al. [8] proposed the central color coordinate system and edge-based color coordinate system based on the diagonal-offset model. It is of great importance for selecting a suitable color space or proposing a new color space with color invariance to improve the robustness of color descriptors to illumination variation.

2.1.2. Physical Reflectance Model. Image information is modeled by means of the Kubelka-Munk theory [47] for colorant layers. The physical reflectance model resulting from the Kubelka-Munk theory is given by [48]. The reflected spectrum in the viewing direction is given by $E(\lambda, \vec{x})$:

$$E(\lambda, \vec{x}) = e(\lambda, \vec{x}) \left(1 - \rho_f(\vec{x})\right)^2 R_\infty(\lambda, \vec{x}) + e(\lambda, \vec{x}) \rho_f(\vec{x}), \quad (1)$$

where \vec{x} denotes the position at the imaging plane and λ the wavelength. Further, $e(\lambda, \vec{x})$ denotes the illumination spectrum and $\rho_f(\vec{x})$ the Fresnel reflectance at \vec{x} , and the material reflectivity is denoted by $R_\infty(\lambda, \vec{x})$.

The other common physical illumination model is the Lambertian model:

$$f(\vec{x}) = \int_{\omega} e(\lambda) S(\vec{x}, \lambda) c(\lambda) d\lambda, \quad (2)$$

where \vec{x} denotes the position at the imaging plane, λ the wavelength, and ω the range of visible light; $e(\lambda)$ represents the spectral distribution of the light source; $S(\vec{x}, \lambda)$ denotes the physical reflectivity of the surface of the object at the point \vec{x} the light of wavelength λ ; and photographic function of imaging device $c(\lambda) = (R(\lambda), G(\lambda), B(\lambda))^T$.

Light scattering refers to the phenomenon that a part of light deviates from the original direction when it passes through the inhomogeneous medium. The condition of natural light illumination is mostly the mixture of direct light and scattered light. Shafer [49] added scattered lighting into the Lambertian model to simulate natural light conditions, so that the model could be more universal:

$$f(\vec{x}) = \int_{\omega} e(\lambda) S(\vec{x}, \lambda) c(\lambda) d\lambda + \int_{\omega} \phi(\lambda) c(\lambda) d\lambda, \quad (3)$$

where $\phi(\lambda)$ indicates scattered lighting. This improvement is more suitable for outdoor scenes. Similarly, Shafer obtained a dichromatic reflection model [49] based on the Kubelka-Munk model (KM).

Many new color descriptors are derived from these physical reflection models by adding other influence indicators. The algorithm derived from the physical model has photometric invariance, but it sacrifices the discriminant power of the descriptors.

2.1.3. Mathematical Model. Moments are widely used in image processing, and the moment sets that are calculated from an image describe the global features of the shape of the image and provide much information about the different types of geometric features of the image. Hu [29] proposed seven moment invariants of the first order to third order, and a series of improved algorithms for Hu moments are applied to object recognition and image retrieval [50, 51]. Hu moments have been proved to be robust to translation, rotation, and scaling. According to different color channels and Hu moments, color moment invariants can be deduced, which has good robustness to illumination change. In addition, color moment invariants can be obtained by the Lie group approach [52]. Note that the color moment invariants' dimension is small so that they are suitable for image recognition and retrieval.

2.1.4. Information Theory. It is well known that a pixel value, which is captured in a color space that can be mapped to the same value in the photometric invariance color space, can sacrifice the discriminative power of the color descriptors

derived from the physical reflectance model. In order to find the balance of illumination invariance and discriminative power, a divisive information-theoretic feature clustering algorithm (DITC) proposed by Dhillon [31] is used to construct color descriptors with high discriminative power. The DITC algorithm introduces the concept of information entropy and Kullback-Leibler (KL) divergence information theory. The color descriptors based on the information theory are employed to find an optimum clustering approach in the color space with higher discriminative power.

2.2. Brief Introduction of Current Color Descriptors. In the following, color descriptors are presented and we categorize them according to the four theoretical foundations mentioned above.

2.2.1. Color Descriptors Based on the Selection of the Color Space

(1) Histogram.

(i) RGB histogram

The RGB histogram is a combination of three 1D histograms based on the channel of R, G, and B in the RGB color space. The spatial structure of the RGB color space is not consistent with people's subjective judgment of color similarity, and it possesses no invariance to illumination changes.

(ii) rg histogram

It is the normalization process of the RGB color space, and the chromaticity components r and g represent the color information in images:

$$\begin{pmatrix} r \\ g \end{pmatrix} = \begin{pmatrix} R/(R+G+B) \\ G/(R+G+B) \end{pmatrix}, \quad (4)$$

Due to the normalization, the components r and g have been proved to be invariant to scaling, light intensity changes, shadows, and shading.

(iii) Hue histogram

When mapping images to the HSV color space, it is proved that the hue becomes unstable near the gray axis. Weijer et al. [53] have discovered that the certainty of the hue is inversely proportional to saturation so that the hue histogram is made more robust by weighting each sample of the hue by its saturation. Only the color channel H [53] is scale-invariant and shift-invariant with respect to light intensity.

(i) Opponent histogram

The opponent histogram, in which images are mapped to the opponent color space, combines the three 1D channels of histograms:

$$(O_1 O_2 O_3)^T = \begin{pmatrix} T - G & R + G - 2B & R + G + B \\ \sqrt{2} & \sqrt{6} & \sqrt{3} \end{pmatrix}^T, \quad (5)$$

where the components O_1 and O_2 represent color information, and the component O_3 denotes intensity information. van de Sande et al. [36] have proved that the components O_1 and O_2 are shift-invariant with respect to light intensity. The intensity channel O_3 has no invariance properties.

(2) Color SIFT Descriptors.

(i) SIFT

Lowe [35] proposed the SIFT to describe local features using edge orientation histograms based on points of interest. SIFT features possess invariance properties to rotation, scaling, and light intensity variation. However, the SIFT is not invariant to light color changes due to the combination of R, G, and B channels in the intensity channel. Note that SIFT is extensible and can be easily combined with other forms of feature vectors.

(ii) RGB-SIFT

RGB-SIFT is the combination of the SIFT computed in the R, G, and B channels separately. It has been proved that RGB-SIFT is scale-invariant, shift-invariant, and invariant to light color changes.

(iii) HSV-SIFT

The method of calculating the SIFT over three channels of the HSV color space independently has been proposed by Bosch et al. [54]. The descriptor has 128-dimensional features for each channel. It is known that the H channel is scale-invariant and shift-invariant with respect to light intensity. However, the combination of HSV channels provided low invariance.

(iv) Hue-SIFT

van de Weijer et al. [53] introduced a concatenation of the hue histogram with the SIFT descriptor. Weighting the hue histogram with saturation solves the problem of instability near the gray axis for HSV-SIFT. The Hue-SIFT possesses invariance to intensity changes.

(v) Opponent-SIFT

Opponent-SIFT is constructed from the transformation between the RGB color space and opponent color space (5); similarly, the o_1 and o_2 channels describe color information while the o_3 channel represents intensity information in images. In [36], opponent-SIFT shows the best result in the context of image classification.

(vi) C-SIFT

C-SIFT [38, 55] is derived from the normalization of 2D opponent color space which offsets the intensity information in the o_1 and o_2 channels: O_1/O_3 and o_2/o_3 . Because of the lack of intensity information, C-SIFT possess lower discriminative power.

(vii) rg-SIFT

According to the normalization of the RGB color space (4), the rg-SIFT can be regarded as an independent computation of SIFT features in the r and g chromaticity components.

(3) Two Famous Color Descriptors.

(i) Robust hue descriptor (HUE)

Weijer and Schmid [55] proposed to use the histogram of the hue that is linearly transformed from the RGB color space to describe image patches:

$$\text{hue} = \arccan\left(\frac{\sqrt{3(R-G)}}{R+G-2B}\right). \quad (6)$$

Weighting the influence of the histogram by saturation of the corresponding pixel to make the hue more stable, the hue descriptor is invariant with respect to lighting geometry and specularities when assuming white illumination.

(ii) Opponent derivative descriptor (OPP)

Weijer and Schmid [55] proposed to represent the image patches by a histogram of opponent derivative:

$$\text{ang}_x^o = \arccan\left(\frac{O_{1x}}{O_{2x}}\right), \quad (7)$$

where O_{1x} and O_{2x} denote the spatial derivative in the color information channels. The descriptor weighted by $\sqrt{O_{1x} + O_{2x}}$ has been proved to be invariant with respect to specularities and diffuse lighting.

2.2.2. Color Descriptors Based on the Physical Reflectance Model

(1) *Color Descriptors Derived from Lambertian Model.* Funt and Finlayson [56] and Gever and Smeulders [57] proposed photometric invariant images' index on the basis of the Lambertian model. However, the color descriptors mentioned above is susceptible to blur. Weijer and Schmid [58] provided a construction scheme of color invariant descriptors to counter the impact of image blur. However, these descriptors all depend on the edge information of the image so that they will lose much of the color information for images with sparse edges.

(2) *Color Descriptors Derived from the Kubelka-Munk Model.* Geusebroek and Boomgaard [48] proposed a framework of five-color invariants (H, C, W, N , and E) to describe object reflectance regarding five different image conditions, which used the Gaussian scale space paradigm of color images. The five different imaging conditions are as follows:

- (a) equal energy and uneven illumination
- (b) equal energy but uneven illumination and matte, and dull surfaces

- (c) equal energy and uniform illumination and matte, dull surfaces and planar objects

- (d) colored but uneven illumination

- (e) arbitrary image conditions

The object reflectance is derived from the Kubelka-Munk theory, extensive experiments show these color measurements to obtain highly discriminative power while maintaining photometric invariance.

2.2.3. *Color Descriptors Based on a Mathematical Model.* A color image can be described by a function I derived from every channel of the RGB color model at the spatial position: $\{x, y\}: I(x, y) = (R(x, y), G(x, y), B(x, y))$. Mindru et al. [59] have provided a definition of the generalized color moment:

$$M_{pq}^{abc} = \int \int x^p y^q [I_R(x, y)]^a [I_G(x, y)]^b [I_B(x, y)]^c dx dy, \quad (8)$$

where M_{pq}^{abc} denotes a generalized color moment of order $p + q$ and degree $a + b + c$. In particular, it means a lack of color information for moments of degree 0, i.e., M_{pq}^{000} , while for moments of order 0, i.e., M_{00}^{abc} , this means that it does not contain any spatial information. The framework can be applied to construct various moments; however, to maintain the stability of color moments, only the generalized color moments up to the first order and second degree are adopted.

(i) Color moment

Generalized color moments up to the second degree and the first order are adopted to construct the color moment (CM) descriptor. Hence, the resulting color moments are a combination of three possible situations for order: $(p, q) \in \{(0, 0), \{1, 0\}, \{0, 1\}\}$ and ten possible situations for degree: $(a, b, c) \in \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (2, 0, 0), (0, 2, 0), (0, 0, 2), (1, 1, 0), (1, 0, 1), (0, 1, 1)\}$.

Consequently, the color moment descriptor has 30 dimensions and it is only robust to shift variation.

(ii) Color moment invariants

Mindru et al. [59] provided various constructions of color moment invariants (CMI) based on the framework of the generalized color moment. In this paper, due to the better results achieved by PSO invariants, we select it. Furthermore, it is deduced under the circumstance that no geometric deformations are present and the photometric transformations are of Type SO. This yields 24 dimensions.

2.2.4. Color Descriptors Based on Information Theory

(i) Color name

Berlin and Kay [60] have defined basic color terms based on a large number of anthropological studies, and they provided eleven basic terms to describe color: CN = {black, blue, brown, grey, green, orange, pink, purple, red, white, yellow}.

According to the theory, Benavente et al. [61] proposed the fuzzy color name (FCN) to model the category of arbitrary color terms by using the concept of fuzzy sets, where the color name descriptor describes the probability of a color stimulus belonging to arbitrary color terms. van de Weijer et al. [62] have proposed a novel color name (CN) descriptor to represent an image:

$$K = \{p(n_1|R), p(n_2|R), \dots, p(n_{11}|R)\},$$

$$\text{with } p(n_i|R) = \frac{1}{N} \sum_{x \in R} p(n_i|f(x)), \quad (9)$$

where n_i is related to i th color name, x denotes the spatial coordinates of N pixels in region R corresponding to the $L^*a^*b^*$ color space, i.e., $f = \{L^*, a^*, b^*\}$, and $p = \{n_i|R\}$ represents the probability of a color name under the corresponding pixel value. It is proved that the color name has good performance in image classification.

(ii) Discriminative color descriptors

The color descriptors derived from a physical-based model can introduce a decline of its discriminative power. Khan et al. [63] have proposed discriminative color descriptors (DD) based on an information theoretic approach. The DD bear some resemblance to the color name, which discretize the color space into eleven parts. The DD are employed to find optimized feature clusters for minimization of global objective function, which in this case, separate the original color space into m color words $W = \{\omega_1, \dots, \omega_m\}$, where $m = 10 \times 20 \times 20 = 4000$ in the $L^*a^*b^*$ color space.

Note that image data with l classes is given by $C = \{c_1, \dots, c_l\}$. Hence, the mutual information describing the discriminative power of color words W regarding the problem of differentiating the classes C is given:

$$I(C, W) = \sum_i \sum_t p(c_i, \omega_t) \log \frac{p(c_i, \omega_t)}{p(c_i)p(\omega_t)}, \quad (10)$$

where $p(c_i, \omega_t)$ and the prior $p(c_i)$ and $p(\omega_t)$ can be measured empirically from image data.

3. Experimental Setup

We first give implementation details for the evaluated color descriptors. We then describe the image datasets used for evaluation. Finally, the evaluation criteria are given.

3.1. Feature Mapping Pipelines. To evaluate the essential characteristics of color descriptors, the bag of features (BOF) model [64] is used to obtain fixed-length feature vectors. The main algorithm used in the BOF model is k -means clustering. In essence, it is a process of clustering the extracted image features to construct a visual codebook and then mapping features against the visual codebook. Hence, the BOF model implements vector quantization of the color descriptors. As described in [64], the process of establishing a visual codebook is as follows:

- (a) Feature extraction and description. We use a grid (16×16) with 50% overlap to extract patches from images, and different color descriptors are calculated on every patch
- (b) Construction of visual codebook. The visual codebook is constructed with the feature patches extracted before, and the k -means algorithm is used to cluster all patches, which are divided into k classes (in this paper $k = 300$).
- (c) Representation of image features. The frequency of each feature word in the k -dimensional visual codebook of test images is counted; then, each test image can be represented as a k -dimensional vector

In addition, when calculating the whole image directly for color descriptors based on histograms, we select 16 bins for each color space channel.

3.2. Performance Evaluation

3.2.1. Dataset. We use the popular color datasets aiming at studying the effects of illumination variance: SFU (Simon Fraser University), ALOI (Amsterdam Library of illumination), Phos, RawFoot (raw food texture database), and THRI2015 (time-lapse hyperspectral radiance images).

(i) Indoor dataset

The 321 SFU image data [65] is provided by Computational Vision Lab of Simon Fraser University, and it contains 330 images of 30 scenes under 11 different artificial lights. Some images were culled from each set due to deficiencies in the calibration data so that 321 images are adopted. It is proved to be germane to study the computational color constancy. The image data is split into four groups: images with minimal specularities (Mondrian, 22 scenes, 223 images); images with nonnegligible dielectric specularities (specular, 9 scenes, 98 images); images with metallic specularities (metallic, 14 scenes, 149 images); and images with fluorescent surfaces (fluorescent, 6 scenes, 59 images). The 321 SFU image data is used for image recognition. In addition, another image data, which is also from Simon Fraser University, consists of 220 images of 20 scenes under 11 different illumination conditions. The ALOI image dataset [66] contains 1200 objects, each one with 12 images obtained under different illumination conditions; we choose the top 30 objects, 12 images of each one, and a total of 360 images for experiments. The Phos image dataset [67] contains 220 images of 15 scenes under 15 different illumination conditions, i.e., various strengths of uniform illumination and different degrees of nonuniform illumination. The RawFoot image dataset [68] includes images of samples of textures, acquired under 46 lighting conditions which may differ in the light direction, in illuminant color, in its intensity, or in a combination of these factors. We choose 460 images of 10 objects under 46 different illumination conditions from RawFoot.

(ii) Outdoor dataset

The sequences of time-lapse hyperspectral radiance images of Natural Scenes 2015 (THRI2015) [41] contains 34 images of four natural scenes under 7-9 different natural light sources based on the passage of time. In other words, the images were acquired at approximately 1-hour intervals for each scene. For convenience of image retrieval, we reject redundant images from each scene to make our data include 24 images of four scenes under 6 time points of natural light.

3.2.2. Evaluation Criteria. In the experiment of image recognition, we use the common k -nearest neighbor classification scheme where the Euclidean distance is adopted as a measurement of the feature's distance between images. Firstly, k images with the closest feature distance are selected for the test images in the training set, if one type of image appears most frequently in the k images, then the test image belongs to this category. In this paper, k values are selected to be 1, 3, and 5 for the assessment of color descriptors. We take the recognition rate of the objects as the evaluation criteria. Suppose that a total of N experiments was conducted, of which M times were correctly identified. Then, the recognition rate (RR) can be defined as follows:

$$\text{RR} = \frac{M}{N} \times 100\%. \quad (11)$$

We also calculate the average RR over three different k values. For the experimental scheme, we choose the leave-one-out approach. Suppose that there are N images in the image data, and every time, only one is taken as the test image, and the remaining $N - 1$ images are used as training images. In this way, different images are selected as test images and then N experiments are conducted.

In the context of image retrieval, the performance of color descriptors is assessed by the normalized average rank for a single query [55]:

$$\text{NAR} = \frac{1}{NN_R} \left(\sum_{i=1}^{N_R} R_i - \frac{N_R(N_R + 1)}{2} \right), \quad (12)$$

where N denotes the total number of image data, N_R is the number of images relevant to the query image, and R_i represents rank of the i th relevant image in the query results. A smaller NAR indicates a better retrieval result. The retrieval result is perfect when $\text{NAR} = 0$, and $\text{NAR} = 0.5$ means that the retrieval is random. We take the average NAR value (ANAR) of all candidates as the final result on the image data. For example, we select the first image on every scene of the ALOI image data (360 images of 30 scenes under 12 different illumination conditions) as a candidate, and the remaining images per scene are the training set so that 10 relevant images are possessed by every candidate. Hence, $N = 330$, $N_R = 11$.

In addition, in order to measure the illumination invariance of features, we also measure the Euclidean distance between these color descriptors of the same scene including indoor and outdoor scenes under different lighting conditions. The indoor image data from the Amsterdam Library

of Object Images (ALOI) [67] contains 12 images of 1 scene under 12 different artificial lights as shown in Figure 1(a), and the outdoor image data extracted from THRI2015 contains 6 images of 1 scene under 6 different natural illuminations as shown in Figure 1(b).

4. Results

This section outlines the discussion of the results obtained in our experiments. The performance of the color descriptors discussed in Section 2 under illumination variation is compared in the context of image recognition and image retrieval. For the indoor image data, we divide images into four kinds of surface conditions to analyze the effect of different surface characteristics on the performance of descriptors.

4.1. Experiment on Image Recognition. For ease of presentation in graph, the classification of descriptors is abbreviated as follows: color descriptors based on the selection of color space (Colorspace), color descriptors based on physical reflectance models (Phy.), color descriptors based on mathematical models (Math.), and color descriptors based on information theory (Inf.).

(i) Indoor image data

Figure 2 shows the results of color descriptors in SFU 321 image data for images with different surface characteristics. From the results in Figure 2(a), the color descriptors derived from the physical reflectance model such as W , C , N , E , and H have a high degree of robustness to illumination color variation, among which W is the best, because the property of W can be interpreted as an edge detector and it does not only represent object properties but also includes the shadow edge information, at the same time, the construction process of W , which is based on the situation that equal energy, uniform illumination and matte, dull, surfaces, makes W more suitable for these images containing minimal specularities. For color space selection, a majority of SIFT and color SIFT descriptors perform much better than histogram-based descriptors. From the results in Figures 2(b) and 2(c), and N followed by W performs better than other color descriptors. The OPP followed by HUE also have good performance, note that OPP, which uses a histogram over the opponent angle to represent image patches, is invariant to specularities and diffuse lighting; therefore, it has satisfactory performance for images with metal reflection. The majority of the histogram-based color descriptors lack illumination invariance except rg histogram, which improves the robustness of the rg color space to light change by normalizing the RGB color space. The color descriptors derived from the physical reflectance model still have good performance with $\text{RR} > 90\%$. It is remarkable that H , which is related to the hue of materials for equal energy but uneven illumination, has a reduced performance with $\text{RR} = 61.52\%$. The RR of color descriptors based on information theory are more than 80%, of which CN performs slightly better than the 50-dimensional DD. Regarding the results in Figure 2(c) for

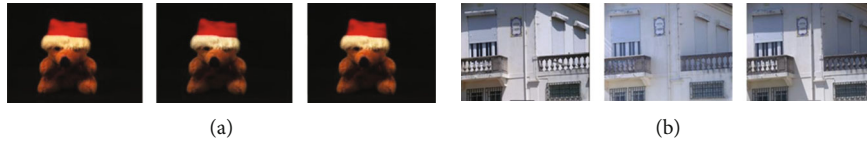


FIGURE 1: Examples of images for invariance measurement. (a) Indoor images from ALOI. (b) Outdoor images from THRI2015.

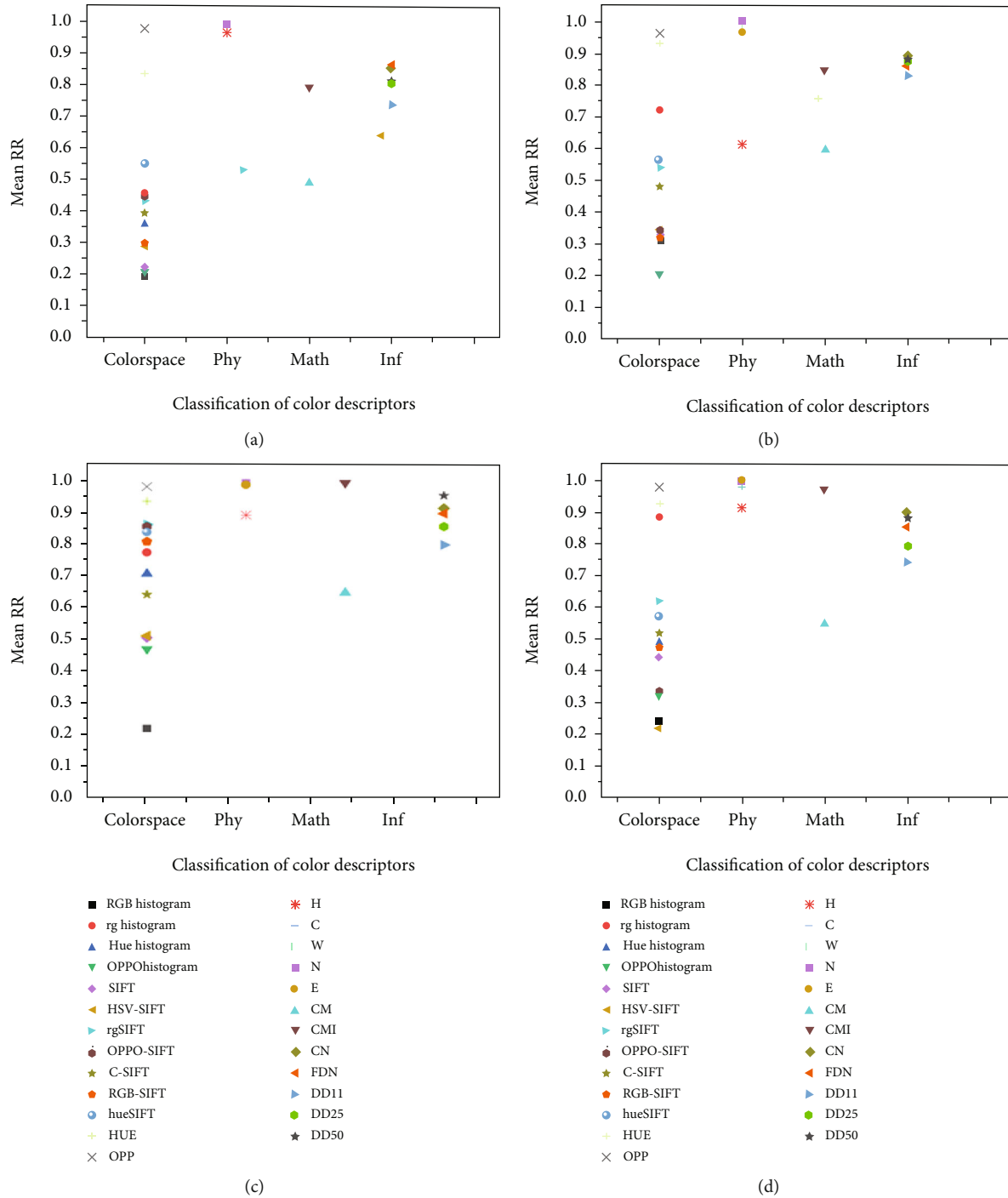


FIGURE 2: Mean RR of different color descriptors under different k values in the SFU 321 image data with different surface characteristics. (a) Images containing minimal specularities. (b) Images containing metallic specularities. (c) Images containing at least on fluorescent surface. (d) Images with nonnegligible dielectric specularities.

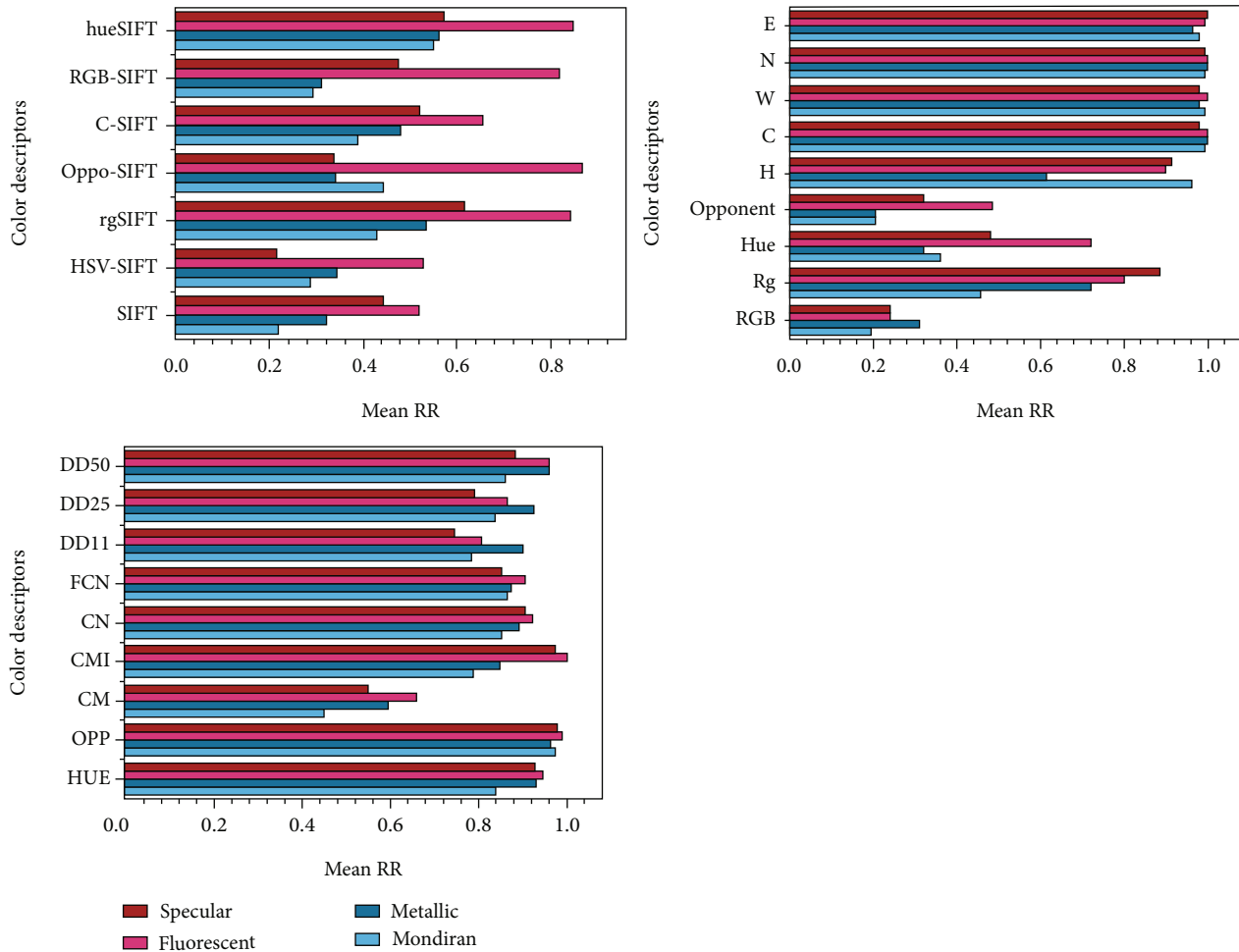


FIGURE 3: The influence of four different scenes' surface characteristics on the value of mean RR for different color descriptors.

images containing at least on fluorescent surface, most of the color descriptors have an RR of more than 80 percent except for histogram-based color descriptors, color moments, DD11, SIFT, HSV-SIFT, and C-SIFT. Observing the results in Figure 2(d) for images with nonnegligible dielectric specularities, *E* performs best slightly reduced by *N* and *C* whereas HSV-SIFT performs worst, which is due to HSV-SIFT lack of invariant properties. Note that for descriptors based on physical reflectance models, *E* has a more wide application than others. Since *N* indicates transition in object reflectance, *N* can keep stable for images with specular reflection. In addition, the RR of both CMI and color descriptors derived from KM are higher than ninety percent. The OPP has better performance than HUE, which is followed by CN. For all the surface characteristics, CMI is better than CM, which is in accordance with the expected theoretical results, and DD shows an obvious improvement in performance with increase of dimensions, among which DD with 50 dimensions is more satisfactory.

According to the comparison for the influence of descriptors in Figure 3, the performance of most color descriptors with respect to the color space selection shows an increase in images with metallic specularities (metallic)

or images with nonnegligible dielectric specularities (specular) compared with images with minimal specularities (Mondrian) except opponent-SIFT. For images with specularities, which may result in appearance of highlight area, the performance of opponent-SIFT, which has been proved that it is not robust to light color change and shift [28], is reduced. Different surface characteristics have the least influence on the color descriptors based on physical reflectance models. In view of information-theory-based color descriptors, they all have better performance in metallic than in Mondrian. Descriptors perform better in specular than in Mondrian except DD50 and FCN. It is not difficult to find that most color descriptors in the metallic category perform better than those in the specular category. The nonnegligible dielectric specularities have a more negative effect on color descriptors than metal-containing specular reflection. Meanwhile, performance of most descriptors shows an improvement for images with fluorescent surfaces, whose key characteristic is that some of the light energy they absorb is reemitted at longer wavelengths except DD25 and DD11.

Table 1 shows the performance of descriptors in indoor image data. For ALOI image data, we can observe that *C*, *W*, *N*, and CMI give perfect results, and *E*, followed by *H*,

TABLE 1: Performance comparison of color descriptors in indoor image data.

Descriptor	ALOI image data				Recognition rate (RR) (%)				RawFoot image data			
	$K = 1$	$K = 3$	$K = 5$	Mean RR (%)	$K = 1$	$K = 3$	$K = 5$	Mean RR (%)	$K = 1$	$K = 3$	$K = 5$	Mean RR (%)
RGB	98.61	97.22	90.56	95.46	66.67	61.33	65.33	64.44	81.52	71.74	67.61	73.62
rg	80.83	78.06	75.56	78.15	88	86.67	89.33	88	70.22	65.87	61.52	65.87
Hue	91.94	87.78	77.22	85.65	66.67	52	33.33	50.67	78.7	75.65	68.26	74.2
OPP	82.78	78.33	73.61	78.24	84	77.33	76	79.11	60.65	61.93	57.61	60.06
SIFT	93.33	89.72	89.44	90.83	32	30.67	38.67	33.78	68.26	61.09	58.48	62.61
HSV-SIFT	40	35.28	36.67	37.32	38.67	45.33	44	42.67	79.13	74.14	73.04	75.44
rg-SIFT	58.89	52.22	47.5	52.87	60	64	58.67	60.89	77.39	75.87	74.78	76.01
OPP-SIFT	36.11	38.89	39.44	38.15	46.67	44	49.33	46.67	75	73.04	74.78	74.27
C-SIFT	40	32.78	28.61	33.8	42.67	40	38.67	40.45	45.43	40.87	39.13	41.81
RGB-SIFT	35.56	36.39	35.56	35.84	52	46.67	41.33	46.67	77.61	77.39	73.48	76.16
Hue-SIFT	69.17	65.56	65.56	66.76	93.33	90.67	90.67	91.56	96.3	94.57	93.48	94.78
HUE	86.94	87.22	86.67	86.94	93.33	86.67	89.33	89.78	91.09	89.57	86.96	89.21
OPP	82.78	85.56	85.28	84.54	94.67	93.33	89.33	92.44	94.35	94.57	93.91	94.28
H	98.33	98.61	96.67	97.87	94.67	94.67	94.67	94.67	74.78	70	67.83	70.87
C	100	100	100	100	92	89.33	85.33	88.89	89.57	91.74	91.09	90.8
W	100	100	100	100	89.33	89.33	81.33	86.66	93.48	93.69	92.39	93.19
N	100	100	100	100	93.33	89.33	90.67	91.11	76.3	78.04	75.22	76.52
E	99.72	99.72	99.72	99.72	86.67	86.67	88	87.11	91.09	89.78	90	90.29
CM	92.5	92.22	90	91.57	18.67	9.3	4	10.66	63.91	55.87	54.13	57.97
CMI	100	100	100	100	60	58.67	57.33	58.67	75.87	70	65.87	70.58
CN	68.61	67.5	67.78	67.86	88	85.33	85.33	86.22	92.61	89.57	89.35	90.51
FCN	70	68.06	68.06	68.71	84	84	73.33	80.44	93.48	91.96	90.43	91.96
DD11	85	85.28	85	85.09	89.33	82.67	74.67	82.22	91.09	89.35	89.13	89.86
DD25	88.33	87.78	87.22	87.78	80	73.33	73.33	75.55	92.39	88.26	87.17	89.27
DD50	78.61	76.94	76.67	77.41	84	80	77.33	80.44	89.13	85.22	84.78	86.38

obtains slightly lower score than the best, but they perform better than all the other color descriptors. The descriptors (C , W , N , E , and H) based on physical reflectance model focus on edge information of color images; hence, it is suitable for ALOI image data with obvious image contour information. SIFT-based color descriptors except for SIFT give the lower score than the other descriptors.

From the results of descriptors in Phos image data, we can find that H , whose property is related to the hue of material, gives the highest score, OPP performs well, i.e., lower than H , but better than most of the other descriptors. Note that CM performs worse than many other descriptors for its' lowest RR. CM does not work well for image data with multiple objects in each image and has limited invariance to illumination changes. CMI performs best among the mathematical-based and CN gives the best results among the information-theory-based.

For RawFoot image data based on texture information, we can observe that Hue-SIFT, which is invariant to light intensity change and shift [36], has the best performance with RR higher than 93%, followed by OPP, W , and the other descriptors. Note that Hue-SIFT performs best among descriptors based on color space selection, W performs best among the physical-model-based, CMI performs best among

the mathematical-based, and FCN performs best among the information-theory-based. Benavente et al. [61] pointed out that the color assignment of FCN is based on physical color samples, which may result in some error when it applied to other stimuli like light.

(ii) Outdoor image data

Table 2 presents the results for THRI2015 image data. OPP obtain the highest score, and a slightly lower score is obtained by HUE. In addition, the CN, FCN, DD25, and DD50 also perform well for the outdoor data. HSV-SIFT, which has poor performance in indoor image data, achieves the lowest recognition rate for below 20 percent. It can also verify the previous conclusion that the combination of HSV color space does not have any robustness. Most color descriptors derived from the physical reflectance model have relatively better performance. The fact that CMI performs much better than CM is consistent with previous results.

4.2. *Experiment on Image Retrieval.* Table 3 presents the image retrieval results for all the image data.

(i) Indoor image data

TABLE 2: Performance comparison of color descriptors in THRI2015.

Descriptors	Recognition rate (RR) (%)			Mean RR (%)
	$K = 1$	$K = 3$	$K = 5$	
RGB-hist	70.83	75	45.83	63.89
Rg-hist	83.33	54.17	70.83	69.44
Hue-hist	50	37.5	29.17	38.89
OPP-hist	20.83	25	25	23.61
SIFT	50	54.17	45.83	50
HSV-SIFT	12.5	12.25	16.67	13.81
rg-SIFT	45.83	45.83	33.33	41.66
OPP-SIFT	54.7	45.83	41.67	47.4
C-SIFT	50	33.33	37.5	40.28
RGB-SIFT	50	37.5	25	37.5
Hue-SIFT	70.83	79.17	79.17	76.39
HUE	95.83	91.67	91.67	93.06
OPP	100	100	91.67	97.22
H	83.33	62.5	66.67	70.83
C	87.5	87.5	75	83.33
W	75	50	45.83	56.94
N	83.33	79.17	75	79.17
E	95.83	87.5	87.5	90.28
CM	83.33	62.5	54.17	66.67
CMI	95.83	66.67	66.67	76.39
CN	100	91.67	83.33	91.67
FCN	100	91.67	87.5	93.06
DD11	91.67	83.33	75	83.33
DD25	95.83	95.83	83.33	91.66
DD50	100	95.83	75	90.28

For ALOI image data, w gives the best score for ANAR = 0, followed by N , C , and other color descriptors. HSV-SIFT fails on this image data for its' highest value of ANAR. We can observe that performance of the information-theory-based is better than other types of color descriptors.

For Phos image data, it is observed that H performs significantly better than other color descriptors, HUE and C obtain a slightly higher value of ANAR than H , but they perform better than all the other descriptors. RGB-histogram gives the largest value of ANAR, which reflects the lack of robustness to illumination variation. Note that Hue histogram, SIFT, HSV-SIFT, E , CM, and DD50 perform worse than many other descriptors.

For RawFoot image data in the context of image retrieval, W obtains the best results with ANAR = 0.0661. E , CM, and DD50 fail on this example with the value of ANAR higher than 0.3. Hue-SIFT also performs well with ANAR = 0.0918, i.e., not as good as W , but better than all other descriptors.

(ii) Outdoor image data

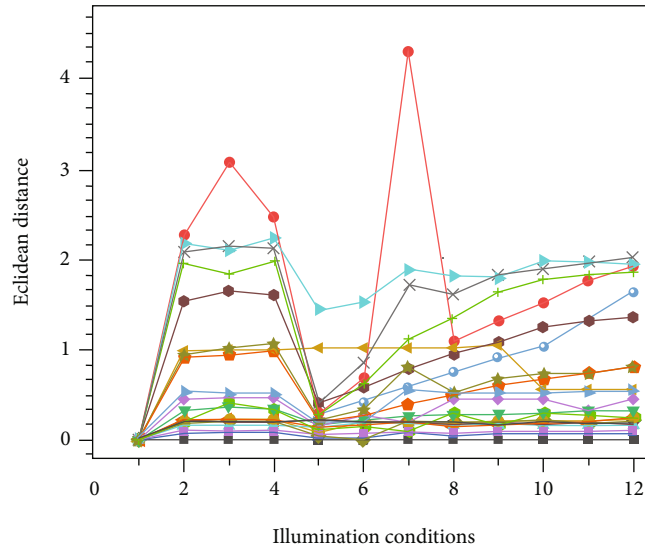
For THRI2015 image data, it can be seen that the Hue-SIFT performs the best with the lowest ANAR = 0.0475. In

TABLE 3: ANAR of color descriptors for all the image data.

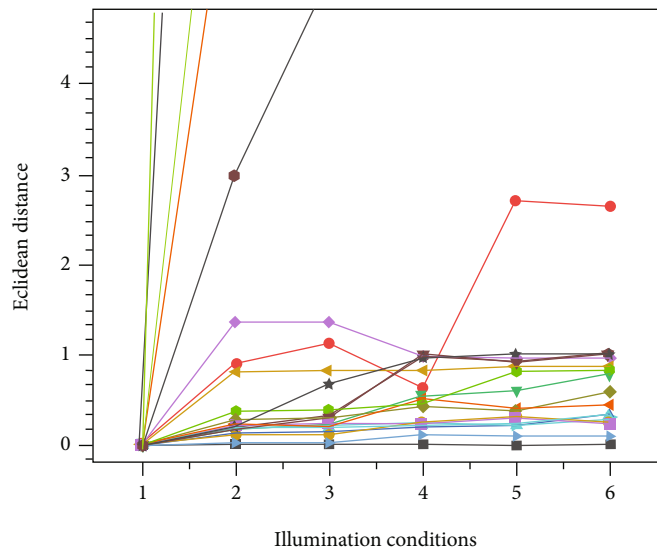
Descriptors	ANAR			
	ALOI	Phos	RaWFoot	THRI2015
RGB-hist	0.2086	0.4000	0.2885	0.22
Rg-hist	0.3747	0.1276	0.2446	0.175
Hue-hist	0.2704	0.3231	0.2789	0.3675
OPP-hist	0.3142	0.1933	0.2896	0.3750
SIFT	0.0212	0.3129	0.2857	0.3850
HSV-SIFT	0.3284	0.3300	0.2260	0.3825
rg-SIFT	0.1597	0.1849	0.1863	0.2975
OPP-SIFT	0.1390	0.1735	0.1695	0.32
C-SIFT	0.2772	0.1233	0.2365	0.3450
RGB-SIFT	0.0964	0.2435	0.1868	0.3325
Hue-SIFT	0.1261	0.1727	0.0918	0.0475
HUE	0.1630	0.0714	0.2124	0.19
OPP	0.0236	0.1186	0.2896	0.1275
H	0.2733	0.0578	0.2554	0.36
C	0.0061	0.0812	0.1684	0.28
W	0	0.1012	0.0661	0.3750
N	0.0028	0.1008	0.1865	0.32
E	0.0163	0.3118	0.3107	0.2950
CM	0.0258	0.3961	0.3258	0.3075
CMI	0.0101	0.2345	0.2609	0.12
CN	0.0448	0.1155	0.2253	0.23
FCN	0.0114	0.2353	0.2818	0.2325
DD11	0.0162	0.2769	0.2769	0.28
DD25	0.0108	0.2265	0.3094	0.3350
DD50	0.0696	0.3078	0.3135	0.26

addition, CMI also has better performance than other descriptors for ANAR = 0.12, respectively. When focusing on the comparison of the same type descriptors, the rg-histogram has better performance in histogram-based color descriptors. After Hue-SIFT, OPP obtains the second lowest ANAR value for descriptors based on color space selection. For descriptors derived from the information theory, the CN performs best and a slightly larger score is obtained by the fuzzy color name. Note that CMI performs well.

Observing the results in Figure 4(a) for ALOI image data, CMI performs best for its minimum Euclidean distance between features under different illumination conditions, and the C and W also have better performance for its Euclidean distances are less than 0.01. It can be seen that the CM has the worst performance for its extremely unstable Euclidean distance under different illumination conditions; the DD, OPP, HUE, and Hue histogram also perform poorly for its Euclidean distance are greater than 1. However, from the results in Figure 4(b), the FCN, CN, and DD perform worst for its Euclidean distance are greater than 8 under different illumination conditions, and the CMI still performs best with lowest variation of Euclidean distances under different illumination conditions; in addition, the Hue-SIFT also has better invariance for its Euclidean distance are lower than 0.1.



(a)



(b)

FIGURE 4: Illumination invariance measurement of color descriptors. (a) Indoor scene. (b) Outdoor scene.

All in all, CMI, which combines the powers of pixel coordinates and their intensities in each color bands, uniformly describes the shape and color distribution of images and has strong robustness to illumination variations. The color descriptors derived from information theory, whose limitation

is the fewer vocabulary of color names, obtain high discriminative power at the expense of photometric invariance.

In the unknown image scenes, we recommend giving priority to CMI for its robustness to illumination variation in the context of image recognition and retrieval.

TABLE 4: The mean RR and p value of a paired T -test on dataset SFU321.

Dimensions	Descriptors	SFU 321							
		Mondrian		Metallic.		Fluorescent		Specular	
		Mean RR	p value	Mean RR	p value	Mean RR	p value	Mean RR	p value
300	C	99.25	0.28	100	—	100	—	97.92	0.26
300	W	99.4	—	98.21	0.70	100	0.47	97.96	0.22
300	N	99.1	0.26	100	0.65	100	0.55	99.32	0.24
300	E	98.06	0.32	96.65	0.55	99.42	0.58	100	—
300	H	96.41	0.29	61.52	0.53	90.06	0.64	91.16	0.34
300	OPP-SIFT	44.39	0.35	34	0.66	86.55	0.25	33.67	0.33
300	Hue-SIFT	55.01	0.28	56.15	0.39	84.8	0.28	57.14	0.17
300	rg-SIFT	43.05	0.38	53.53	0.67	84.13	0.26	61.9	0.29
300	RGB-SIFT	29.38	0.39	31.06	0.55	81.87	0.32	47.62	0.31
300	C-SIFT	39.01	0.36	48.1	0.60	65.5	0.26	52.04	0.29
300	HSV-SIFT	28.85	0.40	34.23	0.45	52.63	0.31	21.77	0.33
300	SIFT	21.97	0.43	31.99	0.59	52.04	0.33	44.22	0.29
16 × 16	rg-hist	45.59	0.18	72.26	0.30	79.51	0.18	88.78	0.29
16 × 16	Hue-hist	35.73	<0.01	31.99	0.04	71.93	<0.01	48.54	0.11
16 × 16 × 16	OPP-hist	44.39	<0.01	20.36	0.17	48.54	0.17	31.97	0.26
16 × 16 × 16	RGB-hist	19.58	<0.01	30.87	0.14	23.98	0.04	24.15	0.23
24	CMI	78.92	0.12	84.79	0.12	100	0.16	97.28	0.12
30	CM	44.89	<0.01	59.34	<0.01	66.06	<0.01	54.7	<0.01
11	OPP	97.31	0.47	96.42	0.52	98.87	0.48	97.86	0.33
11	Hue	83.56	0.15	93.06	0.08	94.35	0.27	92.86	0.10
50	DD50	86.1	0.52	95.97	0.33	96.05	0.50	88.1	0.41
25	DD25	83.86	0.38	92.62	0.28	86.44	0.44	79.25	0.40
11	DD11	78.48	0.37	89.93	0.36	80.79	0.29	74.49	0.26
11	CN	85.2	0.46	89.26	0.4	92.09	0.48	90.47	0.42
11	FCN	86.4	0.45	87.47	0.38	90.49	0.57	85.37	0.37

4.3. *T-Test Measurement.* In order to further analyze the performance of color descriptors, Tables 4 and 5 show the p value of a paired T -test, comparing the best color descriptors in each image data with the remaining descriptors. From the comparison of vertical directions in the table, when $p > 0.05$, it reflects that the differences between a color descriptor and the best one are not significant. In Table 4, we can observe that the differences of hue-hist, OPP-hist, and RGB-hist compared with the best descriptor W is significant ($p < 0.01$) for Mondrian, which indicates performance of the histogram-based except rg-hist is poor, whereas the differences between the best descriptor C and hue-hist are not significant ($p < 0.01$) for fluorescent. In addition, there is a significant difference between CM and the best descriptors for all the surface characteristics, which reflects that CM has limited robustness to illumination variation. From the results of Table 5, the differences between the best descriptor N and the remaining descriptors are not significant ($p > 0.05$) except CM ($p < 0.01$) for ALOI image data, and the differences of hue-hist, OPP-hist, RGB-hist, CM, and CMI compared with the best descriptor H are significant ($p < 0.05$) for Phos image data, whereas for outdoor image data, a paired T -test comparing the OPP with the other descriptors

shows a significant difference with all of them except C , N , E , H , Hue-SIFT, rg-SIFT, CMI, Hue, and descriptors based on information theory.

5. Discussion and Conclusion

In this paper, in order to compare the robustness of color descriptors with different scenes and different surface characteristics to illumination variation, we have presented a performance evaluation of different color descriptors in the context of image recognition and image retrieval. The recommended color descriptors for different surface characteristic and different image data are shown in the following:

- (i) Mondrian: W
- (ii) Metallic: N
- (iii) Fluorescent: CMI
- (iv) Specular: E
- (v) ALOI image data: CMI/ W
- (vi) Phos image data: H

TABLE 5: The mean RR and p value of a paired T -test on dataset ALOI, Phos, RawFooT, and THRI2015.

Dimensions	Descriptors	ALOI.		Indoor Phos		RawFooT		Outdoor THRI2015	
		Mean RR	p value	Mean RR	p value	Mean RR	p value	Mean RR	p value
300	C	100	0.54	88.89	0.38	90.8	0.42	83.33	0.23
300	W	100	0.55	86.66	0.34	93.19	0.49	56.94	0.01
300	N	100	—	91.11	0.28	76.52	0.53	79.17	0.13
300	E	99.72	0.46	87.11	0.33	90.29	0.44	90.28	0.32
300	H	97.87	0.52	94.67	—	70.87	0.41	70.83	0.06
300	OPP-SIFT	38.15	0.41	46.67	0.25	74.27	0.02	47.4	<0.01
300	Hue-SIFT	66.76	0.51	91.56	0.26	94.78	—	76.39	0.06
300	rg-SIFT	52.87	0.44	60.89	0.37	76.01	0.11	41.66	0.08
300	RGB-SIFT	35.84	0.43	46.67	0.24	76.16	0.03	37.5	<0.01
300	C-SIFT	33.80	0.47	40.45	0.31	41.81	0.17	40.28	0.04
300	HSV-SIFT	37.32	0.46	42.67	0.28	75.44	0.11	13.81	<0.01
300	SIFT	90.83	0.45	33.78	0.42	62.61	0.11	50	0.02
16 × 16	rg-hist	78.15	0.29	88	0.53	65.87	0.12	69.44	<0.01
16 × 16	Hue-hist	85.65	0.35	50.67	<0.01	74.20	0.21	38.89	<0.01
16 × 16 × 16	OPP-hist	78.24	0.29	79.11	<0.01	60.06	0.27	23.61	<0.01
16 × 16 × 16	RGB-hist	95.46	0.33	64.44	<0.01	73.62	0.22	63.89	<0.01
24	CMI	100	0.16	58.67	0.02	70.58	0.52	76.39	0.10
30	CM	91.57	<0.01	10.66	<0.01	57.97	0.03	66.67	<0.01
11	OPP	84.54	0.45	92.44	0.24	94.28	0.52	97.22	—
11	Hue	86.94	0.48	89.78	0.18	89.21	0.55	93.06	0.12
50	DD50	77.41	0.38	80.44	0.21	86.38	0.57	90.28	0.08
25	DD25	87.78	0.40	75.55	0.24	89.27	0.57	91.66	0.25
11	DD11	85.09	0.38	82.22	0.30	89.86	0.52	83.33	0.57
11	CN	67.86	0.33	86.22	0.47	90.51	0.50	91.67	0.35
11	FCN	68.71	0.37	80.44	0.33	91.96	0.55	93.06	0.57

(vii) RawFooT image data: Hue-sift/ W

(viii) THRI2015 image data: OPP/Hue-SIFT

In summary, we can get the following conclusions:

- (i) Compared with objects with metallic specularities, images containing nonnegligible dielectric specularities have a more negative effect on the performance of color descriptors, and for the objects with fluorescent surface, the performance of most descriptors has been improved significantly. Consequently, color descriptors are more suitable for images with fluorescent surfaces, whereas for images with specular or metal reflections, the highlight part would lead to the performance degradation of pure color descriptors. It can be considered to fuse the color information of images with the shape, texture, and other information, or focus on the highlighted part for further research
- (ii) The color moment invariants (CMI) provide the optimal balance between performance and dimensions in most tests. Hence, CMI can be considered

an alternative when the high dimensionality of color descriptors is an issue

(iii) Color descriptors derived from physical reflectance models are more suitable for object recognition and image retrieval

(iv) From the theoretical and experimental results, it can be concluded that we can transform images into a color space with illumination invariance such as the rg color space and then calculate color descriptors, which can improve the robustness of color descriptors to illumination variations. This also provides an effective way to improve existing methods

Color descriptors play an important role in feature extraction; future research may tend to build more distinctive color descriptors and consider the relationship between image pixels. Combining with deep learning methods is also an important trend. Feature extraction in computer vision does not point to a single feature; integrating color descriptors properly with other features will play a greater role in different application scenarios.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

The authors gratefully acknowledge Chunxiao Li's technical support and theoretical support. This research was supported by "the Fundamental Research Funds for the Central Universities" (Grant number BLX202018) and National Science Foundation of China (Grant number 32071680).

References

- [1] D. H. Foster, "Color constancy," *Vision Research*, vol. 51, no. 7, pp. 674–700, 2011.
- [2] K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms—part I: methodology and experiments with synthesized data," *IEEE Transactions on Image Processing*, vol. 11, no. 9, pp. 972–984, 2002.
- [3] K. Barnard, L. Martin, A. Coath, and B. Funt, "A comparison of computational color constancy algorithms—part II: experiments with image data," *IEEE Trans Image Process*, vol. 11, no. 9, pp. 985–996, 2002.
- [4] P. E. Bradley and B. Jutzi, "Improved feature detection in fused intensity-range images with complex SIFT (CSIFT)," *Remote Sensing*, vol. 3, no. 9, pp. 2076–2088, 2011.
- [5] A. Jafari Malekabadi, M. Khojastehpour, and B. Emadi, "A comparative evaluation of combined feature detectors and descriptors in different color spaces for stereo image matching of tree," *Scientia Horticulturae*, vol. 228, pp. 187–195, 2018.
- [6] D. Berwick and S. W. Lee, "Spectral gradients for color-based object recognition and indexing," *Computer Vision and Image Understanding*, vol. 94, no. 1–3, pp. 28–43, 2004.
- [7] D. Muselet and L. Macaire, "Combining color and spatial information for object recognition across illumination changes," *Pattern Recognition Letters*, vol. 28, no. 10, pp. 1176–1185, 2007.
- [8] B. Li, D. Xu, W. H. Xiong, and S. H. Feng, "Illumination-independent descriptors using color moment invariants," *Optical Engineering*, vol. 48, no. 2, article 027005, 2009.
- [9] M. Dhekane, A. Seal, and P. Khanna, "Illumination and expression invariant face recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 31, no. 12, article 1756018, 2017.
- [10] Y. Cheng, L. B. Jiao, X. H. Cao, and Z. Y. Li, "Illumination-insensitive features for face recognition," *The Visual Computer*, vol. 33, no. 11, pp. 1483–1493, 2017.
- [11] D. J. Kim, M. K. Shon, S. Lee, and E. Kim, "Illumination-robust face recognition approach using enhanced preprocessing and feature extraction," *Journal of Nanoelectronics and Optoelectronics*, vol. 11, no. 2, pp. 141–147, 2016.
- [12] X. Cao, W. Shen, L. G. Yu, Y. L. Wang, J. Y. Yang, and Z. W. Zhang, "Illumination invariant extraction for face recognition using neighboring wavelet coefficients," *Pattern Recognition*, vol. 45, no. 4, pp. 1299–1305, 2012.
- [13] O. Arandjelovic, "Colour invariants under a non-linear photometric camera model and their application to face recognition from video," *Pattern Recognition*, vol. 45, no. 7, pp. 2499–2509, 2012.
- [14] G. Y. Zhang, J. Liu, Y. Liu, J. Zhao, L. Tian, and Y. Q. Chen, "Physical blob detector and multi-channel color shape descriptor for human detection," *Journal of Visual Communication and Image Representation*, vol. 52, pp. 13–23, 2018.
- [15] H. Y. Lee, "Spatial color descriptor for image retrieval and video segmentation," *IEEE Transactions on Multimedia*, vol. 5, pp. 358–367, 2003.
- [16] A. Talib, "A weighted dominant color descriptor for content-based image retrieval," *Journal of Visual Communication and Image Representation*, vol. 24, pp. 345–360, 2013.
- [17] B. S. Phadikar, "Content-based image retrieval in DCT compressed domain with MPEG-7 edge descriptor and genetic algorithm," *Pattern Analysis and Applications*, vol. 21, pp. 469–489, 2018.
- [18] X. Y. Wang, "Image retrieval based on exponent moments descriptor and localized angular phase histogram," *Multimedia Tools and Applications*, vol. 76, no. 6, pp. 7633–7659, 2017.
- [19] J. Y. Chang, C. J. Lian, H. C. Fang, and L. G. Chen, "Architecture and analysis of color structure descriptor for real-time video indexing and retrieval," in *Advances in Multimedia Information Processing - PCM*, pp. 130–137, Springer, 2004.
- [20] A. M. Ferman, A. M. Tekalp, and R. Mehrotra, "Robust color histogram descriptors for video segment retrieval and identification," *IEEE Transactions on Image Processing*, vol. 11, no. 5, pp. 497–508, 2002.
- [21] J. Z. Chen, Q. Li, Q. Peng, and K. H. Wong, "CSIFT based locality-constrained linear coding for image classification," *Pattern Analysis and Applications*, vol. 18, no. 2, pp. 441–450, 2015.
- [22] S. Banerji, A. Sinha, and C. Liu, "New image descriptors based on color, texture, shape, and wavelets for object and scene image classification," *Neurocomputing*, vol. 117, pp. 173–185, 2013.
- [23] N. Hor, "Image retrieval approach based on local texture information derived from predefined patterns and spatial domain information," *International Journal of Computer Science Engineering*, vol. 8, pp. 246–254, 2019.
- [24] R. Khwildi and A. Ouled Zaid, "HDR image retrieval by using color-based descriptor and tone mapping operator," *The Visual Computer*, vol. 36, no. 6, pp. 1111–1126, 2020.
- [25] I. Jeena Jacob, K. G. Srinivasagan, P. Ebby Darney, and K. Jayapriya, "Deep learned inter-channel colored texture pattern: a new chromatic-texture descriptor," *Pattern Analysis and Applications*, vol. 23, no. 1, pp. 239–251, 2020.
- [26] E. K. Kumar, P. Kishore, M. T. Kiran Kumar, and D. A. Kumar, "3D sign language recognition with joint distance and angular coded color topographical descriptor on a 2-stream CNN," *Neurocomputing*, vol. 372, pp. 40–54, 2020.
- [27] X. Wei, K. Jia, J. Lan, Y. Li, Y. Zeng, and C. Wang, "Automatic method of fruit object extraction under complex agricultural background for vision system of fruit picking robot," *Optik*, vol. 125, no. 19, pp. 5684–5689, 2014.
- [28] G. Wu, B. Li, Q. Zhu, M. Huang, and Y. Guo, "Using color and 3D geometry features to segment fruit point cloud and improve fruit recognition accuracy," *Computers and Electronics in Agriculture*, vol. 174, article 105475, 2020.
- [29] M. K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, pp. 179–187, 1962.
- [30] A. Khotanzad and Y. H. Hong, "Invariant Image Recognition by Zernike Moments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 489–497, 1990.

- [31] I. Dhillon, "A divisive information theoretic feature clustering algorithm for text classification," *Journal of Machine Learning Research*, vol. 3, pp. 1265–1287, 2003.
- [32] S. Fekri-Ershad and F. Tajeripour, "Impulse-noise resistant color-texture classification approach using hybrid color local binary patterns and Kullback–Leibler divergence," *The Computer Journal*, vol. 60, no. 11, pp. 1633–1648, 2017.
- [33] K. Bowyer, *Empirical Evaluation Techniques in Computer Vision*, IEEE Computer Society Press, 1998.
- [34] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [35] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [36] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1582–1596, 2010.
- [37] C. C. Chang and C. J. Lin, "LIBSVM," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.
- [38] G. J. Burghouts and J. M. Geusebroek, "Performance evaluation of local colour invariants," *Computer Vision and Image Understanding*, vol. 113, no. 1, pp. 48–62, 2009.
- [39] A. Setkov, M. Gouiffès, and C. Jacquemin, "Evaluation of color descriptors for projector-camera systems," *Journal of Visual Communication and Image Representation*, vol. 36, pp. 11–27, 2016.
- [40] F. Bianconi, "Theoretical and experimental comparison of different approaches for color texture classification," *Journal of Electronic Imaging*, vol. 20, no. 4, article 043006, 2011.
- [41] C. Cusano, P. Napoletano, and R. Schettini, "Evaluating color texture descriptors under large variations of controlled lighting conditions," *Journal of the Optical Society of America A*, vol. 33, no. 1, pp. 17–30, 2016.
- [42] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
- [43] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: a survey," *International Journal of Computer Vision*, 2020.
- [44] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *International Journal of Computer Vision*, vol. 127, no. 5, pp. 512–531, 2019.
- [45] X. Jiang, J. Ma, J. Jiang, and X. Guo, "Robust feature matching using spatial clustering with heavy outliers," *IEEE Transactions on Image Processing*, vol. 29, pp. 736–746, 2020.
- [46] A. Koschan and M. A. Abidi, *Digital Color Image Processing*, Wiley-Interscience, Hoboken, NJ, USA, 2008.
- [47] G. N. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods*, Wiley, New York, NY, USA, 1967.
- [48] J. M. Geusebroek and R. V. D. Boomgaard, "Color invariance," *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 23, pp. 1338–1350, 2001.
- [49] S. A. Shafer, "Using color to separate reflection components," *Color Research & Application*, vol. 10, no. 4, pp. 210–218, 1985.
- [50] J. Flusser, "On the independence of rotation moment invariants," *Pattern Recognition*, vol. 33, no. 9, pp. 1405–1410, 2000.
- [51] J. Flusser and T. Suk, "Rotation moment invariants for recognition of symmetric objects," *IEEE Trans Image Process*, vol. 15, no. 12, pp. 3784–3790, 2006.
- [52] L. Van Gool, T. Moons, E. Pauwels, and A. Oosterlinck, "Vision and Lie's approach to invariance," *Image Vision Computing*, vol. 13, no. 4, pp. 259–277, 1995.
- [53] J. van de Weijer, T. Gevers, and A. D. Bagdanov, "Boosting color saliency in image feature detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 150–156, 2006.
- [54] A. Bosch, A. Zisserman, and X. Muñoz, "Scene classification using a hybrid generative/discriminative approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 712–727, 2008.
- [55] J. V. D. Weijer and C. Schmid, "Coloring local feature extraction," in *European Conference on Computer Vision (ECCV)*, pp. 334–348, Graz, Austria, 2006.
- [56] B. V. Funt and G. D. Finlayson, "Color constant color indexing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 522–529, 1995.
- [57] T. Gevers and A. Smeulders, "Color-based object recognition," *Pattern Recognition*, vol. 32, no. 3, pp. 453–464, 1999.
- [58] J. V. D. Weijer and C. Schmid, "Blur robust and color constant image description," in *2006 International Conference on Image Processing*, pp. 993–996, Atlanta, GA, USA, 2006.
- [59] F. Mindru, T. Tuytelaars, L. V. Gool, and T. Moons, "Moment invariants for recognition under changing viewpoint and illumination," *Computer Vision and Image Understanding*, vol. 94, no. 1-3, pp. 3–27, 2004.
- [60] B. Berlin and P. Kay, *Basic Color Terms :Their Universality and Evolution*, University of California Press, Berkeley, CA, USA, 1969.
- [61] R. Benavente, M. Vanrell, and R. Baldrich, "Parametric fuzzy sets for automatic color naming," *Journal of the Optical Society of America A*, vol. 25, no. 10, pp. 2582–2593, 2008.
- [62] J. van de Weijer, C. Schmid, J. Verbeek, and D. Larlus, "Learning color names for real-world applications," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1512–1523, 2009.
- [63] R. Khan, J. V. D. Weijer, and F. S. Khan, "Discriminative color descriptors," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2866–2873, Portland, OR, USA, 2013.
- [64] G. Csurka, C. Dance, L. Fan, J. Williamowski, and C. Bray, "Visual categorization with bags of keypoints," in *2011 4th International Congress on Image and Signal Processing*, pp. 1–22, Shanghai, China, 2004.
- [65] K. Barnard, L. Martin, B. Funt, and A. Coath, "A data set for color research," *Color Research and Application*, vol. 27, no. 3, pp. 147–151, 2002.
- [66] http://www.cs.sfu.ca/~colour/data/colour_constancy_test_images/index.html.
- [67] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, "The Amsterdam Library of object images," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 103–112, 2005.
- [68] V. Vonikakis, D. Chrysostomou, R. Kouskouridas, and A. Gasteratos, "A biologically inspired scale-space for illumination invariant feature detection," *Measurement Science and Technology*, vol. 24, no. 7, article 074024, 2013.