Flower Image Retrieval with Category Attributes

Lin Li*, Yu Qiao*†
*Shenzhen Key Lab of Comp. Vision and Patt. Recog., Shenzhen Institute of Advanced Technology, CAS, P.R.China
†The Chinese University of Hong Kong, Hong Kong
{lin.li, yu.qiao}@siat.ac.cn

***Abstract*—In this paper, we recommend a mid-level representation, category attributes, for content based flower image retrieval. Low-level features have been utilized in images for retrieval. However, even though these features are efficient, the similarity between low-level features may differ from high level human perception, known as semantic gap. In real life, it is very usual to use attributes, a domain specific terminology, to describe the visual appearance of objects. Inspired by this, we utilize category attributes, i.e. daisy, buttercup or iris to construct semantic representation of flower images. For each category attribute, we train a linear SVM based on low level visual features,containing the appearance of color, texture and shape. Outputs of these classifiers are regarded as attribute features. Then we use distances between attribute features for flower retrieval. This method was evaluated on 17 Category Flower Dataset. Experimental results show that attribute based representation outperforms low-level features in terms of mean average precision.**

***Keywords*—Content Based Image Retrieval, Category Attributes, Linear SVM, 17 Category Flower Dataset**

I. Introduction

This paper proposes a system of content based image retrieval [12] for flowers. Given a query image which has focus on one kind of flower, our system aims to return a set of representative images in which similar types of flower appears. The challenge of this problem comes from many aspects, such as, large intra-class appearance variation of flowers viewpoints and illumination etc.

Most approaches to retrieve images of a particular object rely on the bag-of-features (BOF) representation in [3]. This initial method maps local image descriptors into visual words, for instance, it selects for each descriptor the the nearest descriptor from a visual vocabulary, learnt by k-means on a training set. The number of descriptors assigned to each visual word then constructs a histogram, representing BOF model. Then from 2006, Spatial pyramid in [14] has drawn more attention. This model adds approximate global geometric correspondence to BOF model. It separates an image into sub-regions and computes BOF histograms of local features in those sub-regions. In this paper, spatial pyramid is used to describe SIFT features.

A visual vocabulary for flower classification introduced in [2] demonstrates that it distinguishes one flower from another efficiently. It uses clustered HSV color space as a color cue, SIFT [8] as an expression of shape, and MR8 filter [9] for some information of texture, building a flower vocabulary by combining these features.

However, attribute has a higher semantic information than low-level features. In real life, as we claimed before, attributes are more common to be used to express the visual feature of objects. For example, automatic attribute discovery suggested in [10] provides an automatic way to find and identify attributes (e.g. stiletto, sandal and clutch) from pictures and textual information on the Internet, attribute classifier in [7] indicates it is valid to train bi nary classifiers to identify the presence of visual appearance, such as gender, race and age. In this paper, the category attributes from 17 Category Flower Dataset [13] form 17 Support Vector Machine classifiers for each category.

In the past, low-level feature based image retrieval system chose similarity functions like Euclidean or χ^2 distance on the low-level descriptors - features to compare similarity. In this attribute based image retrieval system, comparing the Euclidean distances between scores of attributes learned from the attribute classifiers is more appropriate.

Our approach has the advantages as follows: (1) Our mid-level representation is robust to viewpoint and illumination changes by training with stable features such as HSV, SIFT, LBP. (2) Compared to low-level features, our attribute based representation is closer to human perception of flowers which allows us to retrieve flower more effectively and efficiently.

Our major contribution is as follows: We present a mid-level representation - category attributes - image retrieval method that systematically outperforms low-level features.

The remainder of this paper is organized as follows. Section II explains how the category attributes are used in a content based image retrieval system. Section III describes the experiments we have performed and their results of attribute based image retrieval in comparison to those of low-level features. We report substantial improvement to the baseline using low-level features. Finally, conclusion of this paper is in Section IV.

II. Our Approach

In our approach, the first step is to extract low-level image features, e.g. SIFT [8], HSV, auto-correlogram [6], color-moments [5], wavelet-moment, edge-detection and Local Binary Pattern [1]. The second step is to train mid-level visual features - category attributes with these low-level features . These attributes are simple scores of our category attribute classifiers. Finally, to estimate measurement of image similarity, we compute the Euclidean distances between the query and images from the dataset. The above three steps are formalized as follows:

1) Extract Low-level Features:We extract k low-level features $f_{i=1,\ldots,k}$ for each flower image I, and concatenate these vectors to form a large feature

978-1-4799-4808-6 /14/$31.00 ©2014 IEEE

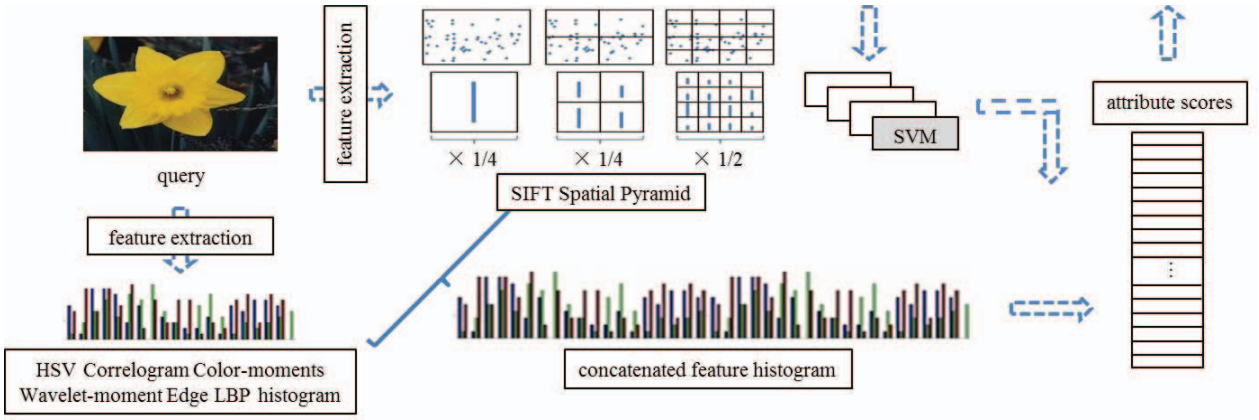

Fig. 1. Low-level features extraction and construction of the category attributes descriptor for a query. Step represented in gray requires a learning stage.

vector $F(I) = \langle f_1(I), \ldots, f_k(I) \rangle$.

2) Construct Category Attributes: For each extracted feature vector $F(I)$, we compute the output of n category attribute classifiers $C_{i=1,\ldots,n}$ in order to produce a category attributes vector $C(I)$ for the image, $C(I) = \langle C_1(F(I)), \ldots, C_n(F(I)) \rangle$.

3) Estimate measurement of image similarity: We utilize Euclidean distances between query image I and those D_j from the dataset, containing m images. The similarity distance function s:

$$s = \sqrt{(C(I) - C(D_j))^2} \; j = 1, \ldots, m \qquad (1)$$

A. Extraction of low-level features

In the following, we present three major types of low-level image features in our approach: color, texture, and shape related features. Fig. 1 illustrates the extraction of these features as well as the category attribute descriptor described in section II-B.

The color existence is sort of hard to identify flowers, because the color distribution is intersection among flowers. That is, it is helpful with color information rather to refine the possible species than identify the flower species. That a white flower could be a windflower or a lily valley but could not be a sunflower is a the best example to illustrate this. So color feature is less weighted than the others.

In addition, with illumination variation, there would be notable appearance differences, i.e. more lighter even white, which lead to confusion between classes. One way to reduce the effect of illumination variation is to use a color space which is less sensitive to it. Hence, we use HSV color space to describe the color of flowers.

For shape representation, it is appropriate to describe each petal in the same way, because petals are almost redundant. We utilize a rotation invariant descriptor SIFT visual word histogram. A dimension of 128 SIFT descriptor for each feature point is computed on a regular grid and patch [8]. A great quantity of SIFT descriptors are clustered by k-means (with k = 200). Given a set of cluster center (visual word) $\omega_i, i = 1, 2, \ldots, n$, each image $I_j, j = 1, 2, \ldots, N$, is

Fig. 2. Flowers with distinctive patterns. From left to right: tiger-lily with distinctive dots, pansy with distinctive stripes, fritillary with distinctive checks.

then represented by a n-dimension frequency histogram. Then the image is subdivided into three different levels, one with original size, one cutting into four equal block, one turning to be nine equal style box. For each level of style box, we count the features that fall in each one. Then, to balance their influence, each style box has been weighted as illustrated in the part of SIFT Spatial Pyramid in Fig. 1. Finally, we concatenate the above histograms to form a single long vector. Spatial Pyramid [14] is more helpful to record the spatial features of SIFT rather than BOF [3].

As shown in Fig. 2, flowers have distinctive petal texture patterns. We describe the texture by Local Binary Pattern (LBP) [1] features. First, a structure containing a mapping table for uniform rotation-invariant LBP codes is generated from a gray image. We set P = 16, R = 2 in [1] then get a histogram of LBP codes.

Before combining these histograms, HSV, LBP, SIFT histograms should be normalized first. Then the normalized histograms are concatenated to form an entire feature vector.

B. Construction of category attributes

Category classifiers $C_{i=1,\ldots,n}$ are used to depict the category attributes of flowers, e.g. daffodils, sunflower, pansy and giant white arum lily in Fig. 3. Thus we train several flower category attribute classifiers, treating attributes classification as a supervised learning problem. For each category, a group of positive and negative images is need for the training. For 17 Category Flower Dataset, each category classifier is a SVM with a linear kernel, trained using libsvm [11]. Examples of some category attributes are shown in Fig. 3.

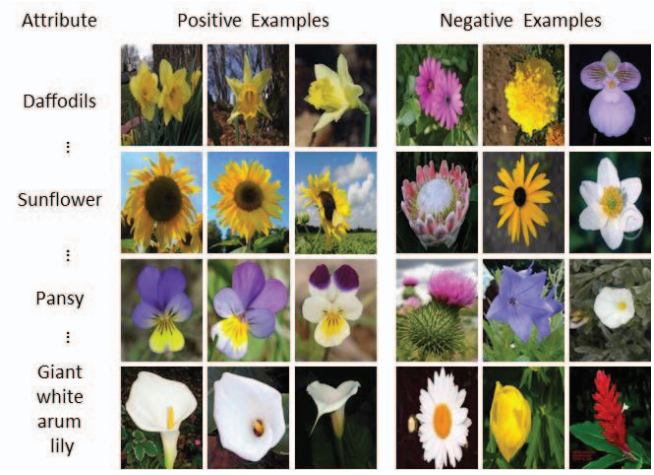

Fig. 3. Category Attributes for Training. Each row shows training examples of flower images that match the given attribute label (positive examples) and those don't (negative examples).

C. Estimate of image similarity measurement

We utilize Euclidean distances between query image and those from the dataset. The similarity distance function is as Eq. 1.

III. Experiments and Results

A. Dataset and evaluation

We present results of 17 Category Flower Dataset to evaluate content based flower image retrieval system. It is available online [13]. This dataset consists of 1,360 labeled images of 17 categories, with 80 images per category. This dataset is very challenging because there are large variation between same category while small variation between different ones considering viewpoint, scale, and illumination.For example, some classes can not be identified only by color, shape or texture. These flower images is from many websites as well as artificial photographs.

B. Results

To evaluate the performance, we compute mean average precision(mAP) as follows:

$$mAP = \frac{\sum_{q=1}^{Q} AP(q)}{Q} \tag{2}$$

where Q is the number of queries, and the average precision(AP) is computed using the query result information in the following:

$$AP = \frac{\sum_{k=1}^{N} (P(k) \times Rel(k))}{N\_R} \tag{3}$$

The relevance of the retrieved top k image (Rel(k)) and the precision (P(k)) in (3) are illustrated as (4) and (5):

$$Rel\,(k) = \begin{cases} 1, & \text{where } k \text{ is relevant} \\ 0, & \text{where } k \text{ is not relevant} \end{cases} \tag{4}$$

Fig. 4. Comparison of flower image retrieval results between our mid-level and low-level features. For the same query left, top right is top 20 retrieval result of category attribute based method, bottom is those of low-level feature based method. Fifteen relevant of category attributes, eleven relevant of low-level features.

$$P(k) = \frac{|\{relevant\ documents \cap retrieved\ documents\}|}{|\{retrieved\ documents\}|} \tag{5}$$

In (4) and (5), k represents the kth retrieved image. In (5), $|N|$ means the norm of variable N, and $relevant\ documents$ means the number of all relevant images from the dataset which contain the similar object as the query while $retrieved\ documents$ means the number of the images retrieved from the query.

In Fig. 5 we show how the performance varies with the number of top-K query images on 17 Category Flower Dataset. Especially when the number of query images increases, our method outperforms low-level feature based image retrieval method. We credit this excellent performance boost to the ability of category attribute to overcome some over-low-level description of the image when large queries are used.

IV. Conclusion and Future Works

In this paper we present a new method for flower image retrieval. We make use of category attributes to construct middle level image representation. This way it construct a new mid-level visual presentation - category attribute - for the query and images in the dataset. This approach captures flower category attribute, is more effective and has lower dimension, so we don't need to rely on costly and inefficiently similarity measurement between low-level feature vectors. This approach achieves good results on 17 Category Flower Dataset. In comparison to low-level feature image retrieval, the proposed method shows substantial performance improvement.

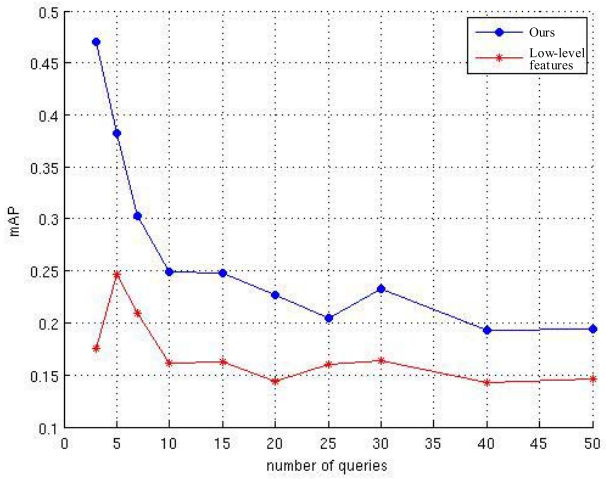

Fig. 5. Performance vs. number of queries using our approach.

V. ACKNOWLEDGEMENT

This work is partly supported by National Natural Science Foundation of China (91320101), Shenzhen Basic Research Program (JC201005270350A, JCYJ20120903092050890, JCYJ20120617114614438), 100 Talents Programme of Chinese Academy of Sciences, and Guangdong Innovative Research Team Program (No.201001D0104648280).

REFERENCES

[1] Ojala, Timo, Matti Pietikainen, and Topi Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." Pattern Analysis and Machine Intelligence, IEEE Transactions on 24.7 (2002): 971-987.

[2] Nilsback, M-E., and Andrew Zisserman. "A visual vocabulary for flower classification." Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. Vol. 2. IEEE, 2006.

[3] Sivic, Josef, and Andrew Zisserman. "Video Google: A text retrieval approach to object matching in videos." Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. IEEE, 2003.

[4] Zhang, Jianguo, et al. "Local features and kernels for classification of texture and object categories: A comprehensive study." International journal of computer vision 73.2 (2007): 213-238.

[5] Stricker, Markus A., and Alexander Dimai. "Color indexing with weak spatial constraints." Electronic Imaging: Science and Technology. International Society for Optics and Photonics, 1996.

[6] Meek, C. E. "An efficient method for analysing ionospheric drifts data." Journal of Atmospheric and Terrestrial Physics 42.9 (1980): 835-839.

[7] Kumar, Neeraj, et al. "Attribute and simile classifiers for face verification." Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009.

[8] Lowe, David G. "Distinctive image features from scale-invariant keypoints." International journal of computer vision 60.2 (2004): 91-110.

[9] Varma, Manik, and Andrew Zisserman. "Classifying images of materials: Achieving viewpoint and illumination independence." Computer Vision–ECCV 2002. Springer Berlin Heidelberg, 2002. 255-271.

[10] Berg, Tamara L., Alexander C. Berg, and Jonathan Shih. "Automaticattribute discovery and characterization from noisy web data." Computer Vision–ECCV 2010. Springer Berlin Heidelberg, 2010. 663-676.

[11] Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: a library for support vector machines." ACM Transactions on Intelligent Systems and Technology (TIST) 2.3 (2011): 27.

[12] Smeulders, Arnold WM, et al. "Content-based image retrieval at the end of the early years." Pattern Analysis and Machine Intelligence, IEEE Transactions on 22.12 (2000): 1349-1380.

[13] M. Nilsback and A. Zisserman. 17 Category Flower Dataset. http://www.robots.ox.ac.uk/∼vgg/data/flowers/17/, 2006.

[14] Lazebnik, Svetlana, Cordelia Schmid, and Jean Ponce. "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories." Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. Vol. 2. IEEE, 2006.