Beyond Eleven Color Names for Image Understanding

Lu Yu · Lichao Zhang · Joost van de Weijer · Fahad Shahbaz Khan · Yongmei Cheng · Alejandro Parraga

Abstract Color description is one of the fundamental problems of image understanding. One of the popular ways to represent colors is by means of color names. Most existing work on color names focuses on only the eleven basic color terms of the English language. This could be limiting the discriminative power of these representations, and representations based on more color names are expected to perform better. However, there exists no clear strategy to choose additional color names.

We collect a dataset of 28 additional color names. To ensure that the resulting color representation has high discriminative power we propose a method to order the additional color names according to their complementary nature with the basic color names. This allows us to compute color name representations with high discriminative power of arbitrary length. In the experiments we show that these new color name descriptors outperform the existing color name descriptor on the task of visual tracking, person re-identification and image classification.

Keywords Color name · discriminative descriptors · image classification · re-identification · tracking

Lu Yu
Key Laboratory of Information Fusion Technology, Northwestern Polytechnical University, xi'an, 710072, China
Tel.: +3-464-416-1086
Fax: +0-000-000-0000
E-mail: luyu@cvc.uab.es
Present address:Computer Vision Center Barcelona, Universitat Autonoma de Barcelona, Barcelona, 08193, Spain

Lichao Zhang
Computer Vision Center Barcelona, Universitat Autonoma de Barcelona, Barcelona, 08193, Spain

Joost van de Weijer
Computer Vision Center Barcelona, Universitat Autonoma de Barcelona, Barcelona, 08193, Spain

Fahad Shahbaz Khan
Computer Vision Laboratory, Linköping University, LinköpingSE-581 83, Sweden

Yongmei Cheng
Key Laboratory of Information Fusion Technology, Northwestern Polytechnical University, xi'an, 710072, China

Alejandro Parraga
Computer Vision Center Barcelona, Universitat Autonoma de Barcelona, Barcelona, 08193, Spain

1 Introduction

The description of color is important for many computer vision applications. The description of color is difficult because of the many factors that influence the color value, such as shadows, specularities, image compression, image blur, etc. One approach to address this problem is by means of photometric invariants [13, 11, 10] which are derived from reflectance models. These are invariant with respect to scene accidental events such as shadows, illuminant changes etc. However, these color descriptors are based on assumptions which are often unrealistic in computer vision applications, such as known gamma compression, and absence of image compression. In addition, they suffer from a drop in discriminative power [42].

Color names are linguistic labels which humans use to communicate the colors in the world. Examples of color names are 'red', 'olive' and 'beige'. Computational color names provide a mapping from color values to corresponding color names [31, 1, 43]. Because of their high discriminative power and robustness to photometric variations they were found to be an excellent color representation. In comparison to other color descriptors, including descriptors based on photometric invariance theory, the color name descriptors were found to obtain superior results in many application, especially for image classification [21], image retrieval [29],

object recognition [20], person reidentification [49], and visual tracking [7].

Berlin and Kay [4] in an influential linguistic study defined the term 'basic color term' as being (among other characteristics) a color name which is not subsumable under one of the other basic color terms. They then identified eleven such terms in English language, namely: black, blue, brown, green, gray, orange, pink, purple, red, white, and yellow. Most work on computational color names follow this convention and compute mappings for the eleven basic color terms [1, 43]. Studies on other color representations found that extending the set to more than eleven dimensions might be beneficial [24]. That resulted in the research questions which is addressed in this paper, how do we extend the color name set and does image understanding benefit from a larger color name set.

There are many possible color names which could be added to the eleven basic color terms [33]. However the problem is how to augment the color name set and what is the ordering for new color terms to be included. Inspired by [43, 24, 33], we propose a method which can augment the basic color terms with additional color names. Given a set of color names, we add color name which is less represented by the color names in the set. This procedure is iterated to produce a ranking of color names to add to the initial basic color terms. As a result we can compute new color name descriptors of arbitrary length (limited only by the size of our color name set). In the experiments we will evaluate color name sets of 15 and 25 and show that they outperform the color representations based on 11 color names. In conclusion the contributions of this paper are:

- We collect a new dataset of images to train an extended set of color names. The set contains a total of 39 color name categories.
- We propose a method which allows us to rank the additional color names, and therefore construct discriminative color name descriptors of arbitrary size. We also show that a naive extension of the color name descriptor leads to unsatisfying mapping of colors to color names, whereas our approach to extend the color name descriptors obtains much more acceptable mappings.
- We evaluate the new color name descriptor on visual tracking, person re-identification and image classification and show that the performance improves over the standard eleven dimensional color name descriptor. In addition, we design a psychophysical experiment which shows that our approach improves agreement to human users when labeling color patches with color names.

Upon publication we will publish the dataset and the newly computed color name mappings.

In the next section we will explain the database collection, and show our approach to ranking the color names. In Section 4 we introduce our approach to extending the color name set beyond eleven color names. In Section 5 we evaluate the color name descriptor and we conclude in Section 6.

2 Related work

Here we briefly summarize the related work on methods for color description in computer vision.

We distinguish between two main methodologies to the color description problem. The first methodology is based on reflection models which describe the interaction of light, material and sensors [13, 11, 10, 17]. From these reflection models photometric invariant descriptions of the material color can be derived. Given certain assumptions these descriptors can overcome the dependence of the color description on scene accidental events. Examples are color descriptions which are invariant to illuminant color, shadow-shading and specularities[9, 10, 40]. The main advantage of these methods is that they do not need training data and therefore do not require a laborious and costly labeling phase. The main drawback of these methods is that the assumptions on which they are based (for example white illumination, known acquisition device, etc) limit their application. Typically they require high-quality images without compression artifacts, and are not very effective for the medium quality images which are currently used in the many large scale data sets which have been collected from the internet.

The second methodology to color description is based on color names. Humans use color names routinely and seemingly without effort. They have been primarily studied in the fields of visual psychology, anthropology and linguistics [12]. Basic color terms have been studied in the influential work of Berlin and Kay [4]. They are defined as those color names in a language which are applied to diverse classes of objects, whose meaning is not subsumable under one of the other basic color terms, and which are used consistently and with consensus by most speakers of the language. Basic color names were found to be shared between languages. The number of basic terms varies from two in some indigenous languages to twelve in for example Russian.

Computational color naming [31, 2, 44] aims to learn a mapping from pixel values to color name labels. A clear example in computer vision where color names are desired is within the context of image retrieval, where a user might want to query for images with "blue sofas". The system recognizes the color name "blue", and orders the retrieved results on "sofa" based on their resemblance to the human usage of "blue'. Later research showed that color names actually also constitute an excellent color descriptor. They were found to be robust to photometric variations, while having in general higher discriminative power than the photometric invariants.

Fig. 1 Example images for the color ochre from the augmented color name dataset.

In recent years, these two approaches to color description, namely, the physics-based and the color name methods, have been compared on a wide variety of computer vision applications. In an earlier conference work, we provided an overview of applications where color names and photometric invariants were compared [41]. Constantly, color names were found to outperform the photometric invariance approaches by a significant margin. Color names have been extensively tested in image classification tasks [19, 21], object recognition [23], person re-identification [48] and action recognition [22]. The main reason for the success of color names is the high discriminative power which they possess, while being robust to photometric variations in images. It motivates us to investigate extending the color name set, with the aim to further improve the performance.

There have also been several attempts to divide the color space into categories using psychophysics, either by focusing on the regions of consensus [39, 5] or the categorical boundaries [35, 36]. All these models are based on a small subset of agreed focal colors.

3 The augmented color name dataset

In this section we explain the collection of the augmented color name dataset. The English language has hundreds of color names apart from the eleven basic color terms. To select a limited set we make use of two recent studies of color names in the English language [33, 32]. These studies investigated which color name words were widely used, had a shared meaning among the speaker population, be salient and therefore identifiable in an array of colors, and can be reliably distinguished in color space. They investigated a total of 28 candidates including beige, burgundy, cyan, fuchsia, lavender, lilac, magenta, maroon, mauve, ochre, olive, peach, plum, rose, salmon, tan, teal, turquoise, violet, burgundy, lilac, lime green, light green, dark green, dark purple, light blue, mustard, olive green, pale yellow and mint green.

The choice of training data to infer the mapping from RGB values to color names should be dictated by its application objective. In this paper we are interested in color name mappings which can be used for image understanding applications which in general are uncalibrated. We therefore also resort to learning the mapping from uncallibrated data crawled from Google similar as [43]. We collect images from Google by using the search query 'colorname + objects', e.g. 'mauve objects'. An example of six images for 'ochre objects' is provided in Fig. 1. The term 'objects' has been added to diversify the query results. A direct query for only the color term leads to color patches which do not represent colors in real-world situations. In total we collect 250 Google image per color name.

All images are considered to be in sRGB and they are gamma corrected accordingly. Even though these images come from a wide range of cameras, the lighting settings are unknown, and image compression is most likely applied. It has been shown that color names learned from such images provide better results in computer vision applications. This is caused by the fact that in computer vision applications also often the lighting is unknown, image or video compression has been applied, etc. For a further discussion on the differences on learning color names from calibrated and uncalibrated images we refer to [3]. To infer the color name from this dataset we transfer the images to histograms in L*a*b space. Pixels are represented by assigning their L*a*b space values into a finite vocabulary by assigning each value to a regular $10 \times 20 \times 20$ grid. [1].

4 Computation and Ranking of Additional Color Names

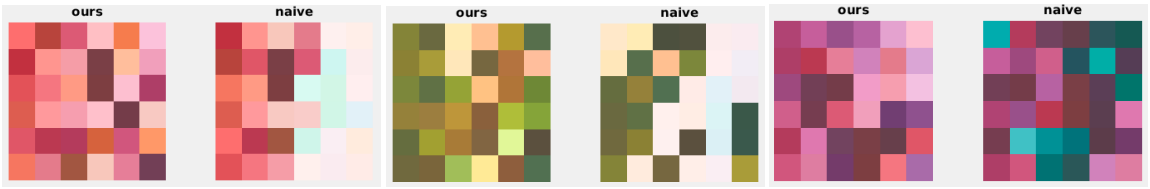
In this section we outline our approach to estimating the color name distribution from this data. We start by explaining the method from Van de Weijer et al. [43] for color name estimation and then we propose our approach to use correlation to rank color names.

4.1 Computation of color name mappings

The objective of computational color naming is to find $p(c|w)$ which is the probability of a color name c, given a color value w, to which we also refer as a color name mapping. For the computation of $p(c|w)$ we will use the algorithm proposed in [43]. However it requires some adaptation to be used for color name sets which include non basic color names.

We apply probabilistic latent semantic analysis (PLSA) [15] to estimate the probability of color values when given a color name. PLSA is a generative model, in this case on how

[1] Because the Lab-space is perceptually uniform we discretize it into equal volume bins. Different quantization levels per channel are chosen because of the different ranges: the intensity axis ranges from 0 to 100, and the chromatic axes range from -100 to 100.



(a) red (b) olive (c) fushsia

Fig. 2 Comparisons of the top 36 retrieved Munsell patches given a color name. We compare results of our method and the naive method to extend the color name set. Results clearly show that the naive approach fails to retrieve all relevant Munsell patches.

images are generated: the model assumes that images consist of a number of topics (in our case color names) which generate words (in our case RGB values). This model allows us to learn from noisy data such as the data set we collected from Google.

We model the distribution of RGB values w in an image i to be a mixture of color name topics c. In PLSA the conditioned distribution $p(w|i)$ is modeled by

$$p(w|i) = \sum_{c \in C} p(w|c)\, p(c|i). \tag{1}$$

where C is the set of color names. Here $p(w|i)$ is the collection of color histograms of the images and is known. Both $p(w|c)$ and $p(c|i)$ are unknown and need to be estimated. This can be done by minimizing the following loss

$$L = \sum_i \sum_w n(i,w) \log p(i,w) \tag{2}$$

with the EM algorithm. Here the joint distribution $p(i,w) = p(i)\, p(w|i)$ where $p(i)$ is considered uniform; $n(i,w)$ is the term frequency and can be directly computed from the training set.

Similar as [43] we introduce an additional term which enforces the color name mappings to be unimodal in LAB space. It enforces the distribution $p(w|c)$ to have a single mode and to decrease monotonically. Enforcing this is appropriate since we consider this to be a property of real color names. It can be obtained by adding a regularization term to the log likelihood:

$$L = \sum_i \sum_w n(i,w) \log p(i,w) - \gamma \sum_c \sum_w \left(p(c|w) - \rho_c(w)\right)^2 \tag{3}$$

here ρ_c is computed from the estimated distribution $p(c|w)$ with a grey scale reconstruction (for more details on this procedure we refer to [43]). The second term which is weighted according to γ enforces the estimated distribution to be close to the unimodal distribution by penalizing their difference.

As a second change to standard PLSA, an adjustment was proposed to allow for the usage of the weak label of the image (the labelling identifying the color name of the image) [43]. This can be done by assuming that $p(c|i)$ is drawn from a Dirichlet distribution of parameter α_{l_i}. Here $\alpha_{l_i}(c) = t \geq 1$ for $c = l_i$, and $\alpha_{l_i}(c) = 1$ otherwise. Here l_i is the label of image i. This leads to the following equation

$$p(c|i) \propto \left(\alpha_{l_i} - 1\right) + \sum_w n(i,w)\, p(c|w,i) \tag{4}$$

The computation of the distributions $p(w|c)$ and $p(c|i)$ is done by iteratively applying an EM-like algorithm, where we iterate between

- minimize Eq. 3 as a function of $p(w|c)$ with a conjugate gradient method,
- compute $p(c|i)$ according to Eq. 4,

until convergence. This provides us with the color name mappings $p(w|c)$ which we were aiming for. We use $t = 2$ and $\gamma = 200$ in our experiments.

4.2 Extending the color names set

One of the hurdles to extending the basic color term set with other color names is that the resulting color name set can no longer be interpreted as a probability distribution, i.e. for C larger than eleven the $\sum_{c \in C} p(c|w) \geq 1$. For example there are colors which can be described as being clearly 'violet', 'plum', 'purple' at the same time. This does not happen with the eleven basic color terms, because one of their main characteristics is that they are not subsumable under one of the other basic color terms. As a consequence the PLSA algorithm cannot be applied to color name sets which are larger than eleven because it is only valid when $\sum_{c \in C} p(c|w) = 1$. The violation of this equality increases with the number of additional color names.

To stress the fact that we are no longer working with probabilities we write $q(c|w)$ to be the membership of the color name c to the color word w. We allow $\sum_{c \in C} q(c|w) \geq 1$ and enforce $0 \leq q(c|w) \leq 1$. For example a color could have a membership of 1 to 'green', and 0.8 to 'lime'.

As mentioned above the violation is smallest in case we only add a single color name to the color name set. We therefore propose the following procedure for the estimation of $q(w|c)$: (1) for the eleven basic color terms we use the

PLSA algorithm, and set $q\left(w|c_{\{1,...,11\}}\right) = p\left(w|c_{\{1,...,11\}}\right)$, (2) for additional color names we compute $q\left(w|c_{\{12,...,39\}}\right)$ by adding a single color name at the time to the basic color terms and apply the PLSA algorithm. E.g. we add color name n to the basic color name set (yielding a total of 12 color names), estimate $p\left(w|c_{\{1,...,11,n\}}\right)$ and set $q\left(w|c_{\{n\}}\right) = p\left(w|c_{\{n\}}\right)$, and repeat this procedure for all color names not in the basic color name set. As a result of this procedure we have the $q(w|c)$ for the 11 basic color terms and the 28 additional color names. Finally, we obtain $q(c|w)$ by applying the Bayes theorem:

$$q(c_j|w) = \frac{q(w|c_j)}{\sum_{i=1}^{11} q(w|c_i)}. \quad (5)$$

where we assume a uniform prior over the color names.

In Fig. 2 we illustrate the importance of the iterative construction of the $q(c|w)$ which we propose here. If we would apply a naive extension of the method proposed in [43] the additional color names will compete with the basic color names, and we enforce $\sum_{c\in C} p(c|w) = 1$ to be true. As a result the borders of color names will move around when adding additional color names. This can be considered an undesired effect since colors which would previously considered to be 'red' with a high probability would suddenly be only considered 'burgundy'. In Fig. 2 we show the top 36 retrieved Munsell [18] patches given three color names (we consider a total of 329 Munsell patches). The retrieval shows the patches with the highest probability given the color name. Similar results were obtained for the other color names. We never observed the naive results to obtain a better selection of color names. Whereas our iterative scheme to compute $q(c|w)$ provides a relevant set of color patches, the naive approach only manages to return part of the relevant Munsell patches. As a consequence, in a retrieval application where a user looks for 'fuchsia' shoes she would only retrieval part of the relevant shoes from the dataset when based on the naive approach.

4.3 Ranking additional color names

In the previous section we have proposed to add 28 new color names to the basic color name set. In this section, we address the question how to rank these new color names. The ranking is of importance for the construction of compact color name representations. For example, if we would like a 15 dimensional color descriptor we would add the first four color names from the ranking to the eleven basic color terms.

When adding color names we would like them to be as different as possible to the ones which have already been selected. A color name which is significantly different from the existing set would increase the discriminative power of the combined color name set, and therefore improve its application to image understanding.

Fig. 3 (top row) The eleven basic color terms. (second and third row) proposed order in which to add 28 additional color names to the basic color term set.

Consider you would like to select the best color name from a color name set C_2 to add to a set of color names C_1. For brevity we will use the notation $B = q(w|c)$ (a matrix of 4000×39) where we use $b_i = q(w|c_i)$ and hence $B = [b_1, ...b_{39}]$. We write $\hat{B} = [\hat{b}_1, ...\hat{b}_{39}]$ to indicate the $L2$ normalized column vectors, and $\hat{B}^C$ to be the matrix B which contains the columns of the indexes included in set C. Given a color name set C_1 we will add the color name j^* from C_2 according to

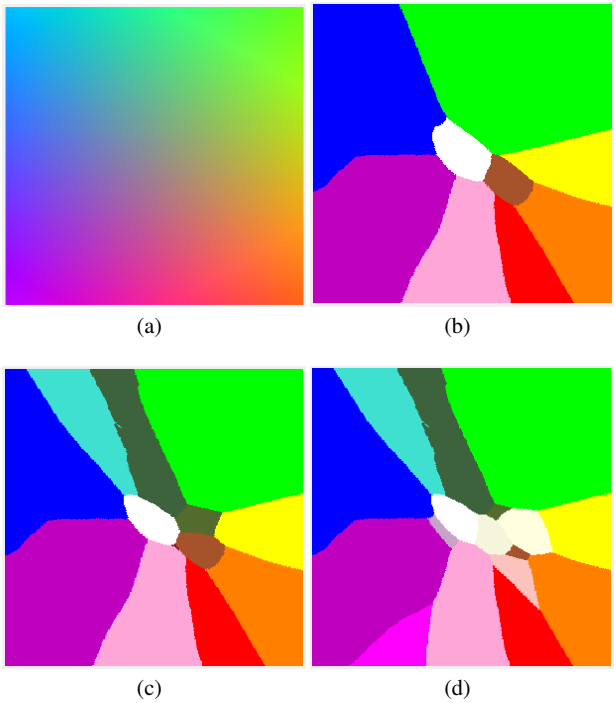
$$j^* = \arg\min_{j\in C_2} \left(\max\left(\left(\hat{B}^{C_1}\right)^{\mathrm{T}} \hat{b}_j\right)\right) \quad (6)$$

This equation considers for each of the potential color names the correlation with all the color names in set C_1. It then selects the color name which has the lowest maximum correlation and could therefore be considered the most different from the existing ones[2]. We initialize the process with $C_1 = \{1, \ldots 11\}$ containing the basic color terms and C_2 all other color names. Next Eq. 6 is applied N times, at each step increasing the color name set C_1 with j^* and removing it from C_2.

In Fig. 3 the results are shown when applying Eq. 6 until all color names have been selected. For example, as a set of 15 color names we would add 'turquoise','olive green', 'mint green' and 'burgundy' to the basic eleven color terms. Note that our approach selects 'turquoise' to be the 12th color name, which is interesting since there are some linguistic studies which suggest that 'turquoise' could be considered as a twelfth basic color term [50].

To illustrate the learned mappings we apply them to the challenging synthetic image with 11, 15 and 25 color names. The results are shown in Fig. 4. Here we only show the color name with the maximum probability for each pixel. Especially on the green-blue border and in the purple-pink-red region new color names are introduced to allow for more precise color descriptions.

[2] We also experimented with selecting the color name with the lowest mean correlation but results were inferior.



(a) (b) (c) (d)

Fig. 4 (a) the original image and the assignment based on (b) the 11 color names mapping; (c) the 15 ranked color name mapping; and the (d) the 25 ranked color name mapping.

5 Experimental results

The eleven basic color names are popular color descriptors and have been shown to obtain excellent results on a large variety of image understanding fields, including image classification [21], action recognition [22], image retrieval [29], person re-identification [49], and visual tracking [7]. In these papers, which compared the color name descriptor against a large variety of color presentations, the color name descriptor came out with superior results. Therefore, in these experiments we will compare our new extended color name descriptor against the standard color name descriptor based on the eleven basic color terms. We will evaluate the descriptor on three relevant computer vision applications namely visual tracking, person re-identification and image classification and we perform an additional user preference experiment.

5.1 Color Naming

In a first experiment we compare the PLSA pipeline we use for color naming against two baselines, namely SVM and k-nearest neighbors. To do so we perform the color name experiment from [44] where the task is to classify pixels from *Ebay dataset* images into the eleven basic color terms. The dataset contains a total of 440 images, consisting of ten images for the eleven color names for four different categories (cars, shoes, dresses, and pottery). All images come with a mask image which identifies the pixels which belong to the named object. Evaluation is only performed for the pixels in the mask. One example of an image and its ground truth mask is given in Fig. 5.

(a) (b)

Fig. 5 (a) example image from EBAY labeled with the color name 'green' and (b) the ground truth mask of the image identifying the pixels which are related with the color name. The results in Table 1 show the percentage of pixels on the mask which are labeled in agreement with the ground truth label.

Table 1 Percentage of correctly classified pixels in the Ebay dataset for various classifiers.

	PLSA	SVM	KNN
Accuracy	72.2%	69.30%	67.66%

All three methods are trained on the L*a*b-histograms of Google images. For the PLSA we use the setup as explained in Section 3. For SVM we use linear kernel [3] where we cross validate for optimal c value. For k-nearest neighbor we optimize for k on the validation set and found 25 to be optimal.

The results of this experiment are provided in Table 1. We can see that the PLSA algorithm obtains superior results compared to both SVM and k-nearest neighbors. In addition we have applied the three methods to a synthetic image and results are provided in Fig. 6. We can see that PLSA manages to obtain smoother edges than k-nearest neighbor and SVM, and that SVM makes many errors for the highly saturated colors (along the borders of the image). These are colors which are less frequent in real images, and therefore have fewer training examples. In conclusion, PLSA based color naming outperforms other popular classifiers for the task of color naming and we will perform the remaining experiments based on the color names which are computed with PLSA.

[3] We found that more complex kernels such as for example intersection did not improve results.

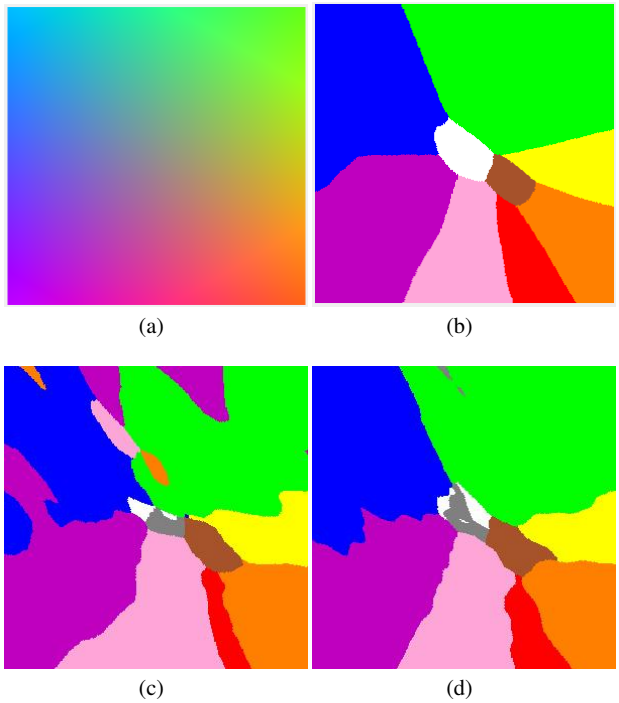


Fig. 6 (a) the original image and the assignment based on (b) PLSA; (c) SVM; and the (d) k-nearest neighbors on the 11 color name mapping.

5.2 Image classification

For image classification we perform experiments on the Oxford Flower102 dataset [34] which contains 8189 images of 102 different kinds of flower (see Fig. 7). It has been selected because of the importance of color for flower classification and the real-world challenges such as significant scale and illumination changes. We follow the standard bag-of-words (BOW) [6] approach. In BOW an image is firstly represented by a collection of local image features, and then each local feature is discretized into a visual vocabulary from the represented cues such as color and shape. Then images are represented as a histogram over visual words. For classification we apply an SVM with intersection kernel.

In a first experiment we compare our proposed method for ranking color names (see Section 4.3) to two baseline methods:

- RANDOM: a color name set with more than eleven color names is constructed by choosing the eleven basic color terms and randomly adding additional color names until the desired number is reached.
- LABCN: this method is derived from the mean LAB values of the color names. Following the notation of Section 4 the mean of each color name is computed according to:

$$\mu_j^{LAB} = \sum_i LAB(w_i)\, p(w_i|c_j) \tag{7}$$

Table 2 Classification accuracy on Oxford Flower102 with different methods.

Accuracy	Ours	LABCN	Random
11	37.23%	37.23%	37.23%
15	37.73%	37.58%	37.60%
25	39.34%	38.61%	38.84%

which is a weighted mean which is computed by multiply the LAB value of the color value w given by $LAB(w_i)$ with the probability of the color value w belonging to the color name c_j. The ranking is then obtained by replacing the selection of Eq. 6 by:

$$j^* = \arg\max_{j \in C_2} dist\left(C_1, \mu_j^{LAB}\right) \tag{8}$$

where the distance between the set of color names C_1 and the color name j given by $dist\left(C_1, \mu_j^{LAB}\right)$ is defined to be equal to the minimum distance of color name j to any of the member of C_1. Thus, the algorithm computes the LAB color name centers, and starting from the eleven color names, adds iteratively that color name which is furthers away from any of the already selected color names.

We test the three different rankings with 15 and 25 color names on flower classification application. Results are shown in Table. 2. As can be seen increasing the set of color names increases the performance, and results improve with 2.1% for our method. A larger number of color terms can enhance the discriminative power but also weaken the photometric invariance. We found that increasing color names beyond 25 color terms did not further improve results.

Next we compare to the two other baselines for ranking the color names. For eleven color names the methods are equal because they all consider the same eleven basic color names. The results show that our method is slightly better than the RANDOM and LABCN baselines when using 15 color names. The difference gets larger when considering 25 color names. When considering the performance gain with respect to eleven color names our method obtains a gain of 2.1% whereas the baseline methods only improve by around 1.4%. The fact the LABCN does not outperform the RANDOM method could be caused by the fact that even though it is selecting color names which describe colors currently not well described by the color name set, it does not take into account the frequency of these colors occurring.

In a second experiment on the Oxford Flower102 dataset we compare the color name descriptor to the discriminative color descriptors (DD) proposed in [24]. These descriptors are not semantic, which are not linked to human color names, but were found to obtain state-of-the-art results. They proposed two sorts of discriminative descriptors: database

Fig. 7 Example images from Flower102 dataset

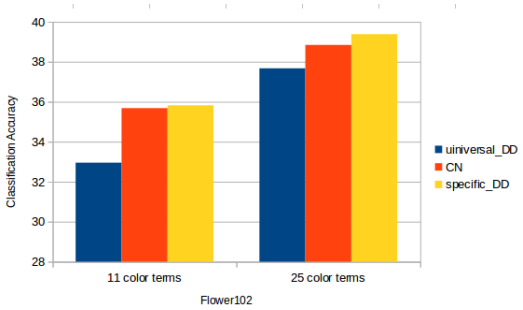


Fig. 8 Classification accuracy on Oxford Flower102 comparing color names with discriminative color descriptors.

specific color name descriptors which are optimized for a specific classification problem and need to be learned from labeled training data, and universal color name descriptors which are learned from several databases and can then be applied without adapting them to the specific dataset. The color names which we propose in this paper are universal color descriptors since they do need to be relearned for new datasets.

The results of classification are provided in Fig. 8. The results show that the descriptor based on the 25 color name set outperforms the universal discriminative descriptor with the same dimension. The dataset specific color descriptor only slightly outperforms this results. Given that the difference is very small, for many applications it might be preferable to apply the 25 color name descriptor which does not require dataset specific training.

5.3 Color naming for tracking

Visual tracking is a challenging problem in computer vision. Recent work has shown that color names provide superior performance when compared to other color representations for visual tracking [7]. Their tracker is based on the CSK

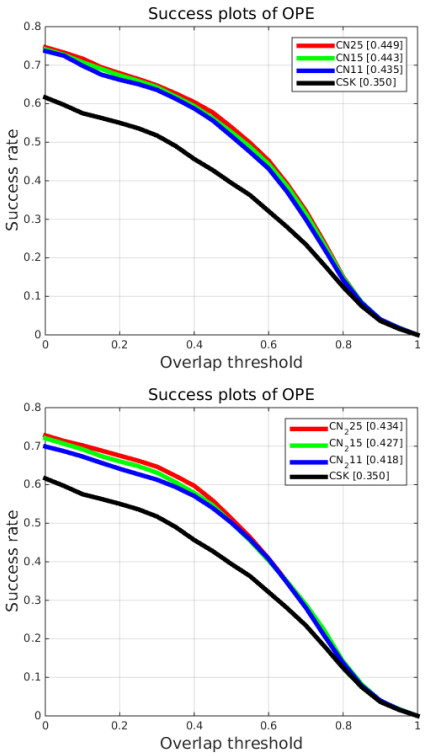


Fig. 9 Success plots for (top) various different color name mappings, and (bottom) various compressed color name mappings.

tracker [14] which is a correlation filter based tracker which only considers the luminance channel. In [7] they show that extending the tracker with color names provides a significant performance improvement. We will apply the same tracker in our experiments, but we will replace the eleven color name mapping with the mappings we have derived here. Results are provided for color name representations with 15 and 25 color names, where the selection is performed with Eq. 6. An additional weighting term λ was introduced to balance the luminance and color channels[4]. Since the introduction of color names in tracking [7], they have been applied in several state-of-the-art trackers [8, 16, 27] showing that color names are among the preferred color representations.

The experiments are performed on an Intel(R) Xeon(R) CPU E5-1620 v3 @ 3.50GHz CPU with 32 GB RAM with a native Matlab implementation. In our approach, we use the same parameter values as suggested by Danelljan et al. [7] for the ACT tracker. We also employ the same dataset, including 35 color sequences used in the evaluation of tracking methods [47] and 6 other color sequences namely: Kitesurf, Shirt, Surfer, Board, Stone and Panda. The sequences used in our experiments pose challenging situations such as motion blur, illumination changes, heavy occlusions, low resolution, fast motion, in-plane and out-of-plane rotations, scale

[4] We found for optimal results were obtained with a $\lambda = 1$ for 11 color names, a $\lambda = 0.9$ for 15 color, and a $\lambda = 0.8$ for 25 color.

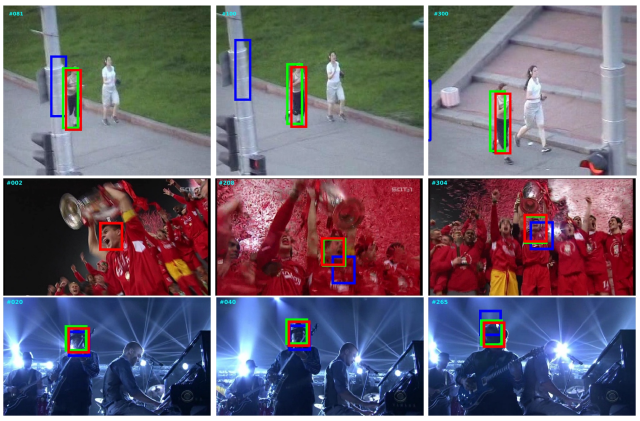


Fig. 10 Comparison of 3 different color name representations for trackers in challenging situations such as illumination variation, occlusion, motion blur and in-plane rotation. The example frames are from the Jogging, Soccer and Shaking sequences respectively. The results of 11D, 15D, and 25D are represented by blue, green and red boxes respectively.

variation, out of view and background clutter. To validate the performance of our approach, we follow the protocol used in [47].

In the first experiment we compare the tracker using three different color name mappings: with the original 11 and with the two new color mappings of 15 and 25. Fig. 9(top) shows the success plots. The success plot contains the overlap precision (OP) over a range of thresholds. OP is defined as the percentage of frames where the bounding box overlap exceeds a threshold $th \in [0,1]$. The trackers are ranked using the area under the curve (AUC). As the channels of the color name mappings increase, the performance of the color tracker improves. The 25 dimensional color mapping obtains a 28% relative gain over the original CSK tracker which obtains 35% OPE score.

Danelljan et al. [7] pointed out that the speed of the tracker decreases with the number of channels and therefore the color name based trackers are significantly slower. However, they proposed to dynamically map the color representation to a lower dimensional representation (they show that 2 dimensions is enough). When we apply the same dynamic dimensionality reduction to our trackers we obtain the results which are presented in Fig. 9(bottom). The results slightly deteriorate with respect to the full representation but the speed increases from 89 to 128 fps for 15 dimensions and from 66 to 110 fps for 25 dimensions. In Fig. 10, we illustrate the results of the trackers on three sequences. Note for example, in the Jogging sequence, occlusion appears in frame #81, and the traditional low dimensional color name representation used in ACT [7] fails to track the woman, but the tracker using high dimensional color names can re-detect the position of the woman after occlusion.

Finally, we show the performance of the three trackers for several attributes as proposed by Wu et al. [47]. The 25 dimensional color name mapping improves over the standard 11 dimensional color mapping for all the eleven attributes. In Fig. 11 the results for four of them are shown. It can be seen that with increasing dimensionality of the color name mapping the performance for illumination variation, low resolution, motion blur and occlusion improves. Especially the performance gains for low resolution (relative gain of 21%) and illumination variation (relative gain of 12%) are noteworthy.

5.4 Color naming for re-identification

In several recent studies [49, 48, 28], color names have been extensively used to encode color information for person re-identification. To validate our approach, we perform the experiments on the challenging Market-1501 dataset [49] for the person re-identification task. The dataset comprises of 32668 annotated bounding boxes of 1501 identities. We follow the bag-of-words pipeline as described in [49]. A visual vocabulary is constructed using the standard K-means algorithm on the training bounding boxes. For fair comparison, we fixed the size of visual vocabulary to 350 words for all color descriptors. For each local color feature, a Multiple Assignment (MA) strategy is employed to locate its nearest neighbor under Euclidean distance. The MA parameter is fixed to 10 visual word indices. The performance is measured by using a cumulative matching characteristic (CMC) curve which plots the probability of correct identification compared to the candidates returned by the method. A rank-1 score is then computed which denotes the expectation of the correct match.

Table 3 shows the performance comparison using different color descriptors on the Market-1501 dataset. The color descriptor with 11 dimensions achieves the mAP score of 12.02% and rank-1 score of 31.80%. The performance improves by increasing the number of color names bins. The best results are obtained using 25 color names with a mAP score of 13.45% and rank-1 score of 34.65%. We have also run the same experiment with the 16 dimensional color name representation proposed in [48] for the task of person re-identification. We found the results to be inferior to ours. In Fig. 12 an example of queries with 11 and 25 color names are provided. The 11 color name representation fails to distinguish between several color tones which are better described in the 25 dimensional representation.

5.5 User preference experiment

In Fig. 2 we illustrated the importance of our proposed method for the computation of extended color name sets when compared to a *naive* approach, which directly applies PLSA to

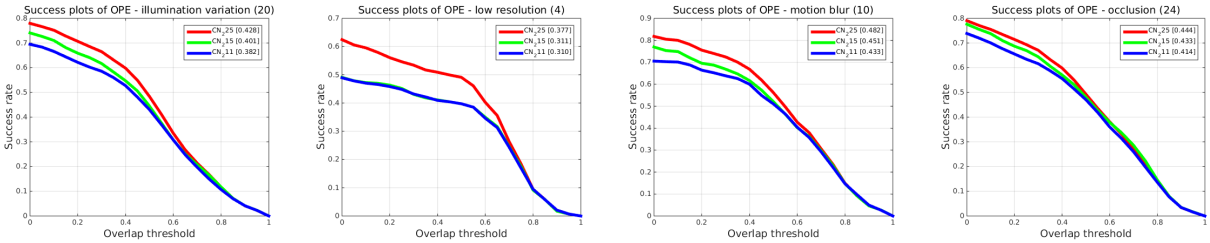


Fig. 11 Success plots for several attributes, including illumination variation, low resolution, motion blur and occlusion. The increased color name representation outperforms the original color name representation for these attributes.

Fig. 12 Examples for top 3 ranking with 11 and 25 color terms. Note that some of the errors which occur when using 11 color names are resolved when using 25 color names.

Fig. 13 Setup for our psychophysical experiment. Color patches and color names were presented on a calibrated CRT monitor and the observer pressed buttons on a gamepad to decide whether the color patch was well described by the name or not. The background was mid-grey and a reference white was provided by a D65-colored frame

Table 3 Re-id performance comparison of different color descriptors on the Mart-1501 dataset. The best results are obtained using 25 color names.

	11	15	16[a]	25
mAP	12.02	12.85	10.93	13.45
r = 1	31.80	33.46	30.34	34.65

[a] yangyang's 16-d color features for re_identification with our experiment settings [48].

the extended color name set. We have designed a psychophysical experiment to quantify the difference between our method and the naive approach. In the experiment we focus on the color names where the two methods do not agree.

We performed a forced choice psychophysical experiment where observers had to decide whether a given color patch was described by a given color name (shown in writing at the top) or not. The stimuli were presented on a calibrated CRT monitor (Sony GDM F500-R) run by a Visage Mk1 stimulus generator. The screen boundary consisted of a 5cm wide frame that acted as reference white (D56, 64 Cd/m^2). The experimental setup was as follows: after 2 minutes of dark adaptation, observers were presented with an image patch centered on a mid-gray screen, and a color name written on top. Their task was to press the left or right buttons of a gamepad to decide whether the name described the color of the patch correctly (yes-no choice). Once observers made their choices (there were no time constraints), the screen was refreshed, the next patch and color name appeared and the trial was repeated. Setup for our psychophysical experiment is shown in Fig. 13. The patches were randomly selected from a list of 162 samples where there was disagreement between the two methods. The color names were obtained from each of the two methods tested, ours and naive method. There were 10 subjects (university students, eight male and two female) and they all had normal color vision tested by the Ishihara color-blindness test. We also made sure all of them were familiarized to the same color terms before the experiment by showing them a series of cards with images of objects (obtained from Google images) categorized under the same color name. We did two runs of the experiment per subject. The first run was considered "training" and was

Table 4 User communication results on Munsell patches.

	Ours	naive
Accuracy	55.5%	45.5%

discarded. The results of the second run are presented in table 4, which shows the percentage of times subjects preferred each method's categorization. Our method's solution was preferred 10% more than that of naive method.

6 Discussion and Conclusions

Color description is an important part of image understanding. It is a difficult problem because colors vary due to accidental events such as shadow, shading, specularities, viewing angle, image compression, etc. The most popular approach to address this problem is by means of photometric invariants derived from reflectance models. However, it was found that descriptors based on color names often obtained better results for computer vision applications. Color names are therefore applied in many applications such as image classification, object detection, action recognition, texture recognition, object tracking, and person re-identification.

Traditionally color name mappings are restricted to the eleven basic color name terms. In this paper, we proposed a method to compute the color name mappings for large color name sets. For this purpose we collected a new data set of 28 additional color names. We have shown that a naive extension of the color name descriptor leads to unsatisfying mapping of colors to color names. To solve this problem we propose an iterative scheme to extend the color name descriptor. In addition, we propose a method to rank the additional color names. Using the ranking we can compute color name representations of arbitrary length. In our experiments we evaluate the impact of increasing the number of color names to visual tracking, person re-identification, and image classification. In all cases adding color names was found to improve the results significantly. In addition, a psychophysical experiments shows that our approach has a larger agreement to human users of labeling patches with color names.

The recent advances of Deep Learning have influenced computer vision research greatly. Driven by the availability of large datasets and improved hardware (GPU computation) these algorithms can jointly learn feature representations and classifiers [25, 26]. They have been shown to be successful on many computer vision application [38] and outperform hand crafted features. They have also been shown to effectively learn attributes of objects, including color, and texture attributes [30, 37, 45]. However, due to the absence of large color name datasets color naming with deep networks has been limited to the eleven basic color names [46]. The dataset proposed in this paper could be used to train networks for larger color name sets. In that case the discussion on the difficulties of extending the color name set beyond the basic color terms (see Section 4.3) should be taken into account when designing the loss function of the network. A simple softmax loss would enforce the probabilities over the color names to sum to one, and would most probably demonstrate some of the shortcomings we have shown in this paper that 'naive' approaches have.

Acknowledgements We acknowledge Dimitris Mylonas for his helpful suggestion on extending the color name set. We also acknowledge Nicole Walasek who has been of great help in the data set collection and PLSA code preparation. Lu Yu acknowledges the Chinese Scholarship Council (CSC) grant No.201506290126. This work was supported by TIN2013-41751-P and TIN2016-79717-R of the Spanish Ministry and the CERCA Programme / Generalitat de Catalunya.

References

1. R. Benavente, M. Vanrell, and R. Bladrich. A data set for fuzzy colour naming. *COLOR research and application*, 31(1):48–56, 2006.
2. R. Benavente, M. Vanrell, and R. Baldrich. Parametric fuzzy sets for automatic color naming. *Journal of the Optical Society of America A*, 25(10):2582–2593, 2008.
3. Robert Benavente, Joost Van de Weijer, Maria Vanrell, Cordelia Schmid, Ramon Baldrich, Jakob Verbeek, and Diane Larlus. Color names. In *Color in Computer Vision*. Wiley, 2012.
4. B. Berlin and P. Kay. *Basic color terms: their universality and evolution*. Berkeley: University of California, 1969.
5. Robert M Boynton and Conrad X Olson. Locating basic colors in the osa space. *Color Research & Application*, 12(2):94–105, 1987.
6. Gabriella Csurka, Christopher Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, volume 1, pages 1–2. Prague, 2004.
7. Martin Danelljan, Fahad Shahbaz Khan, Michael Felsberg, and Joost van de Weijer. Adaptive color attributes for real-time visual tracking. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1090–1097. IEEE, 2014.
8. Martin Danelljan, Gustav Häger, Fahad Shahbaz Khan, and Michael Felsberg. Learning spatially regularized correlation filters for visual tracking. In *ICCV*, 2015.
9. Graham D. Finlayson, Bernt Schiele, and James L. Crowley. Comprehensive colour image normalization. In *ECCV '98: ProcP of the 5th European Conference on Computer Vision-Volume I*, pages 475–490. Springer-Verlag, 1998. ISBN 3-540-64569-1.

10. J.M. Geusebroek, R. van den Boomgaard, A.W.M. Smeulders, and H. Geerts. Color invariance. *PAMI*, 23(12):1338–1350, 2001.
11. Th. Gevers and A. Smeulders. Color based object recognition. *Pattern Recognition*, 32:453–464, 1999.
12. C.L. Hardin and L. Maffi, editors. *Color Categories in Thought and Language*. Cambridge University Press, 1997.
13. G. Healey. Segmenting images using normalized color. *IEEE Trans. Syst., Man, Cybern.*, 22:64–73, 1992.
14. João F Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista. Exploiting the circulant structure of tracking-by-detection with kernels. In *European conference on computer vision*, pages 702–715. Springer, 2012.
15. Thomas Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 50–57. ACM, 1999.
16. Zhibin Hong, Zhe Chen, Chaohui Wang, Xue Mei, Danil Prokhorov, and Dacheng Tao. Multi-store tracker (muster): A cognitive psychology inspired approach to object tracking. In *CVPR*, 2015.
17. Cong Phuoc Huynh and Antonio Robles-Kelly. A solution of the dichromatic model for multispectral photometric invariance. *International Journal of Computer Vision*, 90(1):1–27, 2010.
18. Kenneth Low Kelly and Deane Brewster Judd. *Color: universal language and dictionary of names*, volume 440. US Department of Commerce, National Bureau of Standards, 1976.
19. Fahad S Khan, Joost Weijer, Andrew D Bagdanov, and Maria Vanrell. Portmanteau vocabularies for multi-cue image representation. In *Advances in neural information processing systems*, pages 1323–1331, 2011.
20. Fahad Shahbaz Khan, Rao Muhammad Anwer, Joost van de Weijer, Andrew D Bagdanov, Maria Vanrell, and Antonio M Lopez. Color attributes for object detection. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3306–3313. IEEE, 2012.
21. Fahad Shahbaz Khan, Joost Van de Weijer, and Maria Vanrell. Modulating shape features by color attention for object recognition. *International Journal of Computer Vision*, 98(1):49–64, 2012.
22. Fahad Shahbaz Khan, Rao Muhammad Anwer, Joost van de Weijer, Andrew D Bagdanov, Antonio M Lopez, and Michael Felsberg. Coloring action recognition in still images. *International journal of computer vision*, 105(3):205–221, 2013.
23. F.S. Khan, R.M. Anwer, J. van de Weijer, A.D. Bagdanov, M. Vanrell, and A.M. Lopez. Color attributes for object detection. In *IEEE conference on Computer Vision and Patter Recognition*, 2012.
24. Rahat Khan, Joost Van de Weijer, Fahad Shahbaz Khan, Damien Muselet, Christophe Ducottet, and Cecile Barat. Discriminative color descriptors. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 2866–2873. IEEE, 2013.
25. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
26. Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4): 541–551, 1989.
27. Yang Li and Jianke Zhu. A scale adaptive kernel correlation filter tracker with feature integration. In *ECCV Workshop*, 2014.
28. Xiaokai Liu, Hongyu Wang, Yi Wu, Jimei Yang, and Ming Hsuan Yang. An ensemble color model for human re-identification. In *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*, pages 868–875. IEEE, 2015.
29. Ying Liu, Dengsheng Zhang, Guojun Lu, and Wei-Ying Ma. Region-based image retrieval with high-level semantic color names. In *Multimedia Modelling Conference, 2005. MMM 2005. Proceedings of the 11th International*, pages 180–187. IEEE, 2005.
30. Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3730–3738, 2015.
31. A. Mojsilovic. A computational model for color naming and describing color composition of images. *IEEE Transactions on Image Processing*, 14(5):690–699, 2005.
32. D. Mylonas, D. Griffin, L.D. Purver, P. Katemake, and Davidoff J. The role of primary colours in colour naming. *Under review*, 2016.
33. Dimitris Mylonas and Lindsay MacDonald. Augmenting basic colour terms in english. *Color Research & Application*, 2015.
34. Maria-Elena Nilsback and Andrew Zisserman. Automated flower classification over a large number of classes. In *Computer Vision, Graphics & Image Processing, 2008. ICVGIP'08. Sixth Indian Conference on*, pages 722–729. IEEE, 2008.
35. C Párraga, R Benavente, R Baldrich, and M Vanrell. Psychophysical measurements to model intercolor regions of color-naming space. *Journal of Imaging Science and Technology*, 53(3):31106–1, 2009.
36. C Alejandro Parraga and Arash Akbarinia. Nice: A computational solution to close the gap from colour

perception to colour categorization. *PloS one*, 11(3): e0149538, 2016.
37. Gabriel Schwartz and Ko Nishino. Discovering perceptual attributes in a deep local material recognition network. *arXiv preprint arXiv:1604.01345*, 2016.
38. Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 806–813, 2014.
39. Julia Sturges and TW Whitfield. Locating basic colours in the munsell space. *Color Research & Application*, 20 (6):364–376, 1995.
40. Koen E. A. van de Sande, Theo Gevers, and Cees G. M. Snoek. Evaluating color descriptors for object and scene recognition. *PAMI*, 32(9):1582–1596, 2010.
41. Joost Van De Weijer and Fahad Shahbaz Khan. An overview of color name applications in computer vision. In *International Workshop on Computational Color Imaging*, pages 16–22. Springer, 2015.
42. Joost van de Weijer and Cordelia Schmid. Applying color names to image description. In *IEEE International Conference on Image Processing (ICIP)*, San Antonio, USA, 2007.
43. Joost Van De Weijer, Cordelia Schmid, Jakob Verbeek, and Diane Larlus. Learning color names for real-world applications. *Image Processing, IEEE Transactions on*, 18(7):1512–1523, 2009.
44. Joost van de Weijer, Cordelia Schmid, Jakob Verbeek, and Diane Larlus. Learning color names for real-world applications. *IEEE Transactions on Image Processing*, 18(7):1512–1524, july 2009.
45. Yan Wang, Sheng Li, and Alex C Kot. On branded handbag recognition. *IEEE Transactions on Multimedia*, 18(9):1869–1881, 2016.
46. Yuhang Wang, Jing Liu, Jinqiao Wang, Yong Li, and Hanqing Lu. Color names learning using convolutional neural networks. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 217–221. IEEE, 2015.
47. Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. Online object tracking: A benchmark [c]. computer vision and pattern recogniton (cvpr). In *2013 IEEE Conference on, IEEE*, 2013.
48. Yang Yang, Jimei Yang, Junjie Yan, Shengcai Liao, Dong Yi, and Stan Li. Salient color names for person re-identification. In *Computer Vision (ECCV), 2014 European Conference on*, pages 536–551. Springer, 2014.
49. Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Computer Vision (ICCV), IEEE International Conference on*, 2015.
50. Heinrich Zollinger. Why just turquoise? remarks on the evolution of color terms. *Psychological Research*, 46 (4):403–409, 1984.