Perceptual Image Retrieval by Adding Color Information to the Shape Context Descriptor

Marçal Rusiñol, Farshad Noorbakhsh, Dimosthenis Karatzas, Ernest Valveny and Josep Lladós
Computer Vision Center, Dept. Ciències de la Computació
Edifici O, Univ. Autònoma de Barcelona
08193 Bellaterra (Barcelona), Spain
$\{$*marcal,farshad,dimos,ernest,josep*$\}$*@cvc.uab.cat*

Abstract

In this paper we present a method for the retrieval of images in terms of perceptual similarity. Local color information is added to the shape context descriptor in order to obtain an object description tackling with shape and color as visual cues. We use a color naming algorithm in order to represent the color information from a perceptual point of view. The proposed method has been tested in two different applications, a queried by color sketches object retrieval scenario and a color trademark retrieval problem. Experimental results show that the addition of the color information significantly outperforms the shape context descriptor.

1 Introduction

In the last two decades, with the popularization of the internet, a huge amount of information resources have emerged. However, not all the information is easily accessible. More and more, an explosive growth of information is stored in image formats. Since search engines index and retrieve information in terms of textual queries, there is a lack of accessibility to this particular kind of information. Textual search of images rely on the metadata they have associated instead of analyzing the actual contents of the images. In order to tackle this problem, in the last years a lot of efforts have been devoted to the problem of content-based image retrieval (CBIR).

One of the possible query paradigms in CBIR applications is known as *query-by-sketch*. The user creates the query image with a drawing tool. The systems working with sketched queries must be able to handle the severe deformations of the sketches in order to retrieve the images which are perceptually similar to the sketches. In this particular scenario, most of the literature just rely on shape information. These approaches try to match the sketches with the object's contours as in [5, 4]. The addition of color as a discriminant visual cue when trying to retrieve images by perceptual similarity seems important. However, just a few works dealing with colored sketches can be found in the literature, as for instance [7, 2].

Inspired by the work of Diplaros et al. [3], we propose to enhance the shape context descriptor with color information. A color naming algorithm is used in order to represent the color information from a perceptual point of view. We tested the proposed method in two different applications, an object retrieval scenario queried by color sketches, and a color trademark retrieval problem.

The paper is organized as follows. In the next section we give the basic background of the descriptors we use and we detail the method. In Section 3, the experimental evaluation of the method and the results are provided. Finally, in Section 4 we provide some conclusions about the work.

2 Image description

We give in this section the basic background on the shape context descriptor and the color naming algorithm that we use. Then we will detail how these two descriptors are combined and how the images are matched.

2.1 Shape Context

The shape context (SC) descriptor was proposed by Belongie et al. in [1]. It allows to measure shape similarity by recovering point correspondences between two objects. In the first step, a set of interest points are selected. Edge elements from the shape are sampled in order to obtain a fixed number of n points p_i. In the next step, a histogram using log-polar coordinates captures the distribution of points within the plane relative to each point of the shape. For each point p_i of the shape, a histogram

$$S_i(k) = \#\{q \neq p_i : (q - p_i) \in bin(k)\} \qquad (1)$$

is computed. In our experimental setup, we have chosen 5 bins for $\log r$ and 12 bins for θ. Translational invariance comes naturally to shape context since all the histograms are computed from reference points. Scale invariance is obtained by normalizing all radial distances by the mean distance between all the point pairs in the shape. Angles at each point are measured relative to the direction of the tangent at that point to provide invariance to rotation.

2.2 Color Naming

In order to represent the perceptual color information of the objects, we apply a color naming model. We use the method proposed by van de Weijer et al. in [8]. A probabilistic latent semantic analysis (PLSA) model learned on a set of images retrived from Google, results in a $32 \times 32 \times 32$ lookup table which allows to map pixel values over color names. Eleven basic color terms are considered in this approach, namely black, white, red, green, yellow, blue, brown, orange, pink, purple and gray. By applying this color naming algorithm to an image we obtain a color quantization based on how humans would perceive and describe the color information.

We define the color descriptor K as the vector containing the probability of the color names for a particular point of the image as

$$K = \{p(n_1|f(x)), p(n_2|f(x)), ..., p(n_{11}|f(x))\} \quad (2)$$

where n_i is the i-th color name and $f(x)$ the color value of a given pixel x. $p(n_i|f(x))$ is then the probability of a color name given a pixel value.

2.3 Local Color Names Histograms

For each sampled point p_i of the image we obtain a local description of the shape arround this point by using the shape context histogram S_i. In order to add color information to this shape description we apply the color naming model locally at the same point p_i. A circular mask is defined as the region of interest centered

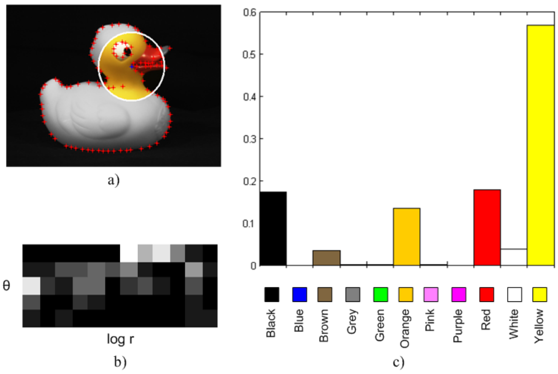


Figure 1. The color shape context descriptor. a) The object under analysis and a given point p_i; b) shape context descriptor and c) local color names histogram.

at p_i. The size of the mask is computed with relation to the mean distance between all the point pairs of the shape.

All the pixels x_j in the region of interest of a point p_i have an associated color descriptor K_j. The local color name histogram C_i is then defined the accumulation of probabilies of each of the eleven color names computed as

$$C_i(k) = \frac{1}{N}\sum_{j=1}^{N} p(n_k|f(x_j)) \qquad (3)$$

where N is the total number of pixels in the mask and k is each of the eleven color names.

The color shape context descriptor (CSC) is the combination of both descriptors S_i and C_i at each point of the shape. We can see an visual example of color shape context descriptors in Fig. 1. For the point p_i (plotted in blue) a mask is defined in order to compute the shape description in Fig. 1b) and the local color names distribution in Fig. 1c).

2.4 Matching

In order to match two shapes we have to find the point correspondences. The way to compute the matching among the two set of points is by using a bipartite graph matching approach that puts in correspondence points having similar shape and color descriptions. The distance between a couple of points p_i and p_j is computed as

$$d(p_i, p_j) = \chi^2(S_i, S_j) \times \chi^2(C_i, C_j) \qquad (4)$$

by using the χ^2 distance

$$\chi^2(A,B) = \frac{1}{2}\sum_{m=1}^{k}\frac{[A(m)-B(m)]^2}{A(m)+B(m)+\epsilon}. \quad (5)$$

By multiplying the distance of the shape context descriptor and the local color descriptor in eq. 4, we reinforce the matches which are similar in shape and color and we hinder the cases where we have similar color but different spatial point distribution or viceversa. Given a set of costs $d(p_i, p_j)$ between all pairs of points, the final distance between two shapes is determined by minimizing the total cost of matching

$$H(\pi) = \sum_i d(p_i, p_{\pi(i)}) \quad (6)$$

where π is a permutation of points and H can be computed by applying the Hungarian method. In order to obtain a more robust matching, the most usual techniques involve the computation of an affine transform that matches the set of points from one shape to another. However, in our application scenarios we do not have to face this kind of transformations.

3 Experimental results

To evaluate our proposed work, we have chosen two different datasets. The ALOI dataset [6] includes in 1000 objects, each one with 12 different illuminations. Twenty objects from the ALOI dataset were sketched by eleven diffent users (details can be found in [2]) and are taken as queries. The second dataset consits of 323 color trademarks organized in 124 classes. Fourteen classes are taken as models and three different logos from each class act as a query to the logo database.

We can see in Fig. 2 some qualitative results of the object retrieval by sketches experiment. As we can see, most of the results are visually similar in the case of the CSC method, whereas a lot of false alarms appear in the case of just using shape as visual cue to describe the objects. In Fig. 3 we provide the quantitative results of the whole experiment by giving the ROC curves, the area under curve (AUC) and mean average precision ($AveP$) measures in Table 1. As we can appreciate the CSC method outperforms in all the cases the SC method.

For the second dataset we propose a psychophysic evaluation procedure. Ten users were asked to rank the five top similar logos for a given query and a set of logos extracted from the results of both the SC and the CSC. In this case we can see that even if the user judgement of similarity is usually scattered among the set of logos, a small set of logos are selected by most of the users.

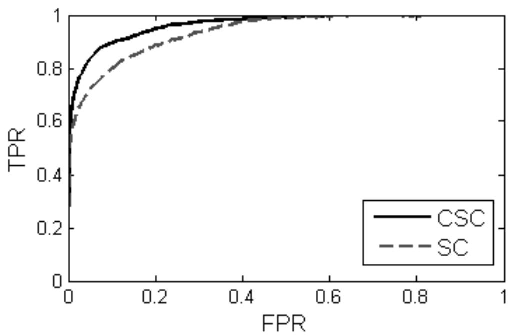


Figure 3. ROC curve for the object retrieval from the ALOI dataset by color sketched queries.

We took these logos as the groundtruth and we computed the amount of these logos present in the results. Retrieving logos using the SC descriptor, a 35.5% of the top fifty results were marked as similar by humans. In the case of using the CSC method the score went to 37.3%. We can see in Fig. 4 an example of this labelling where the green boundingboxes correspond to positive logos marked by the users.

Table 1. Evaluation measures for the ALOI experiment.

Descriptor	AUC (%)	AveP (%)
CSC	96.48	42.15
SC	93.71	40.81

4 Conclusions

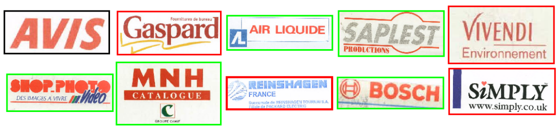
In this paper we have developped and tested a method for the image retrieval based on the perception of color and shaoe by defining a new description strategy. The method has been tested on two different datasets with different natures respect into scratches and trademark models. With the psychophysics evaluation we concluded that combining the local color features with shape context descriptor gives a better result than just using shape as a visual cue.

Acknowledgments

This work has been partially supported by the Spanish projects TIN2006-15694-C02-02, TIN2009-14633-C03-03 and CONSOLIDER - INGENIO 2010 (CSD2007-00018).

a) Color shape context b) Shape context

Figure 2. Object retrieval from the ALOI dataset by color sketched queries. First 15 results for the sketch queries.



a) Color shape context

b) Shape context

Figure 4. First ten results when retrieving the trademark AVIS. Framed in green the logos that the users consider similar to the query.

References

[1] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, 2002.

[2] A. Borràs. *Contributions to the content-based image retrieval using pictorial queries*. PhD thesis, Universitat Autònoma de Barcelona, 2009.

[3] A. Diplaros, T. Gevers, and I. Patras. Combining color and shape information for illumination-viewpoint invariant object recognition. *IEEE Transactions on Image Processing*, 15(1):1–11, 2006.

[4] V. Ferrari, T. Tuytelaars, and L. V. Gool. Object detection by contour segment networks. In *Computer Vision, ECCV*, volume 3953 of *Lecture Notes on Computer Science*, pages 14–28. Springer-Verlag, 2006.

[5] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: The QBIC system. *Computer*, 28(9):23–32, 1995.

[6] J. Geusebroek, G. Burghouts, and A. Smeulders. The Amsterdam library of object images. *International Journal of Computer Vision*, 61(1):103–112, 2005.

[7] C. Jacobs, A. Finkelstein, and D. Salesin. Fast multiresolution image querying. In *Proceedings of the Twenty-second Annual Conference on Computer Graphics and Interactive Ttechniques*, pages 277–286. ACM, 1995.

[8] J. van de Weijer, C. Schmid, J. Verbeek, and D. Larlus. Learning color names for real-world applications. *IEEE Transactions on Image Processing*, 18(7):1512–1523, 2009.