

# APPLYING COLOR NAMES TO IMAGE DESCRIPTION

*Joost van de Weijer\* , Cordelia Schmid*

LEAR TEAM

INRIA Rhône-Alpes

655, avenue de l'Europe, 38330 Montbonnet, France

{joost.van-de-weijer, cordelia.schmid}@inrialpes.fr

## ABSTRACT

Photometric invariance is a desired property for color image descriptors. It ensures that the description has a certain robustness with respect to scene incidental variations such as changes in viewpoint, object orientation, and illuminant color. A drawback of photometric invariance is that the discriminative power of the description reduces while increasing the photometric invariance. In this paper, we look into the use of color names for the purpose of image description. Color names are linguistic labels that humans attach to colors. They display a certain amount of photometric invariance, and as an additional advantage allow the description of the achromatic colors, which are undistinguishable in a photometric invariant representation. Experiments on an image classification task show that color description based on color names outperforms description based on photometric invariants.

Index terms: image color analysis, image classification, image representations, object recognition.

## 1. INTRODUCTION

There exists broad agreement that local features are an efficient tool for image classification due to their robustness with respect to occlusion and geometrical transformations [1]. Of the multiple approaches which have been proposed to describe the shape of local features the SIFT descriptor [2] was found to be among the best [3], and is currently the most used shape descriptor. Only recently people have started to enrich local image descriptors with color information [4, 5, 6].

The main challenge for color description is to obtain robustness with respect to photometric variations as are common in the real world, such as shadow and shading variations and changes of the light source color. For this purpose color descriptors are generally based on photometric invariants [4, 5], such as hue and normalized RGB. In increasing the amount of invariance one should always consider the loss of discriminative power. Photometric invariants are, for instance, blind to the achromatic colors, black, grey, and white,

because from a photometric point of view these could be produced from the same patch by varying the intensity. It is however questionable if for real-world applications full photometric invariance is optimal, and the negative effect due to the loss of discriminative power is not too high.

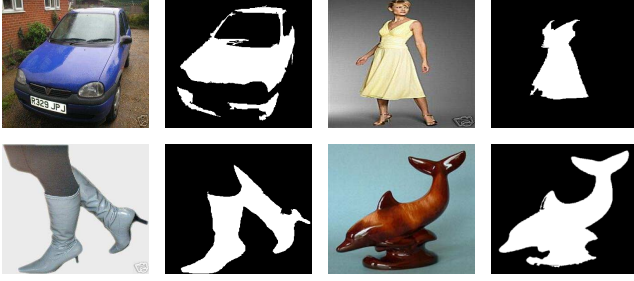
In describing the colors of objects in the real-world people make use of color names such as "red", "black" and "olive". Color names have been primarily studied in the fields of visual psychology, anthropology and linguistics [7]. Within an image understanding context color naming is the action of assigning a linguistic color label to image pixels [8, 9], and has been mainly used in image retrieval applications [10]. Color names possess a certain degree of photometric invariance. In addition color names include labels for the achromatic colors: "black", "grey" and "white", which from a photometric invariance point of view are impossible to distinguish.

In this paper, we explore the applicability of color names as a basis for color image description. As discussed they possess robustness to photometric variations, while preserving the discriminative power to distinguish achromatic colors. The aim is to find out if the traditional choice to use photometric invariants to describe the color content, should not be replaced by a new research direction, in which partial photometric invariance and discriminative power are balanced in a more sophisticated way.

## 2. COLOR NAME DESCRIPTOR

The set of color names used in English is vast and includes labels such as "white", "green", "pastel", and "light blue". In this paper we use the 11 basic color terms of the English language: black, blue, brown, grey, green, orange, pink, purple, red, white, and yellow. The basic color terms were defined in the influential work on color naming of Berlin and Kay [11]. A basic color term of a language is defined as being not subsumable to other basic color terms (e.g. turquoise can be said to be a light greenish blue, and is therefore not a basic color term). We define the color descriptor  $\mathbf{K}$  as the vector containing the probability of the color names given an image

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



**Fig. 1.** Ebay images and hand-segmentations: a "blue" car, a "yellow" dress, "grey" shoes, and "brown" pottery.

region  $R$

$$\mathbf{K} = \{p(n_1 | R), p(n_2 | R), \dots, p(n_{11} | R)\} \quad (1)$$

with

$$p(n_i | R) = \frac{1}{N} \sum_{x \in R} p(n_i | \mathbf{f}(x)) \quad (2)$$

where  $n_i$  is the  $i$ -th color name,  $x$  are the spatial coordinates of the  $N$  pixels in region  $R$ ,  $\mathbf{f} = \{L^*, a^*, b^*\}$ , and  $p(n_i | \mathbf{f})$  is the probability of a color name given a pixel value.

We compute the probabilities  $p(n_i | \mathbf{f})$  from a set of manually annotated images. Here, we use images from the Ebay auction website. On this site users describe objects for sell with an explanatory text, often containing color names. We have assembled 440 such images, 40 images per color name, and hand-segmented the regions corresponding to the color label (see Fig. 1). All images are gamma corrected. Next histograms of the pixels in the segmentation regions are computed in  $L^*a^*b^*$ -space with cubic interpolation (we use a histogram size of 10x20x20 bins in  $L^*a^*b^*$ -space). The 40 normalized histograms of each color name are averaged to compute  $p(\mathbf{f} | n_i)$ . Subsequently,  $p(n_i | \mathbf{f})$  is computed with Bayes Law by assuming an equal prior over the color names. We experimented with simple color constancy algorithms (Grey-World and max-RGB) to compensate for the illuminant color, but found that they did not improve classification result.

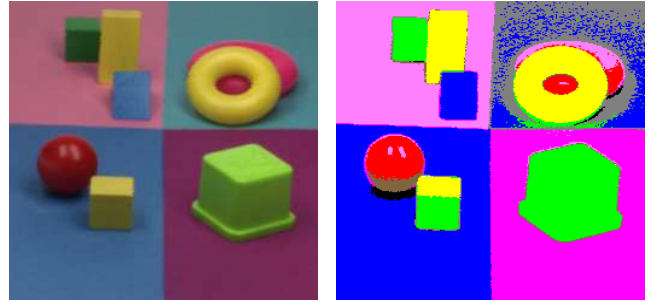
We proposed an alternative approach to compute the color name distributions in [12]. There the distributions are automatically learned from weakly labelled images which are retrieved with Google image search.

### 3. BALANCING PHOTOMETRIC INVARIANCE VERSUS DISCRIMINATIVE POWER

In the introduction, we questioned whether in balancing photometric invariance versus discriminative power, existing color descriptors do not focus too much on photometric invariance. In Table 1 the occurrence of color names in about 40.000 images of the Corel Data set is given. The color name distribution is computed by appointing each pixel to its most probable color name, with  $\arg \max_{n_i} p(n_i | \mathbf{f}(x))$ . Striking is the

| black | blue | brown | grey | green | orange | pink | purple | red | white | yellow |
|-------|------|-------|------|-------|--------|------|--------|-----|-------|--------|
| 19    | 12   | 23    | 19   | 10    | 2      | 2    | 2      | 4   | 6     | 1      |

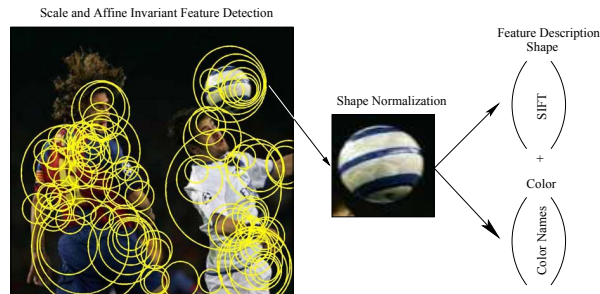
**Table 1.** Percentage of pixels assigned to color name in Corel data set.



**Fig. 2.** Example of color name assignment on toy-image. The color names are represented by their corresponding color, e.g. the pixels on the ball are assigned to red, pink and brown. Note that color names display a certain amount of photometric robustness.

abundance of the achromatic colors: 44% of the pixels is appointed to an achromatic color name. The dark color "brown" is most common and only "blue" and "green" can be said to be frequent bright colors. For photometric invariant representations achromatic colors are undefined, and consequently remain undescribed by the descriptors, resulting in a considerable loss of discriminative power.

On the other hand, a certain degree of photometric invariance is desired. It avoids that many instances of the object, under various viewing angles and lighting conditions, have to be represented in the training set. As an alternative to photometric invariants we propose to use color names, which exhibit robustness with respect to photometric variations, while remaining capable to distinguish between the achromatic colors. In Fig. 2 an illustration of the photometric robustness of color names is shown: the most likely color name for each pixel is depicted. Most objects are given a single color name independent of shading and specular effects, e.g. the blocks in the upper left quarter are described by a single color name. However the photometric robustness is limited: multiple color names are assigned to the yellow block and the red ball in the left bottom quarter. A full photometric invariant such as *hue* would consider all pixels of the yellow block and red ball as belonging to the same color. In the experimental section we investigate if the loss in photometric invariance in going from full photometric invariants to color names is compensated by a gain in discriminative power.



**Fig. 3.** Overview of bag-of-words approach. In the feature detection phase a set of informative points at various scales are detected. Each feature is subsequently described by its shape and color information in the feature description phase.

#### 4. BAG-OF-WORDS APPROACH TO IMAGE CLASSIFICATION

In the experimental phase the color name descriptor is tested for image classification. In image classification the task is to predict the presence of an object in a test image. The bag-of-words approach is generally considered the most successful approach to image classification [13]. The approach has been motivated on the bag-of-words approach to document analysis. In the case of image classification the words are replaced by visual words representing frequent structures in images such as blobs, corners, T-junctions, etc. The method starts by detecting a set of image regions (we apply a Harris-Laplace detector [1]) and their corresponding scales in an image (see Fig. 3). Next, in the description phase, all patches are normalized to a standard size and a descriptor is computed for all regions. The descriptors are clustered by a K-means algorithm to form the set of visual words. Subsequently, each image is represented by a frequency histogram of the visual words. Based on these histograms a classifier is trained for all classes (we apply a linear SVM). A test image is classified with all classifiers, and is appointed to the class for which it obtained the highest SVM score.

The majority of the bag-of-words methods are based on shape alone, and ignore color information. Van de Weijer and Schmid [4] extended the model to also include color information. Their color description is based on histograms of photometric invariants. In this paper, we aim to improve the discriminative power of the color description. For this purpose we base the color description on color names instead of photometric invariants. The combined shape and color descriptor is computed as follows. For the shape description,  $\mathbf{S}$ , we use the SIFT descriptor [2] which is concatenated to the color name descriptor  $\mathbf{K}$  according to:

$$\mathbf{B} = \left( \hat{\mathbf{F}}, \hat{\mathbf{K}} \right) \quad (3)$$

where  $\hat{\cdot}$  indicates that the vector is normalized. The visual

| data set    | soccer |       |               | flowers |       |               |
|-------------|--------|-------|---------------|---------|-------|---------------|
|             | color  | shape | color & shape | color   | shape | color & shape |
| HSV-SIFT    | -      | -     | 77            | -       | -     | 78            |
| hue         | 75     | 58    | 84            | 40      | 65    | 79            |
| opponent    | 75     |       | 85            | 39      |       | 79            |
| color names | 86     |       | 89            | 57      |       | 81            |

**Table 2.** Classification rates on soccer and flower data set for hue, opponent color derivative, HSV-sift and color names. The results are given for only color, only shape (SIFT) and the combination of color and shape.

words are learned in this combined shape-color space and can be thought to contain red corners on a black background, blue blobs on a yellow background, etc.

#### 5. RESULTS

We compare the color name descriptor to the following descriptors from literature: the hue descriptor and the opponent color derivative descriptor proposed in [4], and the HSV-SIFT descriptor proposed in [6]. The hue and opponent derivative descriptor describe the patch with a color histogram of 36 bins. The HSV-SIFT descriptor is constructed by subsequently applying the SIFT descriptor to the H, S, V channel after which the three SIFT descriptors, of length 128, are combined to form one descriptor of length 384. The descriptors are tested on two online available<sup>1</sup> data sets: the soccer team set [4] and the flower data set [14].

##### 5.1. Soccer Teams Classification

The soccer data set consists of seven classes with 25 training images and 15 test images per class. The results for only color, only shape and combined color and shape are given in Table 2. The use of color for this classification problem is crucial: the SIFT shape description obtains a performance of 58%. Most striking are the good results in the case of only color; the color name descriptor improves performance by 10% compared to the *hue* and *opponent* descriptor. In a combination with shape the gain is still 5%. The HSV-SIFT descriptor obtains unsatisfactory results for this data set.

The capability to distinguish between the achromatic colors plays an important role in the classification of this soccer team set, e.g. the outfits of AC Milan are black-red, of PSV black-white and of Liverpool red (see Fig.4). The invariant description loses too much discriminative power, as can be seen from the confusion matrix Table. 3. The color name descriptors's capacity to distinguish black from white in combination with its photometric robustness proved fruitful.

<sup>1</sup>The soccer set at <http://lear.inrialpes.fr/data> and the flower set at <http://www.robots.ox.ac.uk/vgg/research/flowers/>.

| hue  | R         | R-BL      | R-W       | W         | W-BL      | B  | B-P |
|------|-----------|-----------|-----------|-----------|-----------|----|-----|
| R    | 67        | <u>20</u> | <u>13</u> |           |           |    |     |
| R-BL | <u>13</u> | 73        | <u>7</u>  | 7         |           |    |     |
| R-W  | <u>13</u> | <u>13</u> | 73        |           |           |    |     |
| W    |           |           |           | 73        | <u>27</u> |    |     |
| W-BL |           |           |           | <u>20</u> | 73        |    | 7   |
| B    |           |           | 7         |           |           | 73 | 20  |
| B-P  |           | 7         |           |           |           |    | 93  |

| CN   | R   | R-BL      | R-W      | W  | W-BL      | B  | B-P |
|------|-----|-----------|----------|----|-----------|----|-----|
| R    | 100 |           |          |    |           |    |     |
| R-BL |     | 93        |          | 7  |           |    |     |
| R-W  |     | <u>13</u> | <u>7</u> | 67 | 7         |    | 7   |
| W    |     |           |          | 87 | <u>13</u> |    |     |
| W-BL |     |           | 7        |    | <u>7</u>  | 87 |     |
| B    |     |           | 7        | 7  |           | 67 | 20  |
| B-P  |     |           |          |    |           |    | 100 |

**Table 3.** Confusion Matrices for soccer data. Left: hue based color descriptor. Right: color name based descriptor. The soccer teams are abbreviated with the colors of their outfit: Liverpool (Red), AC Milan (Red-Black), PSV (Red-White), Madrid (White), Juventus (White-Black), Chelsea (Blue), Barcelona (Blue-Purple). Note the drop in black and white related confusions as indicated by the underlined numbers.

## 5.2. Flower Classification

The flower data set consists of 17 classes varying in color, texture, and shape [14]. Of each class 40 train and 40 test images exist. The results are summarized in Table 2. For color alone, the color name descriptor obtains significantly better results than the full photometric descriptors. For the combined color and shape description, the performance of the color name descriptor is still slightly better than the other methods.

## 6. CONCLUSIONS AND DISCUSSION

The research effort to enrich local image description with color information has been primarily focused on appending photometric invariant information to shape descriptors. In this paper, we show that full photometric invariance is not optimal due to the loss of discriminative power. The proposed descriptor based on color names outperforms the photometric invariant representations. This indicates that the loss of photometric invariance in going from photometric invariants to color names is more than compensated by a gain in discriminative power. However, we do not believe color names to provide the optimal balance between the two, and from this perspective we see this work as a motivation to further investigate color descriptors with only partial photometric invariance. Nevertheless, results show that the proposed descriptor significantly outperforms existing descriptors for description based on color alone, and does moderately improve description based on color and shape.

## 7. REFERENCES

[1] K. Mikolajczyk and C. Schmid, “Scale and affine invariant interest point detectors,” *International Journal of Computer Vision*, vol. 60, no. 1, pp. 62–86, 2004.



**Fig. 4.** Example images of the soccer team and flower data base. First row: AC Milan, Liverpool, PSV, and Juventus. Second row: daffodil, snowdrop, bluebell, and crocus.

[2] D.G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[3] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[4] J. van de Weijer and C. Schmid, “Coloring local feature extraction,” in *Proc. of the European Conference on Computer Vision*, Graz, Austria, 2006, vol. 2, pp. 334–348.

[5] J.M. Geusebroek, “Compact object descriptors from local colour invariant histograms,” in *British Machine Vision Conference*, 2006.

[6] A. Bosch, A. Zisserman, and X. Munoz, “Scene classification via pLSA,” in *ProcP of the European Conference on Computer Vision*, 2006.

[7] C.L. Hardin and L. Maffi, Eds., *Color Categories in Thought and Language*, Cambridge University Press, 1997.

[8] A. Mojsilovic, “A computational model for color naming and describing color composition of images,” *IEEE Transactions on Image Processing*, vol. 14, no. 5, pp. 690–699, 2005.

[9] R. Benavente, M. Vanrell, and R. Bladrich, “A data set for fuzzy colour naming,” *COLOR research and application*, vol. 31, no. 1, pp. 48–56, 2006.

[10] Y. Liu, D. Zhang, G. Lu, and W-Y. Ma, “Region-based image retrieval with high-level semantic color names,” in *Proc. 11th Int. Conf. on Multimedia Modelling*, 2005.

[11] B. Berlin and P. Kay, *Basic color terms: their universality and evolution*, Berkeley: University of California, 1969.

[12] J. van de Weijer, C. Schmid, and J. Verbeek, “Learning color names from real-world images,” in *Proc. of the Computer Vision and Pattern Recognition*, Minneapolis, Minnesota, USA, 2007.

[13] J. Willamowski, D. Arregui, G. Csurka, C. R. Dance, and L. Fan, “Categorizing nine visual classes using local appearance descriptors,” in *IWLAVS*, 2004.

[14] M-E. Nilsback and A. Zisserman, “A visual vocabulary for flower classification,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.