IMAGE RETRIEVAL BY SUBSPACE-PROJECTED COLOR AND TEXTURE FEATURES

Weidi Liu⋆ *Wei Li*⋆ *Yan Huang*† *Jingliang Peng*⋆

⋆ School of Computer Science and Technology, Shandong University, Jinan, Shandong, 250101, China
† Shenzhen Realis Multimedia Technology Co., Ltd, Shenzhen, Guangdong, 518000, China

ABSTRACT

In this work, we propose to retrieve images according to a subspace-projected color and texture (SuPCAT) feature descriptor. Firstly, we propose a novel contrast and color distribution (CoCD) descriptor to characterize the pixels' colors in an image. Thereafter, we combine the proposed CoCD color feature descriptor with four effective texture feature descriptors and conduct subspace projection on the combined feature descriptor for reduced dimensionality and increased discriminativeness, which leads to the proposed SuPCAT feature descriptor. By integrating the SuPCAT feature descriptor with a basic Euclidean distance metric, we construct an image retrieval scheme and conduct experiments to demontrate its outstanding performance.

Index Terms— image retrieval, contrast and color distribution (CoCD), subspace projected color and texture (SuPCAT)

1. INTRODUCTION

Two-dimensional digital images (or images for short) have been widely used in numerous fields. Creation and distribution of such data have also become easier and easier, especially considering the picturing capability and wireless network access of commodity mobile devices. Naturally, it has been an important while challenging task to precisely retrieve desired images from a vast pool for a specific query.

There have been text based and content based approaches to image retrieval. In a text based image retrieval system, database images are often manually annotated by text descriptors. This system gives a relevant set of images which are matched to the annotation of a query image. The text based approach requires tedious annotation effort and the text annotation is often not sufficiently expressive. To address these issues, content-based image retrieval (CBIR) has been intensively researched. In a CBIR system, images are usually indexed according to low-level features (*e.g.*, color and texture features) that are automatically extracted from the images, and the image retrieval is based on the matching of these image features. In this paper, we focus on CBIR for its wider applicability and higher effectiveness.

Authors' email addresses: lwd0812@foxmail.com (W. Liu), liweilw813@mail.sdu.edu.cn (W. Li), huang.yan74@gmail.com (Y. Huang) and jingliap@gmail.com (J. Peng). J. Peng is the corresponding author.

1.1. Related Work

The field of CBIR has drawn intensive attention in the past two decades [1, 2, 3]. For the feature description, color and texture analyses have been widely used. Color is the most widely used visual information which is robust to scale, orientation, resolution and noise. The traditional color histogram has been applied well in CBIR systems, which, however, has difficulty in characterizing the spatial information. Therefore, several color descriptors have been proposed to exploit spatial structures in images as well, including the dominant color descriptor (DCD) [4], the color layout descriptor (CLD) [5] and the scalable color descriptor (SCD) [6]. Texture provides another important type of information for image depiction. Various ways have been proposed to describe the texture feature of an image, *e.g.*, the local binary pattern (LBP) [7], the Gabor filter [8], the histogram of oriented gradients (HOG) [9], the GIST feature [10] and the gray level co-occurrence matrix (GLCM) [11].

Further, researchers have proposed to use multiple features for the image description. Ahmad *et al.* [12] introduced a saliency structure histogram (SSH) to describe colors and edge orientations for each image. Liu and Yang [13] proposed the color difference histogram (CDH) which integrates the advantages of co-occurrence histogram. Besides, the structure element descriptor (SED) [14], the micro-structure descriptor (MSD) [15] and the multi-texton histogram (MTH) [16] all take multiple features into account. These methods combine separate descriptors into one long descriptor, which results in improved retrieval precision but increases the computing complexity. Nevertheless, they have made limited efforts to choose the most discriminative dimensions of the features used.

1.2. Contributions

In this work, we contribute mainly in feature description for image retrieval. We firstly propose a novel contrast and color distribution (CoCD) feature descriptor, secondly choose

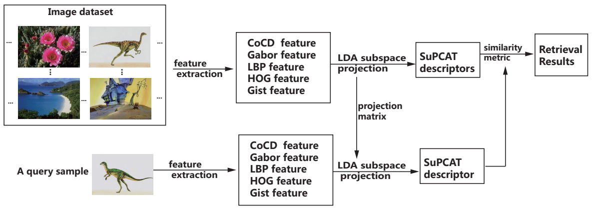

Fig. 1. Flowchart of the proposed image retrieval system.

a descriptive set of texture feature descriptors to combine with the CoCD descriptor, and thirdly find the most discriminative dimensions of the combined feature descriptors through subspace projection, leading to the proposed subspace-projected color and texture (SuPCAT) feature descriptor. We then construct an image retrieval scheme based on the SuPCAT feature descriptor, which significantly outperforms the state-of-the-art according to our experiments.

As shown by the flowchart in Fig. 1, the proposed scheme is learning-based. In the training stage, a discriminative subspace projection matrix is learned from the training samples with the combined feature descriptors, and the SuPCAT feature descriptors are computed for all the images in the database. In the query stage, the SuPCAT feature descriptor is computed for each query sample, which is matched against those in the database according to a specific distance metric (Euclidean distance metric in this work).

2. SuPCAT DESCRIPTOR

We propose a novel CoCD descriptor to characterize the pixels' colors of an image, choose a set of texture feature descriptors and combine them with the CoCD descriptor to form a long descriptor. Thereafter, we learn a subspace projection from the training samples to extract the most discriminative dimensions of the combined feature descriptor. The learned projection is then applied to construct the SuPCAT feature descriptor for every image.

2.1. CoCD Descriptor

Inspired by the CoLD (contrast and luminance distribution) descriptor [2] that characterizes the pixels' luminances in an image, we propose the CoCD (contrast and color distribution) descriptor to characterize the full colors of the pixels in an image.

As proposed in the reference [2], the CoLD descriptor for an image is a feature vector, F, defined as

$$F = (d_0, d_1, \ldots, d_{n-1}, h_0, h_1, \ldots, h_{n-1}) \tag{1}$$

where $n-1$ indicates the maximum luminance level, d_i is the percentage of pixel pairs with the contrast (luminance difference) of i ($i \in [0, n)$) and h_i is the percentage of pixels of the i-th ($i \in [0, n)$) luminance level. Although luminance information provides important hints for shapes and textures, other color information helps signify specific objects in the image as well. As such, we propose to extend the CoLD descriptor by considering full color information in the description, as detailed below.

We base our feature description on the HSV color model, as it closely simulates humans' perception of colors and has the ability to separate chromatic and achromatic components [17]. In this color model, each color is represented by three components, *i.e.*, hue (H), saturation (S) and value (V), and the formulae to convert an RGB color to a HSV color are given below, where $R, G, B \in [0, 255]$ and $t = max(R, G, B)$.

$$\begin{aligned} H &= \begin{cases} \frac{60(G-B)}{t-min(R,G,B)} & if\ t = R \\ 120 + \frac{60(B-R)}{t-min(R,G,B)} & if\ t = G \\ 240 + \frac{60(R-G)}{t-min(R,G,B)} & if\ t = B \end{cases} \\ S &= \begin{cases} \frac{t-min(R,G,B)}{t} & if\ t \neq 0 \\ 0 & otherwise \end{cases} \\ V &= \frac{t}{255} \end{aligned} \tag{2}$$

Further, if $H < 0$, we set $H = H + 360$. As computed above, we have $H \in [0, 360]$, $S \in [0, 1]$ and $V \in [0, 1]$. Thereafter, we quantize the full ranges of S, and V to Q_s and Q_v levels, respectively, and compute one-dimensional feature vector named L [18]: $L = \lfloor (HQ_sQ_v + SQ_s + V)/4 \rfloor$ for quantized S and V of each pixel. In our experiments, we set $Q_s = Q_v = 8$, and the integral L values fall in the range of $[0, 127]$. Using thus computed L values of all the pixels, we construct our CoCD feature descriptor, F, still using the formula in Eq. 1 but with $n-1$ being the upper bound of the L value, d_i the percentage of pixel pairs with the contrast (difference in L) of i ($i \in [0, n)$) and h_i the percentage of pixels with $L = i$ ($i \in [0, n)$).

2.2. Texture Features

Texture features have been widely exploited for various computer vision tasks. Based on our survey, we select four texture feature descriptors, which are abbreviated as HOG [9], GIST [10], Gabor [8] and LBP [7], for our use, as the effectiveness of them has been well proved in the literature. HOG [9] is formed by calculating the gradient direction histograms of local regions in the image. GIST [10] indicates the dominant spatial structure of a scene using spectral and coarsely localized information. Gabor [8] effectively extracts local and spatial information of the target. LBP [7] describes the local texture feature of an image which is invariant to rotation and gray scale. All the four texture features have been proved effective for describing the local and spatial information of images. We combine the proposed CoCD feature descriptor with the selected texture feature descriptors into one feature vector which is denoted as $CFV \in R^d$ ($d = 1011$ in this work) and used as a raw image feature descriptor.

Fig. 2. Thumbnailed image samples from the Corel5k database.

2.3. Subspace Projection via LDA

The combined CFV feature has high dimensions and redundant information, which will influence the efficiency and precision of retrieval. In order to obtain a lower dimensional and more discriminative feature descriptor, we transform the CFV features to a subspace so as to increase the inter-class and reduce the intra-class distances with training samples using the linear discriminant analysis (LDA) [19]. Suppose that the d-dimensional feature vectors, $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n$, belong to C image classes. LDA seeks a transform matrix W that maximizes the ratio of the between-class distance to the within-class distance as given by

$$J(W) = \frac{|W^T S_B W|}{|W^T S_W W|}, \tag{3}$$

where S_B and S_W are the between-class and the within-class scatter matrices, respectively. In essence, the LDA process finds a subspace spanned by $\{\mathbf{w}_i\}_{i=1}^{C-1}$ and projects each feature vector, $\mathbf{f}_i$, onto this subspace to obtain $\mathbf{z}_i$.

The finally projected feature vector for each image is called subspace-projected color and texture feature descriptor (SuPCAT $\in R^k$, k=49) in this work.

3. SIMILARITY METRIC

We employ the basic Euclidean distance metric to form our similarity metric. Specifically, we measure the similarity, $S(X_1, X_2)$, between two k-dimensional feature descriptors, $X_1 = (x_{1,1}, x_{1,2}, \ldots, x_{1,k})$ and $X_2 = (x_{2,1}, x_{2,2}, \ldots, x_{2,k})$, by $S(X_1, X_2) = -D(X_1, X_2)$ with

$$D(X_1, X_2) = \sqrt{\sum_{i=1}^{k} (x_{1,i} - x_{2,i})^2}$$

.

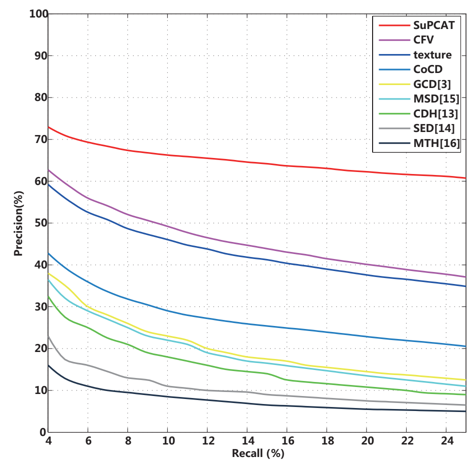

Fig. 3. P-R curves of different methods. The curves for GCD [3], MSD [15], CDH [13], SED [14] and MTH [16] are reconstructed from those in the reference [3].

4. EXPERIMENTS

We experiment on the Corel-5K database which is a baseline database of images ranging from animals and outdoor sports to natural scenes. All images in the database were obtained from Corel Gallery Magic 20, 0000 (8 CDs). The Corel-5k database contains 50 categories covering 5,000 images including diverse content such as fireworks, barks, microscopy images, tiles, food textures, trees, waves, pills and stained glasses. Every category contains 100 images of size 384×256 or 256×384 in the JPEG format. A collage of thumbnailed images from the Corel-5K database, one for each category, is shown in Fig. 2.

In our experiments, we adopt the same experiment setting as the references [3, 23] for fair comparison. Specifically, we randomly choose 10 images from each category as query

Table 1. The mAP statistics of different feature descriptors (%).

Descriptor	GCM [20]	Gabor [8]	LBP [7]	GLCM [11]	Hu [21]	GIST [10]	HOG [9]	CoLD [2]	CoCD
Dimension	81	37	120	16	7	512	128	128	256
mAP	15.00	11.21	13.90	8.44	5.45	16.32	23.21	11.55	14.72

Table 2. The average retrieval precision and recall statistics of different methods. The statistics of TCM [22], MTH [16], MSD [15], SSH [12] and MTSD [23] are taken from the reference [23].

Methods	TCM [22]	MTH [16]	MSD [15]	SSH [12]	MTSD [23]	CoLD[2]	SuPCAT
Precision (%)	33.92	51.84	57.92	61.45	62.98	25.06	67.89
Recall (%)	5.09	7.78	8.69	9.22	9.45	4.01	10.86

images, and evaluate the performance using the precision and recall curves (R-D curves) with precision and recall for the i-th query image defined as follows:

$$P(i, N) = \frac{I_N}{N}$$

$$R(i, N) = \frac{I_N}{M}$$

where I_N is the number of truly relevant images in the N retrieved ones and M is the total number of relevant images in the database.

The mean average precision (mAP) statistics of various feature descriptors are shown in Tab.1. The Euclidean distance is used for all these results. From Tab.1, we observe that the proposed CoCD outperforms most of the other descriptors. In particular, it improves the CoLD descriptor. When evaluated separately, the CoCD descriptor does not work as well as the GIST and the HOG descriptors. It should be noted that, as a color feature descriptor, the CoCD descriptor is complementary to the texture feature descriptors (*e.g.*, HOG and GIST) and promotes the precision of retrieval when added to the texture descriptors, as demonstrated below.

We compare our image retrieval method (formed by the SuPCAT descriptor and the Euclidean distance metric) with several benchmark methods. The R-D curves of these methods are plotted in Fig. 3. From Fig. 3, we observe that our proposed SuPCAT method outperforms all the other ones. Comparing the R-D curves of the CFV (plus the Euclidean distance) method and the textures (plus the Euclidean distance) method, we observe the promotion of performance when the proposed CoCD color feature descriptor is added to the texture feature descriptors (*i.e.*, HOG [9], GIST [10], Gabor [8] and LBP [7]). Comparing the R-D curves of the SuPCAT method and the CFV method, we observe the promotion of performance when the subspace projection is applied.

Further, the average retrieval precision and recall statistics of various benchmark methods are listed in Tab. 2. As shown in Tab. 2, the proposed SuPCAT method outperforms all the other six methods.

5. CONCLUSION

In this work, we have proposed a novel CoCD color feature descriptor, combined it with a selected set of texture feature descriptors, and learned and applied a subspace projection to produce a lower-dimensional and more discriminative SuPCAT feature descriptor. Integrating the SuPCAT feature descriptor with the Euclidean distance metric, we have constructed an image retrieval scheme and demonstrated its outstanding performance through experiments. In the future, we will test applying the subspace projection approach to more image feature descriptors and seek for more effective similarity metrics.

Acknowledgement

This work was funded by NSFC (National Natural Science Foundation of China) grant 61472223.

6. REFERENCES

[1] Bin Xu, Jiajun Bu, Chun Chen, Deng Cai, Xiaofei He, Wei Liu, and Jiebo Luo, "Efficient manifold ranking for image retrieval," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. ACM, 2011, pp. 525–534.

[2] Qingkun Su, Yan Huang, and Jingliang Peng, "Coldimage: Contrast and luminance distribution for content-based image retrieval," in *2011 International Conference on Image Analysis and Signal Processing*. IEEE, 2011, pp. 143–146.

[3] Lin Feng, Jun Wu, Shenglan Liu, and Hongwei Zhang, "Global correlation descriptor: a novel image representation for image retrieval," *Journal of Visual Communication and Image Representation*, vol. 33, pp. 104–114, 2015.

[4] Kanwal Preet Kaur, "On comparative performance analysis of color, edge and texture based histograms for

content based color image retrieval," in *Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), 2014 3rd International Conference on*. IEEE, 2014, pp. 1–6.

[5] Eiji Kasutani and Akio Yamada, "The mpeg-7 color layout descriptor: a compact image feature description for high-speed image/video segment retrieval," in *Image Processing, 2001. Proceedings. 2001 International Conference on*. IEEE, 2001, vol. 1, pp. 674–677.

[6] Jan Theeuwes, "Visual selective attention: A theoretical analysis," *Acta psychologica*, vol. 83, no. 2, pp. 93–154, 1993.

[7] Timo Ojala, Matti Pietikainen, and Topi Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[8] Zhi-Chun Huang, Patrick PK Chan, Wing WY Ng, and Daniel S Yeung, "Content-based image retrieval using color moment and gabor texture feature," in *2010 International Conference on Machine Learning and Cybernetics*. IEEE, 2010, vol. 2, pp. 719–724.

[9] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. IEEE, 2005, vol. 1, pp. 886–893.

[10] Matthijs Douze, Hervé Jégou, Harsimrat Sandhawalia, Laurent Amsaleg, and Cordelia Schmid, "Evaluation of gist descriptors for web-scale image search," in *Proceedings of the ACM International Conference on Image and Video Retrieval*. ACM, 2009, p. 19.

[11] Robert M Haralick, Karthikeyan Shanmugam, et al., "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, , no. 6, pp. 610–621, 1973.

[12] Jamil Ahmad, Muhammad Sajjad, Irfan Mehmood, and Sung Wook Baik, "Ssh: Salient structures histogram for content based image retrieval," in *Network-Based Information Systems (NBiS), 2015 18th International Conference on*. IEEE, 2015, pp. 212–217.

[13] Guang-Hai Liu and Jing-Yu Yang, "Content-based image retrieval using color difference histogram," *Pattern Recognition*, vol. 46, no. 1, pp. 188–198, 2013.

[14] Xingyuan Wang and Zongyu Wang, "A novel method for image retrieval based on structure elements descriptor," *Journal of Visual Communication and Image Representation*, vol. 24, no. 1, pp. 63–74, 2013.

[15] Guang-Hai Liu, Zuo-Yong Li, Lei Zhang, and Yong Xu, "Image retrieval based on micro-structure descriptor," *Pattern Recognition*, vol. 44, no. 9, pp. 2123–2133, 2011.

[16] Guang-Hai Liu, Lei Zhang, Ying-Kun Hou, Zuo-Yong Li, and Jing-Yu Yang, "Image retrieval based on multi-texton histogram," *Pattern Recognition*, vol. 43, no. 7, pp. 2380–2389, 2010.

[17] Wen Chen, Yun Q Shi, and Guorong Xuan, "Identifying computer graphics using hsv color model and statistical moments of characteristic functions," in *Multimedia and Expo, 2007 IEEE International Conference on*. IEEE, 2007, pp. 1123–1126.

[18] Mai-ling SONG and Huan LI, "An image retrieval technology based on hsv color space [j]," *Computer Knowledge and Technology (Academic Exchange)*, vol. 1, 2007.

[19] Keinosuke Fukunaga, *Introduction to statistical pattern recognition*, Academic press, 2013.

[20] Guang J Zhang and Norman A McFarlane, "Sensitivity of climate simulations to the parameterization of cumulus convection in the canadian climate centre general circulation model," *Atmosphere-ocean*, vol. 33, no. 3, pp. 407–446, 1995.

[21] Qing Chen, Emil Petriu, and Xiaoli Yang, "A comparative study of fourier descriptors and hu's seven moment invariants for image recognition," vol. 1, pp. 103–106, 2004.

[22] Tiezheng Ge, Kaiming He, Qifa Ke, and Jian Sun, "Optimized product quantization," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 4, pp. 744–755, 2014.

[23] Meng Zhao, Huaxiang Zhang, and Jiande Sun, "A novel image retrieval method based on multi-trend structure descriptor," *Journal of Visual Communication and Image Representation*, vol. 38, pp. 73–81, 2016.