DISCRIMINATIVE REGIONAL COLOR CO-OCCURRENCE DESCRIPTOR

Qin Zou[1] *Xianbiao Qi*[2] *Qingquan Li*[*,2] *Song Wang*[3]

[1]School of Computer Science, Wuhan University, P.R. China
[2]Shenzhen Key Laboratory of Spatial Smart Sensing and Service,
Shenzhen University, P.R. China
[3]Department of Computer Science and Engineering, University of South Carolina, USA
qzou@whu.edu.cn, qixianbiao@gmail.com, liqq@szu.edu.cn, songwang@cec.sc.edu

ABSTRACT

Traditional color feature descriptors are focused on color-value distributions in the color space, e.g., color histograms, color bag-of-words, which ignore the spatial location and contextual information of different colors. In this paper, a new regional color co-occurrence feature descriptor (RCC) is proposed to reflect spatial relations of colors in an image. First, we partition an image into a number of disjoint regions using superpixel techniques. Then, we construct a color histogram for each region, based on which we construct a color co-occurrence matrix for each pair of neighboring regions. Finally, all the constructed co-occurrence matrices from an image are summed up and normalized as a color descriptor to represent this image. This new color descriptor reflects the color-collocation patterns in the image. We use this new color descriptor for image/object classification and find that it leads to higher classification accuracies than other competing color descriptors.

Index Terms— color descriptor, co-occurrence feature, color collocation, image classification, object recognition.

1. INTRODUCTION

As in human vision, color plays a critical role in many computer-vision applications, such as image classification, scene recognition and object detection [1, 2, 3, 4, 5]. Typically, we need to first represent the color features of an image by a fixed-dimensional color descriptor and then feed this descriptor into a classifier that is trained for a specific vision application. The most widely used color descriptor is color histogram, which reflects the frequency of each color value in the image. By ignoring the spatial location and contextual information of the colors [6], color-histogram based descriptors may not be sufficiently discriminative for accurate image/object classification.

In practice, by considering the contextual information, the color collocation pattern can be taken as a more discriminative feature for object/image classification. For example, pandas' fur contains collocated black and white regions, sunflowers bloom into yellow petals around a brown core, etc. In this paper we analyze the collocation pattern of the colors in an image and develop an algorithm to quantize such pattern into a color descriptor.

Most closely related to our work is the color-occurrence descriptor developed in [7, 8], where color occurrence is derived for each pair of adjacent pixels in an image and the resulting color-occurrence descriptor is used for sliding-window based object detection. Pixel-level color-occurrence analysis is sensitive to image noise and only capture the collocation pattern in the finest scale (pixel level). In this paper, we perform color and color-occurrence analysis on a smaller set of superpixel regions, instead of individual pixels. This not only improves the robustness against the image noise but also captures the color collocation pattern in a higher scale. In the experiments, we justify this point by taking [7, 8] into comparison.

The remainder of this paper is organized as follows. Section 2 briefly reviews the related work. Section 3 introduces a framework for analyzing the regional feature co-occurrence. In this framework, Section 4 describes in detail the proposed regional color co-occurrence descriptor. Section 5 reports the experiment results, followed by a brief conclusion in Section 6.

2. RELATED WORK

Color descriptors In the past two decades, a number of color descriptors have been proposed in the context of image classification and object detection. Most of them are patch-based local color descriptors [9, 10, 11, 12], which are often used in bag-of-visual-words (BoVW) approaches for image representation. In [9], two popular local color descriptors, op-

This research is supported by the National Basic Research Program of China (973 Program) under grant No. 2012CB725303, the National Natural Science Foundation of China (NSFC) under grant No. 61301277 and No. 41371431, the China Postdoctoral Science Foundation funded project under grant No. 2012M521472, and the Hubei Provincial Natural Science Foundation under grant No. 2013CFB299.
*Corresponding author: Dr. Q.Q. Li.

ponent histogram [2] and hue histogram, are proposed and experiments show that these two local color descriptor are insensitive to photometric changes and image quality variation. Based on the progress in linguistics, a new patch-based color descriptor, called Color Names, was proposed in [10]. In [11], a learning-based method is applied to partition the color space into eleven regions, each of which is associated to a basic color name in English language. For each image patch, the color descriptor is defined by the occurrence of each color name in this patch. Color Names was reported to have good photometric invariance in image classification tasks [13, 14]. To seek a balance between the photometric invariance and the discriminative power, Khan et al [12] improved Color Names by using an information-theory based approach to partition the color space. This resulted in a more discriminative color descriptor.

Color co-occurrence descriptor Color co-occurrence descriptors have been studied in [7, 8, 15]. In [8], a color co-occurrence histogram, called Color Correlograms, was computed and applied to image indexing. In [7], color co-occurrence histogram was applied to object detection, by examining the color descriptors in each sliding window. In [15], color co-occurrence histogram was combined with other feature co-occurrence histograms, e.g., CoHoG and CoHED, for an improved image classification performance. As mentioned earlier, these color co-occurrence histograms are pixels based. In this paper, we perform color and color-occurrence analysis on superpixel regions, instead of individual pixels and the proposed color co-occurrence descriptor can capture the color collocation pattern in a much higher scale.

3. A FRAMEWORK FOR REGIONAL FEATURE CO-OCCURRENCE DESCRIPTION

In this section, we introduce the general framework for regional feature co-occurrence description, where region-based features in different regions are analyzed and related based on the region adjacency. In the next section, we will use this framework to develop the proposed color co-occurrence descriptor.

Let $Z = \{z_a | a = 1, 2..., n\}$ be a set of n regions, where for $\forall(a, b)$, $a \neq b$, $z_a \cap z_b = \emptyset$. Let $Q = \{q_i | i = 1, 2..., N\}$ be the set of N possible feature values. For each region z_a, we compute the frequency of every feature value and construct a feature histogram H^a, $a = 1, 2, ..., n$, i.e.,

$$H^a = (H^a_1, H^a_2, ..., H^a_N), \tag{1}$$

where H^a_i is the number of occurrence of feature value q_i in the region z_a.

In this paper, we define the (normalized) feature co-occurrence matrix between region z_a and z_b as

$$\Psi^{a,b} = \mathbf{Norm}\{(H^a)^\mathsf{T} \cdot H^b\}, \tag{2}$$

where **Norm**{} denotes an L_1 normalization to a $N \times N$ matrix such that the summation of the all the matrix element is one. From this definition, the ij-th element in matrix $\Psi^{a,b}$ is product of the frequency of q_i in region z_a and the frequency of q_j in region z_b. This product reflects the co-occurrence of feature value q_i in region z_a and q_j in region z_b.

4. RCC: REGIONAL COLOR CO-OCCURRENCE DESCRIPTOR

In this section, we propose a new regional color co-occurrence descriptor (RCC) by following the framework introduced in Section 3. As detailed below, the computation of RCC from an image is made up of three main steps, i.e., region partitioning, color codebook construction, and regional color co-occurrence matrix calculation.

4.1. Region Partitioning

To compute the regional color co-occurrence matrix for an image, we first partition the image into a number of regions. For better color co-occurrence description, neighboring pixels that are similar in color and homogenous in appearance are desired to be grouped into a same region after the partitioning. In this paper, we use SLIC, a superpixel technique [16] for region partitioning, as it can generate similar-size homogeneous-appearance regions.

4.2. Color Codebook Construction

The goal of construct a color codebook is to limit the number of possible feature values, which are the representative color values in this work. To seek a balance between discriminativeness and generalizablity, we construct the color codebook by the following two steps. First, we collect candidate colors from a collection of images, by taking the average color of each superpixel region as a candidate color. Second, we apply the k-means clustering to the candidate colors, and produce color codes. Specifically, the k-means is initialized with $numCluster$ centers. k-means clustering will group similar color candidate into one same cluster, and thus makes the cluster centers be more representative than the color candidates.

4.3. Regional Color Co-occurrence Matrix Calculation

The color codebook provides to represent the image compactly. Based on the color codebook, we can calculate the regional color co-occurrence matrix for each image. There are mainly four steps, as listed below.

- **colorMapping** Colors in the original image are mapped into color values in the color codebook. For each pixel, it is assigned a color value in the color codebook that is closest to its original color value, where Euclid distance is the measuring metric. In the section of experiment, we will study how the size of the codebook influence the final performance.

- **colorRegionHist** With the color image produced by colorMapping, a color histogram is constructed for each superpixel region. This can be simply achieved by counting the occurrence of each color code. The dimension of the histogram is equal to the size of the color codebook.

- **searchNeighbors** This is to get the neighboring regions of each superpixel region. In this step, two regions sharing a part of their boundaries are counted as neighboring regions to each other.

- **calcRCCMatrix** For each two neighboring superpixel regions, the regional color co-occurrence matrix is calculated by using Eq. 2.

For an image I, an RCC matrix Ψ_I can be produced through the above steps. To further enhance the generalization power, we make it a symmetric matrix and take its upper triangle matrix as the final RCC matrix, as illustrated by Eq. 3,

$$\hat{\Psi}_I = \mathbf{Triu}(\Psi_I^{\mathsf{T}} + \Psi_I). \tag{3}$$

where **Triu**() denotes taking the upper triangle of the matrix. The final RCC matrix $\hat{\Psi}_I$ is taken as the final RCC descriptor for image representation. Note that, $\hat{\Psi}_I$ holds a dimension of $\frac{N(N+1)}{2}$, where N denotes the size of the color codebook.

5. EXPERIMENTS AND RESULTS

5.1. Datasets

The proposed RCC descriptor[1] is evaluated on five diverse datasets: Flowers102 [17], Birds200 [18], Pascal'07 [19], Dogs120 [20], and PFID61 [21].

Flowers102 The Oxford Flowers 102 dataset contains 8189 images of 102 flower categories. It has been proved that foreground masks can boost the classification performance [22, 23]. To focus on color feature description, similar to [24], we apply Grabcut [25] to produce foreground masks. In the experiments, 20 images each category are randomly selected for training and the rest for testing. The procedure is repeated 30 times and the results are averaged.

Birds200 The Caltech-UCSD Birds 200-2010 [18] consists of 6033 bird images which are from 200 kinds of birds. The training and testing samples have been fixed by the dataset. We use the coarse foreground provided by the dataset for color feature extraction.

[1] Codes are available at https://sites.google.com/site/qinzoucn/documents

Pascal'07 The PASCAL VOC Challenge 2007 [19] dataset contains 9963 images of 20 different object categories. Each image is provided bounding boxes indicating one or more objects within the image. The dataset is divided into a predefined training set (5011 images) and testing set (4952 images).

Dogs120 The Stanford Dogs dataset [20] contains 120 dog species and 20580 images, where bounding boxes are provided. There are 100 images per class for training and about 80 images for testing. We use fixed training and testing samples which have been specified along the dataset.

PFID61 The Pittsburgh Food Image Dataset [21] contains 1098 fast-food images from 61 food categories, with masked foreground. The experimental protocol in [26, 27] is used.

5.2. The effectiveness of RCC

To demonstrate the effectiveness of the proposed RCC for image/object classification, we compare RCC with other state-of-the-art color descriptors, using only the color information. Exactly, five different color descriptors are considered for comparison, i.e., normalized RGB (rg histogram) [28], hue histogram (HH) [9], Color Names (CN) [11], the discriminative color descriptor (DD) [12], and the color co-occurrence descriptor (CC) [7, 8]. We evaluate the CC and the proposed RCC with three settings, namely using 11, 25, and 128 color codes, and non-linear SVM using the χ^2 kernel for classification. Results are obtained with a specialized version of each color descriptors. Table 1 shows the performance of each color descriptor on three datasets, in terms of classification accuracy (or mean average precision for Pascal'07). Note that, we directly use results reported in [12] for the first four competing color descriptors. As can be seen from Table 1, for the case of 11 color codes, RCC achieves higher results on Pascal'07 and Birds200, but has slightly lower results than DD(11) on Flowers102. For the case of 25 color codes, RCC outperforms all the other descriptors, with largely improved results. Exactly, RCC gets a further improved performance at a color-codes number of 128, that are 53.3, 16.9, and 25.9 on

Table 1. Image classification performances using only color descriptors.

Descriptor	Flowers102	Birds200	Pascal'07
rg	38.6	4.3	10.6
HH	32.8	3.5	10.1
CN	40.2	7.7	11.6
DD(11)	43.7	8.0	12.2
DD(25)	47.0	8.7	12.6
CC(11)	41.4	6.0	13.5
CC(25)	47.6	11.3	16.9
RCC(11)	42.3	8.4	17.5
RCC(25)	50.7	12.5	19.7

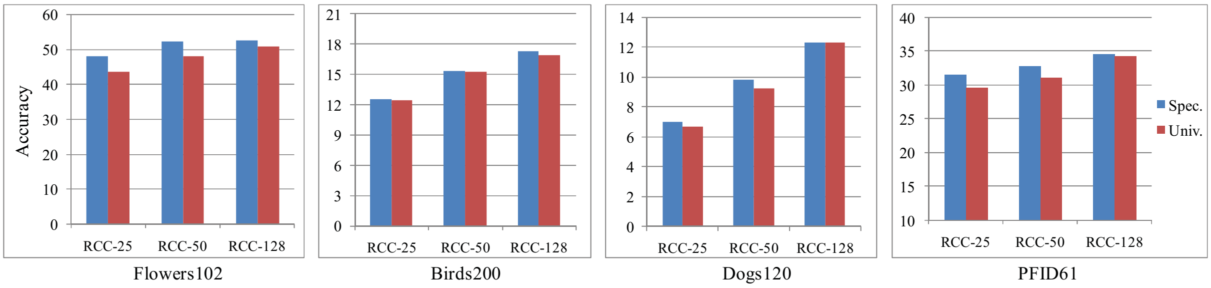

Fig. 1. Comparison of RCC's performance using specialized color codes and universal color codes on four datasets. Note that, RCC-25, RCC-50 and RCC-128 denote RCC with 25, 50, and 128 color codes in the color codebook, respectively.

Table 2. Evaluation of RCC using color only and using color+shape, with universal color codes

Method	Flowers102		Birds200		Dogs120		PFID61	
Shape alone	73.7		16.4		39.3		43.8	
	color only	color+shape	color only	color+shape	color only	color+shape	color only	color+shape
RCC(25)	43.35	78.52	12.43	24.39	6.70	41.34	29.69	45.36
RCC(50)	47.90	80.77	15.24	26.04	9.22	42.06	31.06	46.63
RCC(128)	50.86	81.41	16.34	26.05	12.28	42.87	34.34	47.91

the three datasets, respectively.

5.3. The universality of RCC

There are two kinds of sources that color codes can be drawn from, one is from a specific dataset, and the other is from one/more universal datasets. In the former, color codes are specified as specialized color codes, which are only used for the specific dataset where they are from. While in the later, color codes are specified as universal color codes, which can be used for any other dataset. Proper universal color codes would bring a great convenience, and thus universality is a desired property [29].

We make comparisons to RCC using specialized color codes and universal color codes on the above four datasets. For RCC(25), RCC(50) and RCC(128), universal color codes are produced on the Pascal'07 dataset with the color codebook construction approach as presented in Section 4.2. Since Pascal'07 provides the source images, we keep it out from evaluation. The region size for superpixel segmentation in RCC is kept constant as 100, for all datasets. In the classification step, a same non-linear SVM using χ^2 kernel is applied to all descriptors. It can be seen from Fig. 1, for each dataset and each descriptor, the performance using specialized color codes is slightly higher than that using universal color codes. The difference in performance becomes smaller when the number of the color codes increases from 25 to 128. It can be observed that, at a color-codes number of 128, the performances of RCC using universal color codes are very close to that using specialized color codes, which indicates a good universality property of RCC.

5.4. The additivity of RCC

We evaluate the additivity of the proposed RCC by combining it with shape description in image classification. For shape, we apply multi-scale dense SIFT for shape feature description, and Fisher Vector (FV) [30] as the coding strategy. The FV is equipped with 256 Gaussian models, and the feature is reduced to 70 dimension by PCA. For RCC, we extract universal color codes from the Pascal'07 dataset, with 25, 50, and 128 codes, respectively. Table 2 gives results for each dataset. We can see, for each dataset, improvement can be gained by combining RCC with shape descriptor, which demonstrates a good additivity property of RCC.

6. CONCLUSION

In this paper, we developed a new regional color co-occurrence descriptor (RCC) and used it to image and object classification. By modeling the color distribution within each superpixel region and the color co-distribution between each pair of neighboring superpixel regions, RCC reflects the color collocation patterns of an image in a large scale. By quantizing the image color into a reasonable-size codebook, RCC can gain a balance between the discriminativeness and the generalizablity. We conducted experiments on five popular benchmark datasets and found that the proposed RCC leads to higher image classification accuracies than other state-of-the-art color descriptors. Experiments also showed that RCC can be combined with shape descriptors to boost the accuracy of image classification. By constructing codebook on one benchmark and testing on different benchmarks, we also found that the proposed RCC shows excellent universality.

7. REFERENCES

[1] B. Funt and G. Finlayson, "Color constant color indexing," *IEEE PAMI*, vol. 17, pp. 522–529, 1995.

[2] Theo Gevers and Arnold W.M. Smeulders, "Color-based object recognition," *Pattern Recognition*, vol. 32, pp. 453–464, 1999.

[3] Theo Gevers and Harro Stokman, "Robust histogram construction from color invariants for object recognition," *IEEE TPAMI*, vol. 26, no. 1, pp. 113–118, 2004.

[4] 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 *CVPR*, 2012.

[5] Fahad Shahbaz Khan, Joost van de Weijer, and Maria Vanrell, "Modulating shape features by color attention for object recognition," in *IJCV*, 2012.

[6] Hongxing Wang, Junsong Yuan, and Ying Wu, "Context-aware discovery of visual co-occurrence patterns," *IEEE TIP*, 2014.

[7] Peng Chang and John Krumm, "Object recognition with color cooccurrence histograms," in *ICCV*, 1999.

[8] J. Huang, S.R. Kumar, M. Mitra, W.J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *CVPR*, 1997.

[9] Joost van de Weijer and Cordelia Schmid, "Coloring local feature extraction," in *ECCV*, 2006.

[10] Robert Benavente, Maria Vanrell, and Ramon Baldrich, "Parametric fuzzy sets for automatic color naming," *Journal of the Optical Society of America*, vol. 25, no. 10, pp. 2582–2593, 2008.

[11] Joost van de Weijer, Cordelia Schmid, Jakob Verbeek, and Diane Larlus, "Learning color names for real-world applications," *IEEE TIP*, vol. 18, no. 7, pp. 1512–1524, 2009.

[12] Rahat Khan, Joost van de Weijer, Fahad Khan, Damien Muselet, Christophe Ducottet, and Cecile Barat, "Discriminative color descriptors," in *CVPR*, 2013.

[13] F. Khan, R. Anwer, J. Van de Weijer, A. Bagdanov, M. Vanrell, and A. Lopez, "Color attributes for object detection," in *CVPR*, 2012.

[14] F. S. Khan, J. van de Weijer, and M. Vanrell, "Modulating shape features by color attention for object recognition," *IJCV*, vol. 98, pp. 49–64, 2012.

[15] S. Ito and S. Kubota, "Object classification using heterogeneous co-occurrence features," in *ECCV*, 2010.

[16] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE TPAMI*, vol. 34, no. 11, pp. 2274–2282, 2012.

[17] M.-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in *Indian Conference on CVGIP*, 2008.

[18] P. Welinder, S. Branson, T. Mita, C. Wah, F. Schroff, S. Belongie, and P. Perona, "Caltech-UCSD Birds 200," in *Technical Report CNS-TR-2010-001, Caltech*, 2010.

[19] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes challenge 2007 results," 2007.

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR*, 2009.

[21] M. Chen, K. Dhingra, W. Wu, L. Yang, R. Sukthankar, and J. Yang, "PFID: Pittsburgh fast-food image dataset," in *ICIP*, 2009.

[22] Yuning Chai, Victor Lempitsky, and Andrew Zisserman, "BiCoS: a bi-level co-segmentation method for image classification," in *ICCV*, 2011.

[23] Yuning Chai, Esa Rahtu, Victor Lempitsky, Luc Van Gool, and Andrew Zisserman, "TriCoS: a tri-level class-discriminative co-segmentation method for image classification," in *ECCV*, 2012.

[24] Xianbiao Qi, Rong Xiao, Jun Guo, and Lei Zhang, "Pairwise rotation invariant co-occurrence local binary pattern," in *ECCV*, 2012.

[25] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: interactive foreground extraction using iterated graph cuts," *ACM ToG*, vol. 23, no. 3, pp. 309–314, 2004.

[26] Shulin Yang, Mei Chen, Dean Pomerleau, and Rahul Sukthankar, "Food recognition using statistics of pairwise local features," in *CVPR*, 2010.

[27] Bingbing Ni, Mengdi Xu, Jinhui Tang, Shuicheng Yan, and Pierre Moulin, "Omni-range spatial contexts for visual classification," in *CVPR*, 2012.

[28] K. Sande, T. Gevers, and C. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE TPAMI*, vol. 32, no. 9, pp. 1582–1596, 2010.

[29] A. Torralba and A. Efros, "Unbiased look at dataset bias," in *CVPR*, 2011.

[30] Florent Perronnin, Jorge Sánchez, and Thomas Mensink, "Improving the fisher kernel for large-scale image classification," in *ECCV*, 2010.