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
Email: {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 integrating both 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, an object retrieval scenario based on color sketch queries and a color trademark retrieval problem. Experimental results show that the addition of the color information significantly outperforms the sole use of the shape context descriptor.**

***Keywords*-Graphics recognition; multimedia retrieval; perceptual description; shape context; color naming.**

I. 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 explosively growing amount 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 effort has 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 relies on shape information. These approaches try to match the sketches with the object's contours as in [1], [2]. The addition of color as a discriminant visual cue when trying to retrieve images by perceptual similarity seems important. However, a limited amount of work dealing with colored sketches can be found in the literature, as for instance [3], [4]. The same applies to the problem of logo recognition and retrieval. When registering a new trademark it is important to avoid trademark conflicts and to be sure that no other trademark that looks similar to the new one is already registered. In that particular scenario, the retrieval of logos by visual similarity has great interest. However, most of the works on logo retrieval are focused on the analysis of the logos' contours or regions without taking into account the color information.

Inspired by the work of Diplaros et al. [5], 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 based on color sketch queries, 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.

II. Image description

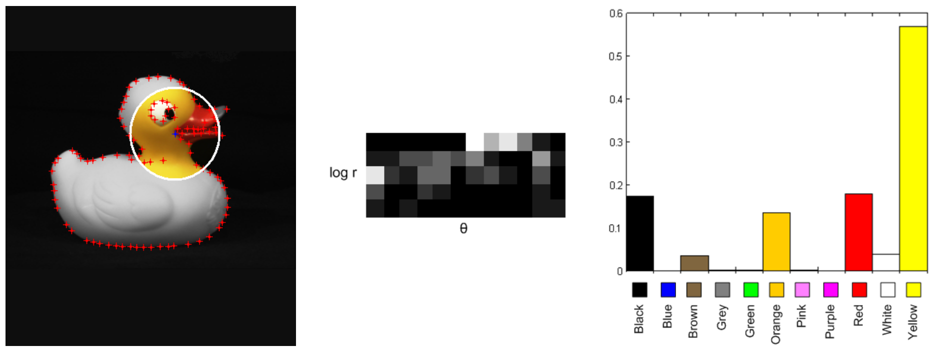
We give in this section the basic background on the shape context descriptor and the color naming algorithm that we use. Then we will detail how these two descriptors are combined and how the images are matched.

A. Shape Context

The shape context (SC) descriptor was proposed by Belongie et al. in [6]. 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

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

with 5 bins for the radial distance $\log r$ and 12 bins for the angles θ is computed. Each bin k counts the occurrences of all the points q of the shape that fall into it. Translational invariance comes naturally to shape context since all the



(a) (b) (c)

Figure 1. Example of the color shape context descriptor.

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.

B. Color Naming

In order to represent the perceptual color information of the objects, we apply a color naming model. A color naming method provides a way to map a color value to one of the predefined number of semantic color groups corresponding to the color names used in everyday communications. Color naming models can be defined through psychophysical measurements [7] or through statistical learning [8]. 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 retrieved from Google, results in a $32 \times 32 \times 32$ lookup table which allows to map pixel values to 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.

C. Local Color Names Histograms

For each sampled point p_i of the image we obtain a local description of the shape around 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 at p_i. In order to keep invariance to scale, the size of the mask is computed with relation to the mean distance between all the points pairs in 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 evidences of each of the eleven color names computed as

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

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 a visual example of color shape context descriptors in Fig. 1. For the point p_i (plotted in blue) in Fig. 1(a) the shape context descriptor [6] is shown in Fig. 1(b). The local color names distribution in Fig. 1(c) comes from the application of the color naming model to the pixels from the mask.

D. Matching

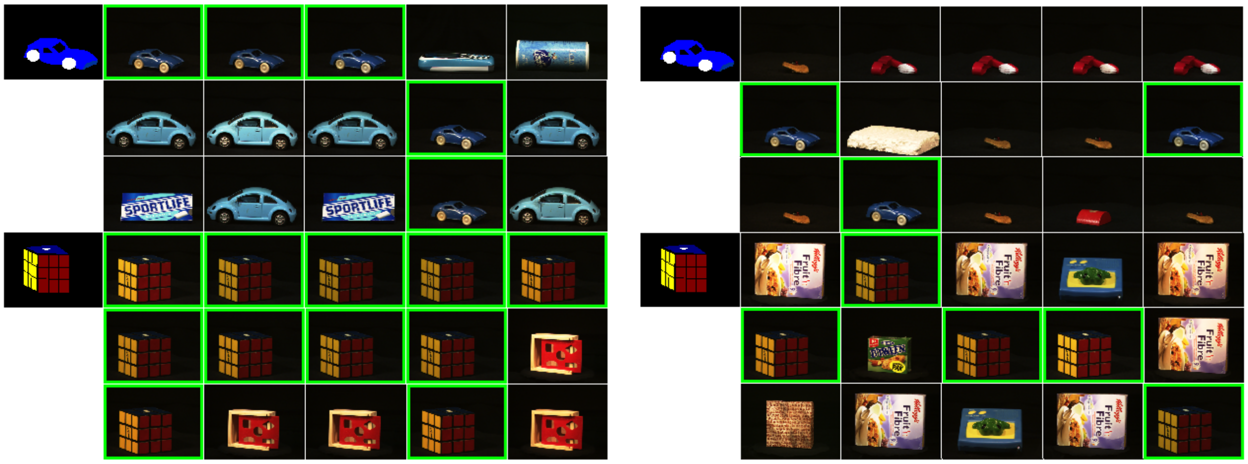
In order to match the sketch with the images in the collection we have to find the point correspondences. The way to compute the matching among the two set of points is by using a bipartite graph matching approach that puts in correspondence points having similar shape and color descriptions. The distance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

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

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



(a) (b)

Figure 2. Results for the ALOI experiment. First 15 results for the sketch queries. (a) Color shape context. (b) Shape context.

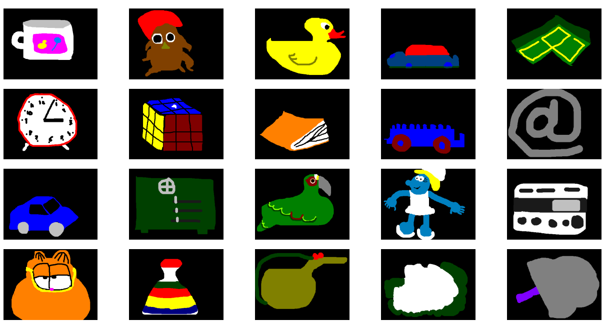


Figure 3. Example of the sketch queries for the ALOI dataset

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. In addition, the distance is kept normalized in $[0, 1]$. We tried other combination methods such as the minimum of both distances or the inverse of the absolute distance difference, obtaining worse performances than with the multiplication. Given a set of local distances $d(p_i, p_j)$ between all pairs of points, the final distance between the sketch query and the image is determined by minimizing the total cost of matching

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

where π is a permutation of points and H is 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.

III. Experimental results

To evaluate our proposed work, we have chosen two different datasets. The ALOI dataset [9] includes in 1000 objects, each one with 12 different illuminations. Twenty objects from the ALOI dataset were sketched by eleven different users (details can be found in [4]) and are taken as queries, we can appreciate the difficulty of using this kind of perceptual query representation in a retrieval framework in Fig. 3. The second dataset consists of 323 color trademarks. Fourteen classes are taken as models and three queries are used per class.

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. 4 we provide the quantitative results of the whole experiment by giving the receiver operating characteristic (ROC) curves plotting the true positive rates (TPR) and the false positive rates (FPR), the area under curve (AUC) and mean average precision ($AveP$) measures in Table I. Note that these results are computed by looking at the objects' class, regardless of the similarity among them. The blue cars in Fig. 2(a) that are not highlighted in green are counted as false alarms even if they are perceptually similar. Even despite this fact, we can appreciate the CSC method outperforms in all the cases the SC method.

For the second dataset the issue is different as no exact matches are sought for, but similar logos. Such a scenario would be useful in a trademark search case in a intellectual property office when the users want to retrieve similar logos than the ones they want to register. In this case there is a need for a different evaluation procedure since a groundtruth is not avaliable, we propose to use a psychophysical one. Ten users were asked to rank the five top similar logos for

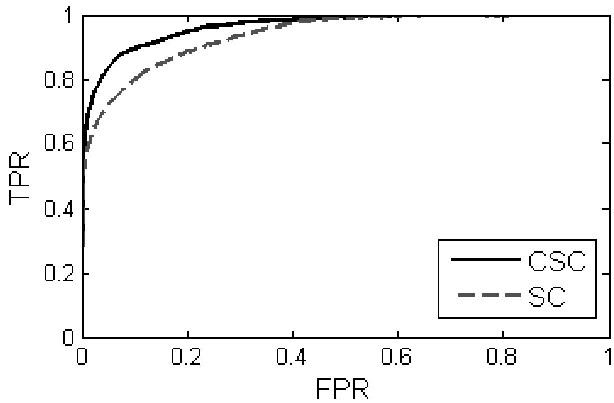


Figure 4. ROC curve for sketched queries over the ALOI dataset

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. 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 up to 37.3%. We can see in Fig. 5 an example of this labelling where the green boundingboxes correspond to positive logos marked by the users.

Table I
PERFORMANCE MEASURES FOR THE ALOI EXPERIMENT.

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

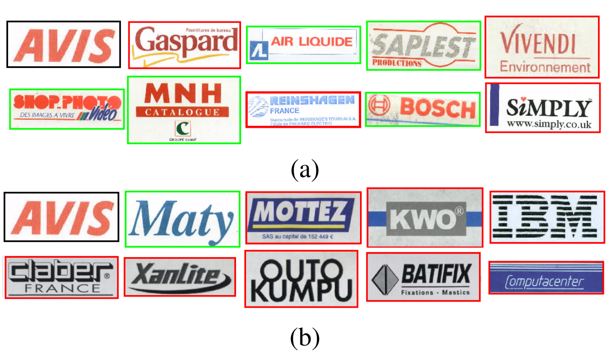


Figure 5. First ten classes for the trademark AVIS. (a) Color shape context. (b) Shape context

IV. Conclusions

In this paper we have developed and tested a method for image retrieval based on the perception of color and shape by defining the color shape context descriptor. The proposed method has proven to give good results in applications where no exact matches are sought for, but in which the user wants to retrieve similar objects to its query. Both the psychophysical evaluation and the qualitative results show that combining the local color features with the shape descriptor gives better results than just using shape as a visual cue.

Acknowledgments

This work has been partially supported by the Spanish projects TIN2006-15694-C02-02, TIN2008-04998, TIN2009-14633-C03-03 and CONSOLIDER - INGENIO 2010 (CSD2007-00018). Logo images were kindly provided by ITESOFT (www.itesoft.fr).

References

[1] 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*, vol. 28, no. 9, pp. 23–32, 1995.

[2] V. Ferrari, T. Tuytelaars, and L. V. Gool, "Object detection by contour segment networks," in *Computer Vision, ECCV*, ser. Lecture Notes on Computer Science. Springer-Verlag, 2006, vol. 3953, pp. 14–28.

[3] C. Jacobs, A. Finkelstein, and D. Salesin, "Fast multiresolution image querying," in *Proceedings of the Twentysecond Annual Conference on Computer Graphics and Interactive Ttechniques*. ACM, 1995, pp. 277–286.

[4] A. Borràs, "Contributions to the content-based image retrieval using pictorial queries," Ph.D. dissertation, Universitat Autònoma de Barcelona, 2009.

[5] A. Diplaros, T. Gevers, and I. Patras, "Combining color and shape information for illumination-viewpoint invariant object recognition," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 1–11, 2006.

[6] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[7] C. Párraga, R. Benavente, M. Vanrell, and R. Baldrich, "Psychophysical measurements to model intercolor regions of color-naming space," *Journal of Imaging Science and Technology*, vol. 53, no. 3, pp. 1–8, 2009.

[8] J. van de Weijer, C. Schmid, J. Verbeek, and D. Larlus, "Learning color names for real-world applications," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1512–1523, 2009.

[9] J. Geusebroek, G. Burghouts, and A. Smeulders, "The Amsterdam library of object images," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 103–112, 2005.