

Evaluating color texture descriptors under large variations of controlled lighting conditions

CLAUDIO CUSANO,^{1,*} PAOLO NAPOLETANO,² AND RAIMONDO SCETTINI²

¹Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Via Ferrata 1, 27100 Pavia, Italy

²Department of Informatics, Systems and Communication, University of Milan-Bicocca, Viale Sarca 336, 20126 Milano, Italy

*Corresponding author: claudio.cusano@unipv.it

Received 7 August 2015; revised 23 October 2015; accepted 25 October 2015; posted 28 October 2015 (Doc. ID 247474); published 9 December 2015

The recognition of color texture under varying lighting conditions remains an open issue. Several features have been proposed for this purpose, ranging from traditional statistical descriptors to features extracted with neural networks. Still, it is not completely clear under what circumstances a feature performs better than others. In this paper, we report an extensive comparison of old and new texture features, with and without a color normalization step, with a particular focus on how these features are affected by small and large variations in the lighting conditions. The evaluation is performed on a new texture database, which includes 68 samples of raw food acquired under 46 conditions that present single and combined variations of light color, direction, and intensity. The database allows us to systematically investigate the robustness of texture descriptors across large variations of imaging conditions. © 2015 Optical Society of America

OCIS codes: (100.2960) Image analysis; (100.5010) Pattern recognition; (100.3008) Image recognition, algorithms and filters; (330.1690) Color.

<http://dx.doi.org/10.1364/JOSAA.33.000017>

1. INTRODUCTION

The role of color in texture classification has been widely debated in the literature. Despite the number and depth of the experimental verifications, it is still not completely clear how much and under what circumstances color information is beneficial. Notable examples of this kind of analysis are in the work by Drimborean and Whelan [1], by Mäenpää and Pietikäinen [2], and by Bianconi *et al.* [3]. They all observed how color can be effective—but only in those cases where illumination conditions do not vary too much between training and test sets. In fact, methods that exploit color information greatly suffer variations in the color of the illuminant. Under these circumstances, the best result is often achieved simply by disregarding color, that is, by reducing all the images to gray scale. The degree of intraclass variability of the images in Fig. 1 suggests why color information, if not properly processed, easily can be deceptive.

A possible strategy to exploit color in texture classification consists in the extraction of image features that are invariant (or at least robust) with respect to changes in the illumination. In scene and object recognition, the approach of specially designing invariant features is rapidly becoming obsolete in favor of features automatically learned from a large amount of data with methods based on deep learning [4]. It is not clear if the same is going to happen in texture recognition as well. A recent

work [5] suggests that a hybrid approach (local features extracted from a convolutional neural network and then aggregated as Fisher vectors) can be the most successful.

The availability of suitable databases of texture images is of primary importance for the research in this field. Therefore, in the past, several texture databases have been collected to assess the performance of texture recognition methods under a variety of conditions. These databases are often focused on the exploration of the variability of texture images under specific variations of imaging conditions [6,7], mostly related to variations in the geometry of the acquisition setup (with little or no variation about the characteristics of the illuminant). For instance, several texture databases include images where the same samples are taken from different points of view. As a result, the images depict the same textures taken at different scales and orientations. By contrast, a more recent work proposed a database of “textures in the wild” [8] to allow texture analysis in completely uncontrolled environments. This approach allows us to implicitly verify the robustness against a multitude of source variations simultaneously. Even though the results on this kind of data set may provide a better indication about the “average” performance of texture recognition methods in several real-world applications, they do not allow a clear analysis of their strengths and weaknesses for specific setups. In fact, there are several application domains where acquisition conditions are

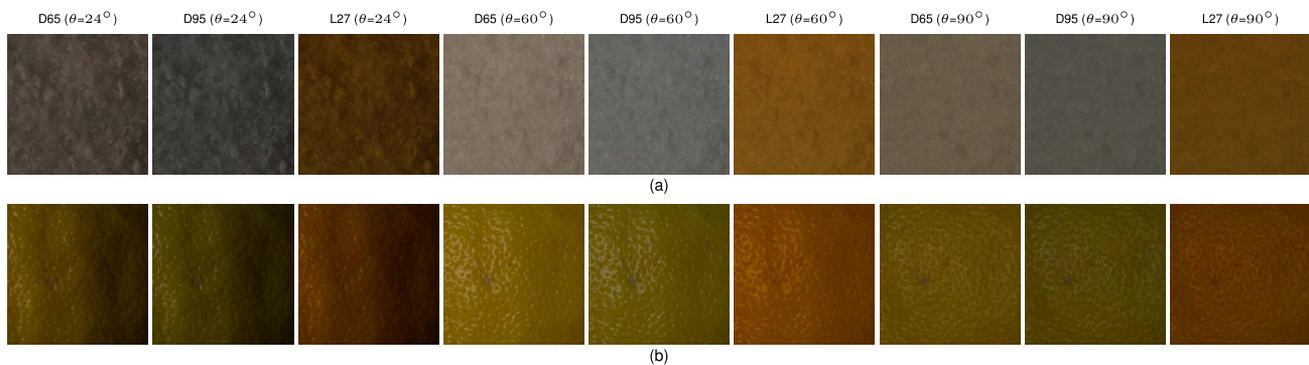


Fig. 1. Examples of two different textures acquired under nine different light conditions. (a) Salt and (b) grapefruit.

indeed very controlled (medical imaging, industrial inspection) and for which the uncertainty inherent to the experimentation in the wild is a serious liability.

In this paper, we address the problem of texture classification under large variations of controlled lighting conditions. We have evaluated and compared several texture and color descriptors with respect to single and combined changes in the lighting conditions. We selected three classes of visual descriptors. The first class includes traditional (hand crafted) descriptors specially designed for texture analysis. The second one includes features that were specially designed for object recognition. The third class includes those features that correspond to intermediate representations computed by convolutional neural networks (CNNs).

Because we addressed the problem of texture classification under varying lighting conditions, we also investigated the use of color normalization methods as preprocessing, thus quantifying how much their application influences the performance of the different descriptors.

Existing texture databases [6,7] do not include, in general, large variations of lighting conditions, and, in particular, they do not allow us to evaluate the goodness of visual descriptors with respect to single and combined lighting condition changes, such as only direction or temperature of the light, direction and temperature of the light, etc. Due to these reasons, we collected a new texture database, which we named “Raw Food Texture” database (RawFooT). This database includes several samples of raw food acquired under 46 conditions differing in light color, direction, and intensity. We choose to focus on raw food because, similarly to other natural surfaces, its color is an intrinsic property. Therefore, the task of classifying this kind of textures does not include the semantic ambiguities that, instead, may arise when dealing with artificial surfaces, where it is possible to argue that samples of different color may be instances of the same class of “textures.” As far as we know, the proposed database is the one featuring the largest amount of variations in the lighting conditions, and the only one where color, direction, and intensity of light are subject to systematic and independent changes. The database is available online at Ref. [9].

In this paper, we address the following issues:

- How effective are handcrafted texture descriptors when acquisition condition variations are so large?

- Can object-recognition descriptors achieve high classification accuracy on pure texture images?
- Do CNN-based descriptors confirm to be powerful also on texture classification tasks?
- Can CNN-based descriptors handle large variations in lighting conditions?
- Is color normalization helpful for texture analysis in case of changes in lighting conditions?

The rest of the paper is organized as follows: Section 2 describes the proposed RawFooT database and compares it against other publicly available data sets; Section 3 reviews the main texture descriptors in the state of the art; Section 4 describes the experimental setup and Section 5 reports the results obtained; finally, Section 6 presents our final considerations and discusses some new directions for our future research on this topic.

2. TEXTURE DATABASES

In the last few years, different research groups have developed a number of databases of texture images ranging from natural textures to man-made materials [6,7,10]. Each database has been designed to study one or several aspects about textures: invariance to acquisition device, invariance to lighting conditions, invariance to image rotation or scale, 3D reconstruction, computer graphics, classification, segmentation, etc. The problems of texture classification and material recognition are closely related. In this paper, we mainly focused on the former because the two problems may require different strategies [11].

We considered the most important texture databases that have been presented in the literature [6,7]; thus, we compiled Table 1 to highlight the most important features of each database. Among the features, the most important are those related to the acquisition setup, such as illumination conditions, sensor angle, image rotation, scaling, and color of the illuminant. In particular, we highlighted four different sources of variations in illumination conditions. One related to the direction of the light, one to the intensity of the light, another to the color temperature of the light, and the fourth related to a mixture of variations, such as temperature and direction, temperature and intensity, etc. As is seen in the table, several databases consider a mixture of variations. The most notable is the OuTex database [17], which is, in fact, the most used in the study of descriptors invariant to lighting conditions [2]. The OuTex

Table 1. Main Features of Existing Texture Databases [6,7]^a

Database	Type of Surface	#Classes #Images	Image Size	Color Repr.	Rotation	Scaling	Light Directions	Light Intensity	Light Temp.	Lighting Variability	Sensor's Angle	Camera Sensor
Brodatz [12]	Natural, artificial	112 classes 112 images	512 × 512	Gray-scale	•	◦	◦	◦	◦	• Controlled unknown	Scanned	–
CUReT [13]	Natural, artificial	61 classes over 14,000 images	512 × 512	Color RGB	•	◦	• 55	◦	◦	◦	7 positions	3 separate CCD
VisTex [14]	Natural, artificial	54 classes 864 images	512 × 512	–	◦	◦	• Unknown directions	◦	◦	• Uncontrolled: daylight, fluorescent, incandescent	Frontal, oblique	–
MeasTex [15]	Natural, artificial	4 classes 944 images	512 × 512	Gray-scale	•	◦	◦	◦	◦	• Daylight direct and indirect, flash	Frontal, oblique	Scanned from 35 mm film
PhoTex [16]	Rough surfaces	64 classes	1280 × 1024	Gray-scale	•	◦	• 4	◦	◦	◦	12 positions	Vosskuhler CCD
OutTex [17]	Natural, artificial	320 classes 51,840 images	516 × 716	Color RGB	•	•	◦	◦	◦	• Fluorescent, t184, incandescent at 3 positions	Frontal	3 CCD, Sony DXC-755P
DynTex [18]	Natural, artificial	36 classes 345 sequences	720 × 576 25 fps	Color RGB	•	◦	◦	◦	◦	◦	–	3 CCD, Sony DCR-TRV890E TRV900E
UIUC [19]	Natural, artificial	25 classes 1000 images	640 × 480	Gray-scale	◦	•	◦	◦	◦	◦	Various	–
KTH-TIPS2 [20]	Natural, food, artificial	44 classes 4608 images	1280 × 960	Color RGB	•	•	• 3	◦	◦	• Incandescent, fluorescent	Frontal, oblique	Olympus C-3030ZOOM
ALOT [21]	Natural, food, artificial	250 classes, over 27,500 images		Color RGB	•	◦	• 5	◦	•	◦	4 positions	Foveon X3 3CMOS
Mondial Marmi [22]	Granite	25 classes	1500 × 1500	Color RGB	•	◦	◦	◦	◦	• White HB LEDs	Frontal	Samsung S850
STex [23]	Natural, artificial	476 classes	1024 × 1024	Color RGB	◦	◦	◦	◦	◦	◦	◦	–
Kylberg-Sintorn [24]	Natural, artificial	25 classes, 135000 images	122 × 122	Gray-scale	•	◦	◦	◦	◦	◦	Frontal	Canon EOS 550D
USPTex [25]	Natural, artificial	191 classes, 2292 images	512 × 384	Color RGB	◦	◦	◦	◦	◦	◦	–	–
Proposed RawFoot	Natural, food	68 classes, 3128 images	800 × 800	Color RGB	◦	◦	• 9	• 5	• 18	• 12	Frontal	Canon 40D

^aThe filled circle indicates that the given feature is present, the empty circle indicates its absence, and the minus sign indicates that information on that feature is not available.

collection includes the OuTex-14 test suite, which contains images that depict textures of 68 different classes acquired under three different light sources, each positioned differently: the 2856 K incandescent CIE A, the 2300 K horizon sunlight, and the 4000 K fluorescent TL84.

Few databases separately consider variations of light direction, intensity, or temperature. In particular, the only database that provides a good number of this kind of variation is the ALOT database [21]. This collection provides 250 classes of textures acquired under several conditions, which were obtained by combining five illumination directions (at 3075 K) and one semi-hemispherical illumination (at 2175 K). Each object was recorded with only one out of five lights turned on, yielding five different illumination angles. One image is recorded with all lights turned on, yielding a sort of hemispherical illumination. All the images are acquired by four cameras positioned differently.

As far as we know, no publicly available texture database has been designed to assess the performance in texture classification under a broad range of variations in the illumination color, direction, and intensity. This is why we collected the RawFooT database.

A. Raw Food Texture Database

The RawFooT database has been specially designed to investigate the robustness of descriptors and classification methods with respect to variations in lighting conditions, with a particular focus on variations in the color of the illuminant. The database includes images of samples of textures, acquired under 46 lighting conditions, which may differ in the light direction, in the illuminant color, in its intensity, or in a combination of these factors.

Psycho-physical studies [26] suggest that in the human visual system, color and pattern information are processed separately. However, it has been observed that their combination can be effective for texture classification. For certain classes of materials, the two kinds of information are clearly independent (e.g., fabrics and other artificial materials). For this reason, we considered samples of texture where the relationship between pattern information and color has not been explicitly designed. Our classes correspond to 68 samples of raw food, including various kinds of meat, fish, cereals, fruit, etc. Therefore, the whole database includes $68 \times 46 = 3128$ images. Figure 2 shows an image of each sample.

Pictures have been acquired in a dark room with a Canon EOS 40D DSLR camera. The camera was placed 48 cm above the sample to be acquired, with the optical axis perpendicular to the surface of the sample. The lenses used had a focal length of 85 mm and a camera aperture of $f/11.3$; each picture has been taken with 4 s of exposition time. For illuminants, we used a pair of monitors (22 in. Samsung SyncMaster LED monitor) positioned above the sample and tilted by 45° , with about 20 cm of space between their upper edges to make room for the camera. By illuminating different regions of the monitors, and, by using different colors (inspired by [27]), we simulated natural and artificial illuminants coming from different directions and at various intensity levels. The two monitors have been colorimetrically characterized using an X-Rite i1 spectral colorimeter in such a way that the device RGB coordinates can be used to accurately render the desired chromaticities. Daylight at 6500 K (D65) has been specified as a white point. Figure 3 shows the setup used for the acquisitions.

For each sample, a program of 46 shots has been followed.



Fig. 2. Overview of the 68 classes included in the Raw Food Texture database. For each class, an image taken under D65 at direction $\theta = 24^\circ$ is shown.

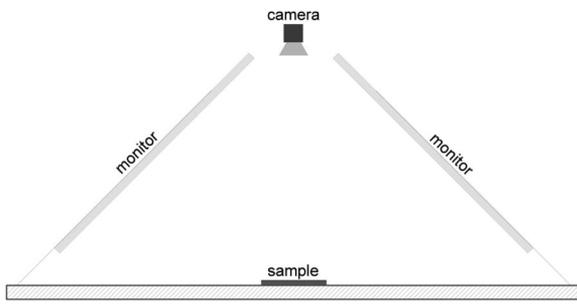


Fig. 3. Setup used to acquire the Raw Food Texture database.

Intensity variations: Four shots have been taken while illuminating the whole monitors with neutral light (D65) at different levels of intensity (100%, 75%, 50%, and 25% of the maximum achievable level).

Light direction: Nine shots have been taken with the light (D65) coming from different angles. In the first eight of these shots, only a band covering 40% of a single monitor has been lit. The angles between the direction of the light coming from the center of the illuminated band and the surface of the sample are 24°, 30°, 36°, 42°, 48°, 54°, 60°, and 66°. For the last shot, two bands covering the upper 20% of each monitor have been lit (because on average the light comes exactly from above the sample, we count it as an angle of 90°).

Daylight: 12 shots have been taken while simulating natural daylight at different color temperatures. To do so, given a color temperature T , we applied the following equations to obtain the corresponding xy chromaticities:

$$\begin{aligned} x &= a_0 + a_1 \frac{10^3}{T} + a_2 \frac{10^6}{T^2} + a_3 \frac{10^9}{T^3}, \\ y &= -3x^2 + 2.87x - 0.275, \end{aligned} \quad (1)$$

where $a_0 = 0.244063$, $a_1 = 0.09911$, $a_2 = 2.9678$, and $a_3 = -4.6070$ if $4000 \text{ K} \leq T \leq 7000 \text{ K}$, and $a_0 = 0.23704$, $a_1 = 0.24748$, $a_2 = 1.9018$, and $a_3 = -2.0064$ if $7000 \text{ K} < T \leq 25000 \text{ K}$ [28]. Chromaticities have been converted in the RGB space with a scaling of the color channels ensuring that their largest value is 255. We considered 12 color temperatures in the range from 4000 to 9500 K with a step of 500 K (we will refer to these as D40, D45, and D95). The whole monitors have been lit during these shots.

Indoor illumination: Six shots have been taken while simulating artificial light with a color temperature of 2700, 3000, 4000, 5000, 5700, and 6500 K on the two whole monitors. We considered LED lights produced by OSRAM, and we computed the corresponding RGB values starting from the chromaticities indicated in the data sheets from the producer's website (<http://www.osram-os.com>). We will refer to these as L27, L30, and L65.

Color and direction: Nine shots have been taken by varying the color and direction of the illuminant. The combinations of three colors (D65, D95, and L27) and of three directions (24°, 60°, and 90°) have been considered.

Multiple illuminants: Three shots have been taken while the sample is illuminated by two illuminants with different colors (D65, D95, or L27). Bands covering the lower 40% of both

monitors have been lit, using two different colors on the two monitors.

Primary colors: Three shots have been taken under pure red, green, and blue illuminants.

Each 3944×2622 picture in the camera space has been converted to standard sRGB, and the final texture images have been obtained by cropping the central region of 800×800 pixels. Figure 4 shows the 46 shots taken for two of the 68 samples. To allow the estimate of the illuminants, we have carried out the program of 46 shots of a 24 squares Macbeth ColorChecker.

3. TEXTURE DESCRIPTORS

A huge variety of texture descriptors have been proposed in the literature. These were traditionally divided into statistical, spectral, structural, and hybrid approaches [29]. Among traditional methods, the most popular are probably those based on histograms, Gabor filters [30], co-occurrence matrices [31], and local binary patterns [32]. These descriptors display different strengths and weaknesses, particularly concerning their invariance with respect to the acquisition conditions.

Traditional descriptors are often designed to capture texture information in uncluttered images taken under controlled conditions. To address those cases where the conditions cannot be controlled, a few attempts have been made to adapt features used for scene or object recognition to the domain of texture classification. For instance, Sharan *et al.* [11] used SIFT and HOG descriptors for material classification, while Sharma *et al.* [33] used a variation of the Fisher vector approach for texture and face classification. Cimpoi *et al.* [8] showed how SIFT descriptors aggregated with the improved Fisher method greatly outperform previous descriptors in the state of the art on a variety of texture classification tasks, including the classification of “textures in the wild.”

Following the trend in image recognition, features extracted from convolutional neural networks (CNNs) have been adopted for texture classification as well. CNNs allow us to leverage very large datasets of labeled images by learning intermediate image representations, which can be used for various image classification problems [34]. For instance, Cimpoi *et al.* [5] used Fisher vectors to pool features computed by a CNN trained for object recognition.

In addition to these general purpose texture descriptors, a variety of descriptors have been specially designed to be robust with respect to specific variations in the acquisition conditions. Khan *et al.* [35], for instance, considered a diagonal/offset model for illumination variations, deduced from it an image normalization transformation, and finally extracted Gabor features from the normalized images. Other color normalization techniques can be used for this purpose. Finlayson *et al.* proposed rank-based features obtained from invariant color representations [36]. Seifi *et al.*, instead, proposed to characterize color textures by analyzing the rank correlation between pixels located in the same neighborhood. The authors obtained a correlation measure, which is related to the colors of the pixels and is not sensitive to illumination changes [37]. Cusano *et al.* [38] proposed a texture descriptor specially designed to deal with the case of variations in the color of the illuminant. The reader can

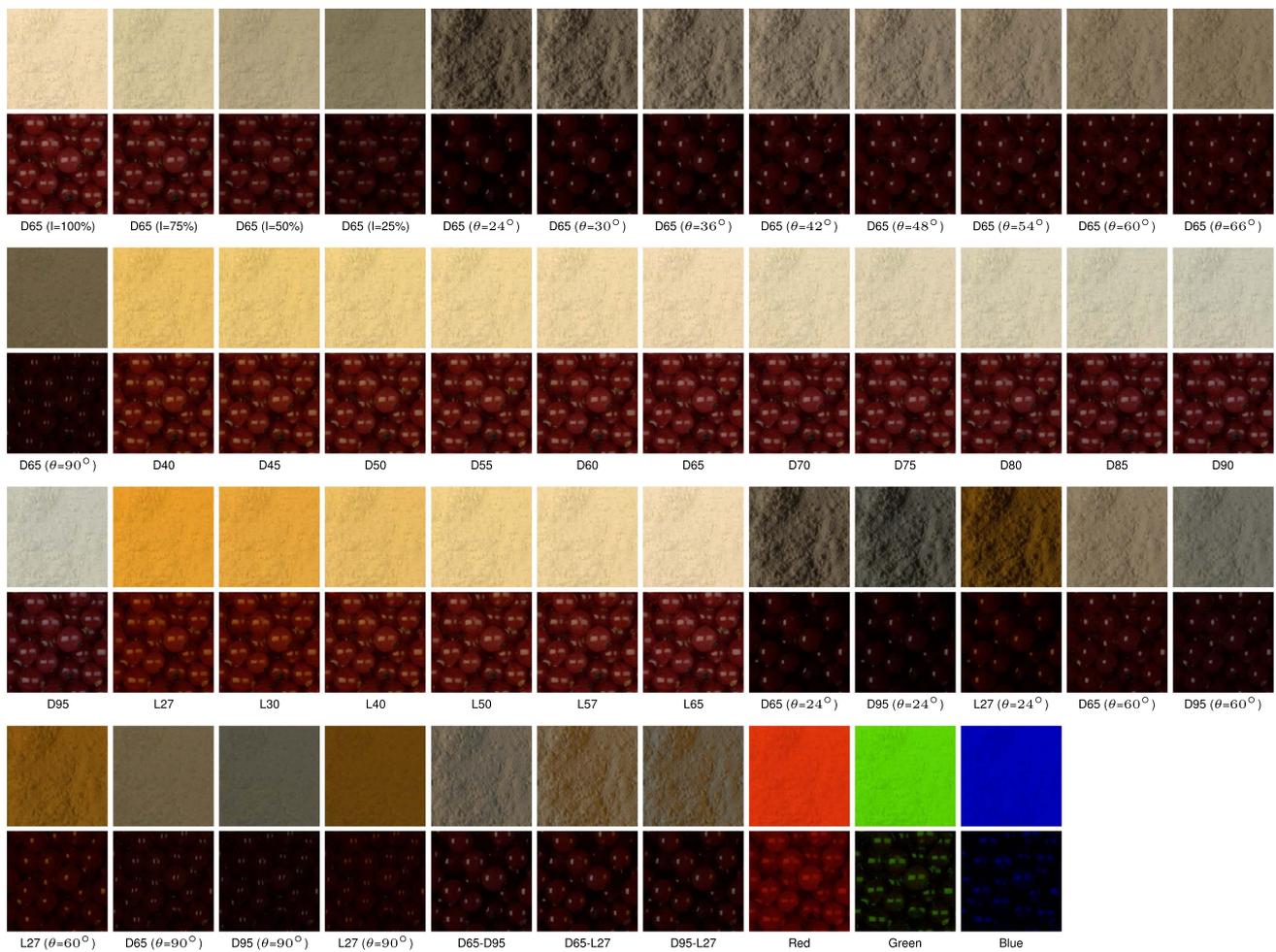


Fig. 4. Overview of the 46 lighting conditions in the Raw Food Texture database. The top rows represent the flour class, while bottom rows represent the currant class.

refer to the work of Drbohlav and Leonardis [39] for a comparative analysis of texture methods under varying viewpoints and illumination and to the work of Kandaswamy *et al.* [40] for a comparison among texture analysis schemes under nonideal conditions.

In this work, we compared several descriptors from the state of the art, by taking a few representative descriptors for each of the previously mentioned approaches. Several descriptors have been applied to color and gray-scale images, where the gray-scale image is defined as the luminance of the image and is obtained by using the standard formula: $L = 0.299R + 0.587G + 0.114B$.

In order to make the results readable, we consider, here, only a selection of all the descriptors evaluated.

A. Traditional Descriptors

- 256-dimensional gray-scale histogram
- 512-dimensional hue and value marginal histogram obtained from the HSV color representation of the image
- 768-dimensional RGB and rgb marginal histograms [41]
- 10-dimensional feature vector composed of normalized chromaticity moments, as defined in [42]

- 15-dimensional feature vector composed of contrast, correlation, energy, entropy, and homogeneity extracted from the co-occurrence matrices of each color channel [43,44]

- 144-dimensional Gabor features composed of mean and standard deviation of six orientations extracted at four frequencies for each color channel [30,3]

- 264-dimensional opponent Gabor feature vector extracted as Gabor features from several inter/intra channel combinations: monochrome features extracted from each channel separately and opponent features extracted from a couple of colors at different frequencies [45]

- 54-dimensional dual tree complex wavelet transform (DT-CWT) features obtained considering four scales, mean and standard deviation, and three color channels [3,46]

- 26-dimensional feature vector obtained calculating morphological operators (granulometries) at four angles and for each color channel [47]

- 512-dimensional Gist features obtained considering eight orientations and four scales for each channel [48]

- 81-dimensional histogram of oriented gradients (HOG) feature vector [49]. Nine histograms with nine bins are concatenated to achieve the final feature vector

- 243-dimensional local binary patterns (LBP) feature vector computed with 16 neighbors, radius 2, and uniform patterns. We applied LBP to gray-scale images and then to the color channels RGB, CIE-Lab, and Ohta's $I_1I_2I_3$ spaces (in these cases, the vector will be 729-dimensional) [2]

- Combination of LBP computed on pairs of color channels, namely, the opponent color LBP (OCLBP) [50]

- LBP combined with the local color contrast descriptor, as described in [38]

- 499-dimensional local color contrast feature vector. It is obtained by concatenating the LBP on the gray images with a quantized measure of color contrast [38]

B. Descriptors for Object Recognition

The features considered here consist in the aggregation of local descriptors according to the quantization defined by a codebook of visual words. As local descriptors, we used 128-dimensional dense SIFT obtained from the gray-scale image by considering a spatial histogram of local gradient orientations. The spatial bins have an extent of 6×6 . The descriptors have been sampled every two pixels and at scales $2^{i/3}$, $i = 0, 1, 2, \dots$

The object recognition features differ for the aggregation method, but all of them are based on a codebook of 1024 visual words built on images from external sources. In particular, we downloaded 20,000 images from Flickr containing various content, such as sunset, countryside, etc., and we used k means to find 1024 representative vectors.

The object recognition features considered here are as follows:

- 1024-dimensional bag of visual words (BoVW)
- 25600-dimensional vector of locally aggregated descriptors (vlad) [8]
- 40960-dimensional Fisher's vectors (fv) of locally aggregated descriptors [51]

C. CNN-Based Descriptors

The CNN-based features have been obtained as the intermediate representations of deep convolutional neural networks originally trained for object recognition. The networks are used to generate a texture descriptor by removing the final softmax nonlinearity and the last fully connected layer, resulting in feature vectors that are L^2 normalized before being used for classification. We considered the most representative CNN architectures in the state of the art [52], each exploring a different accuracy/speed trade-off. All the CNNs have been trained on the ILSVRC-2012 data set using the same protocol as in [53]. In particular, we considered 4096-, 2048-, 1024-, and 128-dimensional feature vectors as follows [34].

- BVLC AlexNet (BVLC AlexNet): AlexNet trained on ILSVRC 2012 [53]

- BVLC Reference CaffeNet (BVLC Ref): AlexNet trained on ILSVRC 2012, with a minor variation from the version as described in [53].

- Fast CNN (Vgg F): This is similar to the one presented in [53] with a reduced number of convolutional layers and the dense connectivity between convolutional layers. The last fully connected layer is 4096-dimensional [54].

- Medium CNN (Vgg M): This is similar to the one presented in [55] with a reduced number of filters in the

convolutional layer four. The last fully connected layer is 4096-dimensional [54].

- Medium CNN (Vgg M-2048-1024-128): Three modifications of the Vgg M network, with lower dimensional last fully connected layer. In particular, we used a feature vector of 2048, 1024, and 128 size [54].

- Slow CNN (Vgg S): This is similar to the one presented in [56] with a reduced number of convolutional layers, less filters in the layer five and the local response normalization. The last fully connected layer is 4096-dimensional [54].

- Vgg Very Deep 19 and 16 layers (Vgg Very Deep 16 and 19): the configuration of these networks has been achieved by increasing the depth to 16 and 19 layers, which results in a substantially deeper network than what has been used in previous studies [57].

Under the premise that the implementations we used are not always optimal, we observed that the computation of most of the descriptors considered here requires from 0.02 to 0.5 s per image on a PC equipped with an Intel Core i5-2500K CPU. The only exceptions are color histograms that are quicker to compute (0.002–0.008 s per image), and granulometry and co-occurrence matrices, which are slower (about 0.9 and 3.5 s per image, respectively).

D. Color Normalization

Invariance with respect to specific changes in acquisition conditions, such as those caused by variations in the illumination, is an important property of visual descriptors. Illumination variations also can be compensated by preprocessing images with a color normalization method. Color normalization methods try to assign a constant color to objects acquired under different illumination conditions.

In order to evaluate this strategy, we have preprocessed the RawFoot database by using several existing normalization methods; next, we have extracted features by using the best color descriptors from the set of descriptors evaluated in Table 2. More precisely, we considered two implementations of the Retinex method described in [58], which improve the computational efficiency while preserving the underlying principles: the McCann99 [59] and the Frankle–McCann [60]. Furthermore, we considered the gray world [61], two variants of edge-based algorithm, the gray edge [62], and the weighted gray-edge method [63].

4. EXPERIMENTS

In all the experiments, we used the nearest neighbor classification strategy: given a patch in the test set, its distance with respect to all the training patches is computed. The prediction of the classifier is the class of the closest element in the training set. For this purpose, after some preliminary tests with several descriptors in which we evaluated the most common distance measures, we decided to use the $L1$ distance: $d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^N |\mathbf{x}_i - \mathbf{y}_i|$, where \mathbf{x} and \mathbf{y} are two feature vectors. All the experiments have been conducted under the maximum ignorance assumption; that is, no information about the lighting conditions of the test patches is available for the classification method and for the descriptors. Performance is reported as classification rate (i.e., the ratio between the number of correctly classified

Table 2. Classification Rates (%) of the Texture Descriptors Considered^a

Features	No Variations		Light Intensity		Light Direction		Daylight Temperature		LED Temperature		Daylight Versus LED		Temp. OR Direction		Temp. and DIRECTION		Multi-Illum	
	avg (min)	avg (min)	avg (min)	avg (min)	avg (min)	avg (min)	avg (min)	avg (min)	avg (min)	avg (min)								
Hist. L	78.32 (60.66)	6.77 (1.47)	20.97 (2.21)	49.94 (11.95)	27.18 (5.88)	38.05 (6.43)	10.45 (1.29)	7.98 (1.29)	47.33 (30.15)									
Hist. H V	96.38 (84.56)	31.45 (14.52)	51.01 (13.97)	49.11 (9.93)	51.56 (23.35)	44.39 (9.19)	16.47 (4.23)	11.36 (4.23)	49.82 (38.42)									
Hist. RGB	94.93 (87.13)	15.89 (3.12)	40.82 (5.70)	56.45 (18.20)	37.51 (12.68)	43.44 (8.00)	15.53 (2.76)	10.21 (2.76)	53.83 (42.28)									
Hist. <i>rgb</i>	97.24 (92.46)	67.08 (36.95)	64.07 (24.63)	37.35 (6.43)	17.38 (3.31)	25.71 (5.15)	20.16 (2.39)	12.57 (2.39)	45.25 (16.36)									
Chrom. mom.	82.54 (58.46)	68.43 (48.90)	53.68 (22.24)	33.41 (4.96)	18.66 (3.68)	24.16 (5.06)	17.03 (2.21)	10.63 (2.21)	33.92 (20.77)									
Co-occ. matr.	35.33 (9.93)	7.20 (2.02)	9.39 (0.55)	23.02 (9.74)	19.01 (6.62)	19.88 (5.61)	3.30 (0.18)	2.89 (0.18)	10.48 (6.07)									
Co-occ. matr. L	18.68 (1.47)	3.32 (0.00)	5.63 (0.55)	16.99 (6.99)	9.49 (3.31)	12.94 (2.85)	2.49 (0.00)	2.57 (0.00)	6.86 (2.57)									
DT-CWT	92.26 (81.62)	21.68 (1.65)	49.48 (12.32)	66.29 (25.92)	42.31 (14.34)	49.77 (15.44)	19.23 (3.12)	13.22 (3.12)	65.93 (55.33)									
DT-CWT L	72.85 (58.09)	10.65 (1.29)	29.29 (5.15)	60.13 (27.39)	32.70 (4.04)	44.06 (5.06)	14.70 (1.47)	12.92 (1.47)	52.73 (37.13)									
Gabor RGB	93.02 (61.76)	66.96 (32.35)	51.66 (11.95)	64.81 (20.77)	38.13 (12.13)	48.03 (12.59)	27.18 (3.49)	16.43 (3.49)	75.18 (59.93)									
Gabor L	72.91 (70.04)	46.57 (18.75)	29.34 (3.86)	68.94 (59.56)	67.62 (58.82)	66.86 (53.40)	27.58 (2.57)	16.09 (2.57)	74.33 (72.79)									
Opp. Gabor RGB	96.15 (59.38)	21.51 (3.49)	50.61 (12.13)	67.75 (22.98)	41.78 (14.34)	50.80 (15.07)	20.22 (3.86)	13.47 (3.86)	58.58 (43.01)									
Gist RGB	66.20 (62.50)	55.06 (31.99)	45.75 (10.85)	55.49 (28.31)	36.78 (13.24)	43.41 (13.79)	25.13 (2.76)	17.30 (2.76)	60.36 (54.78)									
Granulometry	91.98 (51.65)	63.73 (27.76)	48.65 (10.66)	69.80 (21.51)	33.58 (6.80)	48.79 (6.34)	22.20 (1.65)	12.86 (1.65)	74.23 (64.34)									
HoG	46.74 (43.20)	37.52 (24.82)	27.48 (8.46)	41.14 (29.60)	35.29 (22.24)	36.30 (19.30)	16.99 (3.49)	11.71 (3.49)	43.66 (40.62)									
LBP L	80.37 (77.02)	51.15 (17.83)	47.09 (9.19)	77.76 (72.24)	70.77 (54.60)	73.15 (55.06)	29.54 (5.51)	18.66 (5.51)	76.99 (74.26)									
LBP RGB	93.55 (90.81)	68.87 (33.46)	59.48 (13.60)	72.40 (24.63)	48.39 (15.07)	56.08 (16.82)	23.72 (0.55)	14.19 (0.55)	76.81 (67.10)									
LBP Lab	92.90 (88.42)	71.88 (32.54)	61.74 (17.28)	70.61 (24.08)	51.53 (21.69)	56.00 (19.49)	27.55 (3.31)	18.06 (3.31)	77.21 (71.32)									
LBP $I_1I_2I_3$	91.40 (82.90)	66.28 (28.12)	63.05 (17.10)	70.58 (25.92)	49.90 (18.38)	54.76 (17.00)	27.05 (1.10)	18.39 (1.10)	75.55 (64.89)									
OCLBP	95.92 (92.28)	78.75 (51.47)	65.70 (14.52)	67.92 (19.67)	49.94 (15.81)	53.93 (15.81)	25.73 (1.65)	16.99 (1.65)	76.53 (50.00)									
LCC	92.92 (88.60)	62.64 (26.84)	56.15 (12.13)	88.78 (73.71)	74.25 (46.88)	78.82 (50.64)	31.13 (5.15)	19.85 (5.15)	85.91 (84.38)									
BoVW	89.73 (86.58)	87.38 (81.43)	67.38 (16.18)	90.02 (88.05)	88.87 (86.76)	89.53 (87.68)	51.59 (12.68)	39.34 (12.68)	88.60 (87.68)									
VLAD	79.29 (75.18)	76.87 (70.22)	64.73 (20.77)	78.70 (76.65)	77.87 (73.90)	78.31 (75.74)	51.44 (18.20)	42.62 (18.75)	80.58 (79.78)									
FV	85.59 (80.51)	81.02 (71.32)	69.31 (27.39)	86.57 (84.38)	84.58 (78.68)	85.51 (82.35)	55.26 (20.04)	45.95 (20.04)	85.69 (84.56)									
Vgg F	96.94 (89.34)	88.07 (72.61)	83.23 (47.06)	96.27 (86.95)	87.94 (65.26)	89.40 (62.96)	54.89 (13.42)	45.62 (13.42)	95.59 (94.49)									
Vgg M	97.50 (89.89)	89.94 (76.29)	86.14 (52.02)	97.30 (88.97)	90.64 (69.85)	91.91 (71.60)	60.25 (20.04)	51.13 (20.04)	96.23 (94.49)									
Vgg M S	97.53 (91.36)	90.93 (74.82)	86.24 (51.10)	96.16 (85.66)	89.90 (68.75)	90.33 (69.21)	59.02 (16.73)	49.48 (16.73)	95.53 (93.57)									
Vgg M 2048	97.27 (89.89)	89.08 (72.43)	84.74 (48.35)	97.38 (90.81)	90.06 (64.34)	91.41 (69.58)	57.18 (14.15)	47.89 (14.15)	95.01 (93.38)									
Vgg M 1024	96.79 (87.87)	88.53 (69.67)	84.15 (49.08)	96.90 (90.07)	90.14 (63.42)	91.47 (69.76)	56.26 (14.15)	47.26 (14.15)	93.75 (91.91)									
Vgg M 128	93.50 (78.31)	81.54 (59.93)	78.04 (45.04)	93.01 (82.35)	79.72 (46.69)	83.20 (50.92)	49.13 (13.79)	40.75 (13.79)	89.00 (87.50)									
BVLC Ref	95.22 (84.56)	83.23 (60.66)	79.00 (40.81)	94.51 (80.33)	85.11 (58.64)	86.52 (58.00)	49.50 (9.38)	40.63 (9.38)	91.24 (89.15)									
BVLC AlexNet	94.88 (82.17)	85.00 (65.07)	80.25 (43.57)	95.79 (86.03)	88.56 (66.36)	89.64 (65.99)	48.08 (9.38)	39.37 (9.38)	90.44 (88.24)									
Vgg	98.21 (92.83)	94.10 (85.48)	91.23 (66.73)	97.41 (86.21)	93.69 (77.57)	93.67 (75.83)	70.81 (33.82)	63.64 (33.82)	96.60 (94.85)									
Very Deep 16	97.69 (92.10)	93.01 (82.90)	90.32 (62.50)	96.98 (85.85)	93.94 (79.23)	93.53 (74.82)	69.64 (37.68)	62.35 (37.68)	95.59 (94.49)									
Vgg																		
Very Deep 19																		

^aFor each classification task, the best result is reported in bold.

images and the number of test images). Note that more complex classification schemes (e.g., SVMs) would have been viable. We decided to adopt the simplest one in order to focus the evaluation on the features themselves and not on the classifier.

A. RawFoot Database Setup

For each of the 68 classes, we considered 16 patches obtained by dividing the original texture image, which is 800×800 pixels, in 16 nonoverlapping squares of size 200×200 pixels. For each class, we selected eight patches for training and eight for testing alternating themes in a checkerboard pattern. We formed subsets of $68 \times (8 + 8) = 1088$ patches by taking the training and test patches from images taken under different lighting conditions.

In this way, we defined several subsets, grouped in nine texture classification tasks.

1. **No variations:** 46 subsets. Each subset is composed of training and test patches taken under the same lighting condition.
2. **Light intensity:** 12 subsets obtained by combining the four intensity variations. Each subset is composed of training and test patches with different light intensity values.
3. **Light direction:** 72 subsets obtained by combining the nine different light directions. Each subset is composed of training and test patches with different light direction.
4. **Daylight temperature:** 132 subsets obtained by combining all the 12 daylight temperature variations. Each subset is composed of training and test patches with different light temperatures.

5. **LED temperature:** 30 subsets obtained by combining all the six LED temperature variations. Each subset is composed of training and test patches with different light temperatures.

6. **Daylight versus LED:** 72 subsets obtained by combining 12 daylight temperatures with six LED temperatures.

7. **Temperature or direction:** 72 subsets obtained by combining all the nine combinations of color temperatures and light directions. Each subset is composed of training and test patches where either the color or the direction (or both) change.

8. **Temperature and direction:** 36 subsets obtained by combining all the nine combinations of color temperatures and light directions. Each subset is composed of training and test patches where both the color and the direction change.

9. **Multiple illuminant:** Six subsets obtained by combining the three acquisitions with multiple illuminants.

5. RESULTS

Table 2 reports the performance obtained by the descriptors considered as average and minimum accuracy over the nine classification tasks. For the four main tasks (same illuminant, light intensity, light direction, and daylight temperature), the results are shown in greater detail in Fig. 5 but only for some representative methods. When training and tests are taken under the same lighting conditions, the classification rates are generally high, regardless of the specific conditions. CNN features perform very well, with a peak of 98.2% of accuracy obtained with the features extracted by the Vgg Very Deep 16 network. Other, more traditional features perform

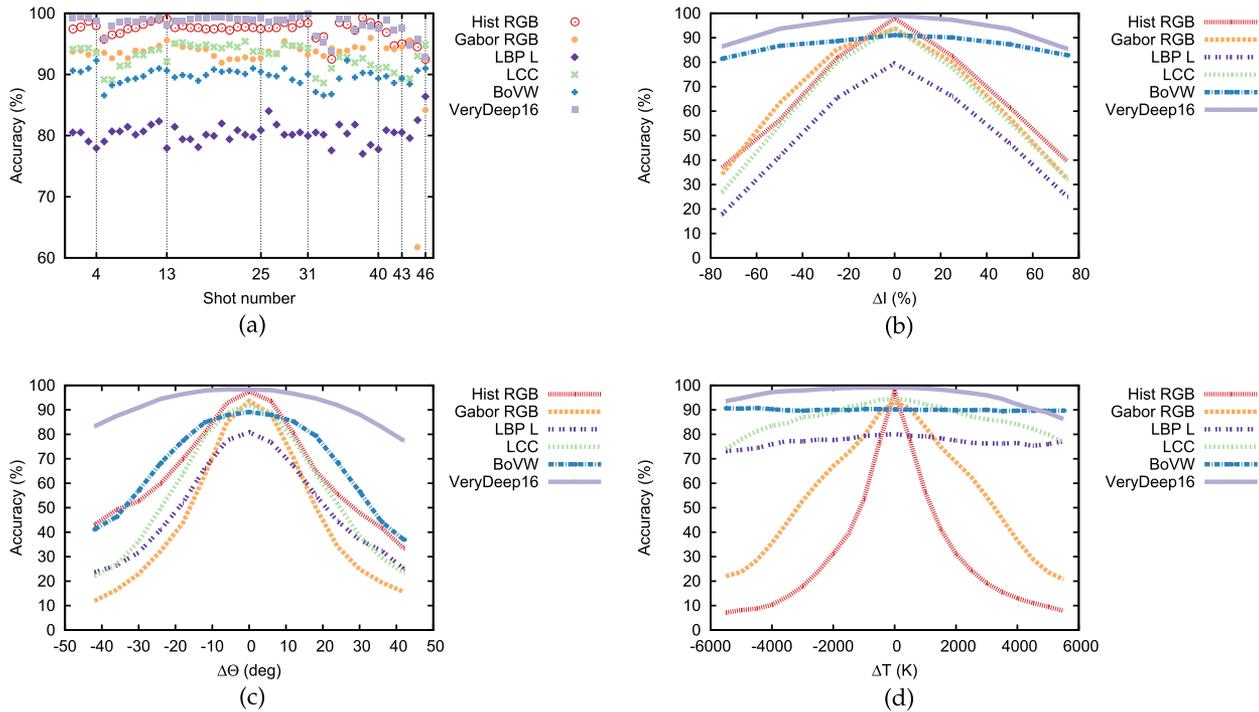


Fig. 5. Detail of the classification rates as functions of the amount of variability in the illumination conditions between the training and test set. (a) Accuracy obtained in the no variations classification task (each point corresponds to one of the 46 shots). (b) Accuracy with respect to the difference ΔI of light intensity. (c) Accuracy obtained varying the difference between the directions of the light. (d) Accuracy with respect to the difference ΔT of daylight temperature.



Fig. 6. Accuracy of the Vgg Very Deep 16 (turquoise) and BoVW (black) over the 68 classes. (a) Training and test images are under different lights and same angle. (b) Training and test images are under the same light and different angles. To map the numbers to the corresponding classes, see Fig. 2.

very well in this scenario (OCLBP at 95.9% Opp. Gabor at 96.2%), and even simple *rgb* histograms achieve an accuracy of 97.2%. It is clear that, under fixed conditions, texture classification is not a very challenging problem [see also Fig. 5(a)].

When training and test patches are taken under variable intensity, the behavior of CNN features and of the descriptors taken from the object recognition literature (BoVW) is stable. Surprisingly, traditional handcrafted features are heavily affected by these kinds of variations, even when they are supposed to be robust to them, as should be the case of LBP and Gabor-based features. This behavior is more evident looking at Fig. 5(b), where only Vgg Very Deep 16 and BoVW have flat curves over changes in the intensity of the light.

Perhaps one of the most challenging variations to take into account is that related to the direction of the light [see Fig. 5(c)]. In this task, all the descriptors suffered a noticeable decrease in performance. However, some CNN features remained, on average, above 90% of accuracy. The performance of all the other features dropped below 70%.

When the illuminant color is allowed to vary between train and test images, the achromatic features are the least affected. In particular, the features from the object recognition literature obtained the same performance of the same illumination task. For other features, such as LBP-L, we observed a decrease in the performance, probably due to the variation in intensity caused by the change of the color temperature. Features that use color information greatly suffer this kind of variability [see Fig. 5(d)]. The most important exception is represented by the CNN features, which have been trained to exploit color, but in such a way to be robust with respect to the large amount of variability in the object categories they were supposed to discriminate.

Very low performance has been obtained when both direction and color change simultaneously. In this case, the best results have been obtained by, again, features from CNNs. However, the highest classification rate is quite low (about 63.6%), and most networks do not allow us to achieve more than 50%. The results for the other features are even worse than that. The last task involved the presence of multiple illuminants. Because their position was stable, we obtained similar results to those of the case of variable color temperature.

Summing up, the challenges of recognizing textures under variable illumination conditions greatly depend on the type of variability involved in the experiments. Features extracted by CNNs significantly outperform the other descriptors considered. Features from the object recognition literature clearly outperform traditional handcrafted texture features in all the scenarios considered. Only under specific circumstances, these last ones outperformed CNN features. For instance, in Fig. 5(d) it can be observed that CNN features fall below the bag of visual word descriptors for extreme variations in color temperature. These circumstances can be better understood by looking at Fig. 6(a), which compares the behavior of Vgg Very Deep 16 and BoVW over the 68 classes. Here, the training and test images have been taken under different lights but all are directed with the same angle. In particular, we averaged the accuracy obtained in three sets of experiments, one for each angle (24°, 60°, and 90°) all including three lights: D65, D95, and L27. It is quite evident that, in this case, object recognition features outperform CNNs, especially for those classes whose appearance is most sensitive to color changes and that contain a more fine-grained texture, such as (3) salt, (12) sliced bread, (36) flour, (53) mango, (61) coconut flakes, (66)

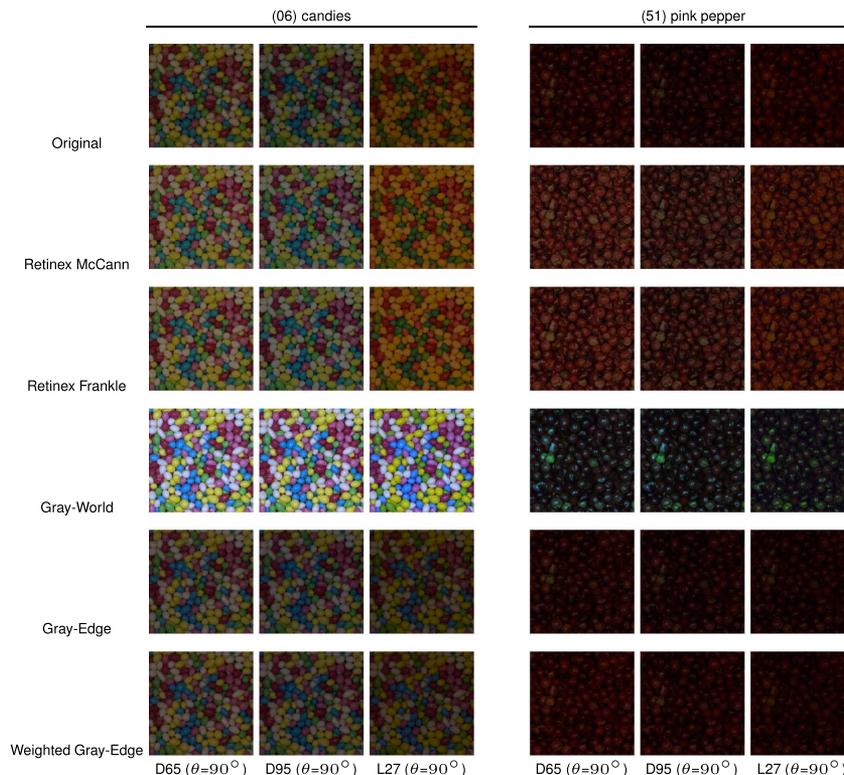


Fig. 7. Color normalization methods applied to two different samples acquired under three different lights.

sugar, etc. This result is due to the fact the CNNs mainly have been trained on images of objects and, thus, contain more coarse details. In contrast, when the training and test images have been taken under the same light but with different light directions, CNN features demonstrate to be more robust than object recognition features [see Fig. 6(b)]. Here, the worst results are obtained by BoVW on coarse-grained texture images, such as (1) chickpeas, (20) basmati rice, (62) chicory, etc.

A. Preprocessing with Color Normalization

We have preprocessed all the images with five state-of-the-art color normalization methods. Examples of preprocessing applied to two different samples are represented in Fig. 7.

Table 3 reports the performance obtained by these color normalization methods combined with a selection of descriptors. It is clear that color normalization helps us to improve performance in the case of CNNs. In particular, the combination of Vgg Very Deep 16 with Retinex Frackle achieves an improvement of 5% in both the cases of temperature and/or direction variations. This result confirms the fact that CNNs have been trained on images without considering changes in illumination conditions. In contrast, the combination of handcrafted features with preprocessing methods, in most of the cases, does not bring any improvements in terms of classification rate. This is due to the fact that those features, except for color histograms, have been designed to be more robust to changes in the temperature of the light.

6. SUMMARY

In order to obtain reliable classification of color textures under uncontrolled conditions, we believe that the descriptor's performance should be assessed under a large set of carefully controlled variations of lighting conditions. We described RawFooT, which is a database of texture images acquired under variable light direction, color, and intensity. The images of the database will be made publicly available together with all the scripts used in our experimentation. We also will disclose the detailed technical specifications of the hardware and software used to acquire the database; this will allow researchers in this area to extend RawFooT or to acquire their own database.

RawFooT allowed us to conduct a variety of experiments in which traditional texture, object recognition, and CNN-based descriptors have been evaluated in terms of their capabilities in dealing with single and combined variations in the lighting conditions. These experiments made very clear the strengths and the weaknesses of the investigated approach and clearly outlined open issues that should be addressed to actually design color texture descriptors robust with respect to unknown variations in the imaging conditions.

In extreme summary, we can conclude the following:

- Traditional texture descriptors are effective only when images have no variations in lighting conditions.
- Object recognition descriptors demonstrated to perform, in most of the cases, better than the traditional ones.
- CNN-based descriptors confirmed to be powerful also on texture classification tasks outperforming the handcrafted traditional and object-oriented features.

- CNN-based descriptors handle most of the variations in lighting conditions. A possible explanation of their good performance is that they capture the spatial layout of the patterns in the texture images. However, for large variations in both color and direction of the light, CNN-based descriptors have demonstrated to be less effective than object recognition descriptors, especially on these classes that are more fine-grained.

- The use of color normalization did not improve any of the handcrafted descriptors, while, for CNN-based descriptors, they demonstrated to be helpful in dealing with complex variations in illumination conditions.

Acknowledgment. We would like to thank the associate editor and the reviewers for their valuable comments and suggestions that allowed us to improve the paper.

REFERENCES

1. A. Drimbarean and P. Whelan, "Experiments in colour texture analysis," *Pattern Recogn. Lett.* **22**, 1161–1167 (2001).
2. T. Mäenpää and M. Pietikäinen, "Classification with color and texture: jointly or separately?" *Pattern Recogn.* **37**, 1629–1640 (2004).
3. F. Bianconi, R. Harvey, P. Southam, and A. Fernández, "Theoretical and experimental comparison of different approaches for color texture classification," *J. Electron. Imaging* **20**, 043006 (2011).
4. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**, 436–444 (2015).
5. M. Cimpoi, S. Maji, and A. Vedaldi, "Deep filter banks for texture recognition and segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 3828–3836.
6. S. Hossain and S. Serikawa, "Texture databases-a comprehensive survey," *Pattern Recogn. Lett.* **34**, 2007–2022 (2013).
7. F. Bianconi and A. Fernández, "An appendix to "texture databases-a comprehensive survey,"" *Pattern Recogn. Lett.* **45**, 33–38 (2014).
8. M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi, "Describing textures in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2014), pp. 3606–3613.
9. "RawFooT DB: Raw Food Texture Database," <http://projects.ivl.disco.unimib.it/rawfoot/> (2015).
10. S. Bell, P. Upchurch, N. Snaveley, and K. Bala, "Material recognition in the wild with the materials in context database," arXiv:1412.0623 (2014).
11. L. Sharan, C. Liu, R. Rosenholtz, and E. H. Adelson, "Recognizing materials using perceptually inspired features," *Int. J. Comput. Vis.* **103**, 348–371 (2013).
12. P. Brodatz, *Textures: a Photographic Album for Artists and Designers* (Dover, 1966), Vol. **66**.
13. K. Dana, B. Van-Ginneken, S. Nayar, and J. Koenderink, "Reflectance and texture of real world surfaces," *ACM Trans. Graph.* **18**, 1–34 (1999).
14. Media Laboratory at Heriot-Watt University, "Vistex database," <http://vismod.media.mit.edu/vismod/imagery/VisionTexture>.
15. G. Smith, "Meastex image texture database and test suite," <http://www.texturesynthesis.com/meastex/meastex.html>.
16. Texturelab Edinburgh at Massachusetts Institute of Technology, "Photex database," <http://www.macs.hw.ac.uk/texturelab/resources/databases/photex/>.
17. T. Ojala, T. Maenpää, M. Pietikäinen, J. Viertola, J. Kyllönen, and S. Huovinen, "Outex-new framework for empirical evaluation of texture analysis algorithms," in *Proceedings of 16th International Conference on Pattern Recognition (IEEE, 2002)*, Vol. **1**, pp. 701–706.
18. R. Péteri, S. Fazekas, and M. J. Huiskes, "Dyntex: A comprehensive database of dynamic textures," *Pattern Recogn. Lett.* **31**, 1627–1632 (2010).

19. S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans. Pattern Anal. Mach. Intell.* **27**, 1265–1278 (2005).
20. B. Caputo, E. Hayman, and P. Mallikarjuna, "Class-specific material categorization," in *10th IEEE International Conference on Computer Vision* (IEEE, 2005), Vol. 2, pp. 1597–1604.
21. G. J. Burghouts and J.-M. Geusebroek, "Material-specific adaptation of color invariant features," *Pattern Recogn. Lett.* **30**, 306–313 (2009).
22. A. Fernández, O. Ghita, E. González, F. Bianconi, and P. F. Whelan, "Evaluation of robustness against rotation of LBP, CCR and ILBP features in granite texture classification," *Mach. Vis. Appl.* **22**, 913–926 (2011).
23. R. Kwitt and P. Meerwald, "Stex: Salzburg texture image database," <http://www.wavelab.at/sources/STex/>.
24. G. Kylberg, "Automatic virus identification using TEM: image segmentation and texture analysis," Ph.D. thesis (Acta Universitatis Upsaliensis, 2014).
25. A. R. Backes, D. Casanova, and O. M. Bruno, "Color texture analysis based on fractal descriptors," *Pattern Recogn.* **45**, 1984–1992 (2012).
26. A. Poirson and B. Wandell, "Pattern-color separable pathways predict sensitivity to simple colored patterns," *Vis. Res.* **36**, 515–526 (1996).
27. R. Khan, J. Van De Weijer, D. Karatzas, and D. Muselet, "Towards multispectral data acquisition with hand-held devices," in *International Conference on Image Processing* (2013), pp. 2053–2057.
28. G. Wyszecki and W. S. Stiles, *Color Science* (Wiley, 1982), Vol. 8.
29. M. Mirmehdi, X. Xie, and J. Suri, *Handbook of Texture Analysis* (Imperial College, 2009).
30. F. Bianconi and A. Fernández, "Evaluation of the effects of Gabor filter parameters on texture classification," *Pattern Recogn.* **40**, 3325–3335 (2007).
31. R. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE* **67**, 786–804 (1979).
32. T. Ojala, M. Pietikäinen, and T. Mänekää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 971–987 (2002).
33. G. Sharma, S. ul Hussain, and F. Jurie, "Local Higher-Order Statistics (lhs) for Texture Categorization and Facial Analysis," in *Lecture Notes in Computer Science* (Springer, 2012), Vol. 7578, pp. 1–12.
34. A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2014), pp. 512–519.
35. R. Khan, D. Muselet, and A. Trémeau, "Classical texture features and illumination color variations," in *Proceedings of third International Conference on Machine Vision* (2010), pp. 280–285.
36. G. Finlayson, S. Hordley, G. Schaefer, and G. Yun Tian, "Illuminant and device invariant colour using histogram equalisation," *Pattern Recogn.* **38**, 179–190 (2005).
37. M. Seifi, X. Song, D. Muselet, and A. Trémeau, "Color texture classification across illumination changes," in *Conference on Colour in Graphics, Imaging, and Vision* (2010), pp. 332–337.
38. C. Cusano, P. Napolitano, and R. Schettini, "Combining local binary patterns and local color contrast for texture classification under varying illumination," *J. Opt. Soc. Am. A* **31**, 1453–1461 (2014).
39. O. Drbohlav and A. Leonardis, "Towards correct and informative evaluation methodology for texture classification under varying viewpoint and illumination," *Comput. Vis. Image Underst.* **114**, 439–449 (2010).
40. U. Kandaswamy, S. A. Schuckers, and D. Adjeroh, "Comparison of texture analysis schemes under nonideal conditions," *IEEE Trans. Image Process.* **20**, 2260–2275 (2011).
41. M. Pietikäinen, S. Nieminen, E. Marszałec, and T. Ojala, "Accurate color discrimination with classification based on feature distributions," in *Proceedings of the 13th International Conference on Pattern Recognition* (1996), Vol. 3, pp. 833–838.
42. G. Paschos, "Fast color texture recognition using chromaticity moments," *Pattern Recogn. Lett.* **21**, 837–841 (2000).
43. V. Arvis, C. Debain, M. Berducat, and A. Benassi, "Generalization of the co-occurrence matrix for colour images: application to colour texture," *Image Anal. Stereol.* **23**, 63–72 (2004).
44. M. Hauta-Kasari, J. Parkkinen, T. Jaaskelainen, and R. Lenz, "Generalized co-occurrence matrix for multispectral texture analysis," in *Proceedings of the 13th International Conference on Pattern Recognition* (1996), Vol. 2, pp. 785–789.
45. A. Jain and G. Healey, "A multiscale representation including opponent color features for texture recognition," *IEEE Trans. Image Process.* **7**, 124–128 (1998).
46. M. Barilla and M. Spann, "Colour-based texture image classification using the complex wavelet transform," in *5th International Conference on Electrical Engineering, Computing Science and Automatic Control* (2008), pp. 358–363.
47. A. Hanbury, U. Kandaswamy, and D. Adjeroh, "Illumination-invariant morphological texture classification," in *Mathematical Morphology: 40 Years On*, C. Ronse, L. Najman, and E. Decencire, eds., Vol. 30 of *Computational Imaging and Vision* (Springer, 2005), pp. 377–386.
48. A. Oliva and A. Torralba, "Modeling the shape of the scene: a holistic representation of the spatial envelope," *Int. J. Comput. Vis.* **42**, 145–175 (2001).
49. O. L. Junior, D. Delgado, V. Gonçalves, and U. Nunes, "Trainable classifier-fusion schemes: an application to pedestrian detection," in *Intelligent Transportation Systems* (2009).
50. C.-H. Chan, J. Kittler, and K. Messer, "Multispectral local binary pattern histogram for component-based color face verification," in *1st IEEE International Conference on Biometrics: Theory, Applications, and Systems* (2007), pp. 1–7.
51. H. Jégou, M. Douze, C. Schmid, and P. Pérez, "Aggregating local descriptors into a compact image representation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2010), pp. 3304–3311.
52. A. Vedaldi and K. Lenc, "MatConvNet—convolutional neural networks for MATLAB," arXiv:1412.4564 (2014).
53. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (2012), pp. 1097–1105.
54. K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: delving deep into convolutional nets," arXiv:1405.3531 (2014).
55. M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision—ECCV 2014* (Springer, 2014), Vol. 8689, pp. 818–833.
56. P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," arXiv:1312.6229 (2013).
57. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv:1409.1556 (2014).
58. B. Funt, F. Ciurea, and J. McCann, "Retinex in MATLAB," *J. Electron. Imaging* **13**, 112–121 (2000).
59. J. McCann, "Lessons learned from Mondrians applied to real images and color gamuts," in *Color and Imaging Conference* (1999), pp. 1–8.
60. J. Frankle and J. McCann, "Method and apparatus for lightness imaging," U.S. patent 4384336 (May 17, 1983).
61. G. Finlayson and E. Trezzi, "Shades of gray and colour constancy," in *Color and Imaging Conference* (2004), pp. 37–41.
62. J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Trans. Image Process.* **16**, 2207–2214 (2007).
63. A. Gijsenij, T. Gevers, and J. van de Weijer, "Improving color constancy by photometric edge weighting," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**, 918–929 (2012).