## Sketch based Image Retrieval using Learned KeyShapes (LKS)

Jose M. Saavedra jose.saavedra@orand.cl Juan Manuel Barrios juan.barrios@orand.cl

A natural alternative for querying in an image retrieval system is by simply drawing what one has in mind. Indeed, drawing was the primitive means of communication between humans. One of the goals of an image retrieval scenario is to provide users a simple modality for querying. Thereby, a drawing means a simple hand-drawn sketch composed only of strokes that users can do easily, lacking color or texture. Examples of hand-drawn sketches are shown in Figure 1.



Figure 1: Examples of hand-drawn sketches. The last two are from the Eitz's dataset [1].

This querying modality leads to the *sketch based image retrieval problem* (SBIR) which is a challenging problem because of two main reasons: (i) images that we want to retrieve are not sketches, (ii) query sketches show certain level of ambiguity by nature that may make a method get confused easily. Consequently, state-of-the-arts SBIR approaches [2, 4] still show low performance.

Therefore, taking some ideas of the human visual perception, we present a novel method for sketch based image retrieval. Our method, is based on detecting the occurrences of mid-level patterns on a sketch. To this end, we figure out a set of patterns (*learned keyshapes*) by means of an unsupervised learning process. We then build a histogram that counts the occurrences of the patterns in the underlying sketch. The histogram is built using soft-voting, spatial division and squared root normalization. We show new state-of-the-art results in two available datasets doubling the precision achieved by current methods.

Our proposal consists of two stages (Figure 2): (1) figure out a set of *keyshapes*, (2) generate the LKS descriptors based on the detected set of *keyshapes*, that will be used later for similarity search.



Figure 2: A scheme of our Learned KeyShapes based proposal (LKS).

**Keyshape Generation**: We got inspired by the work of Lim et al. [3], where reliable contour maps are obtained by detecting a set of *sketch tokens*, that are previously learned from a collection of contour images. For training, we use the Eitz's sketch dataset [1]. We extract one million of  $31 \times 31$  sketch patches, each one centered in a stroke point. Sketch patches are coded by a DAISY descriptors that are then clustered by K-means. In Figure 3, we show a sample of learned *keyshapes* using K = 150.



Figure 3: A sample of learned keyshapes when K = 150.

**Keyshape Generation**: Instead of using low level methods as Canny, we prefer to use the sketch token based approach proposed by Lim et al. [3].

**KeyShape Detection**: We extract sketch patches from the input query as well as from the contour of a test image. Each patch is centered in each stroke point. For

Orand S.A. Estado 360, Of. 702-A, Santiago, Chile

each patch, we search for the *P* nearest keyshapes (P-NK) using DAISY descriptors.

**LKS-Histogram Computation** We build a *K*-dimensional histogram, where *K* is the number of keyshapes. In this process, three steps are involved: (1) Gaussian Kernel based Voting using P-NK, (2) Spatial Division, (3) Pow-based Normalization.

We evaluate our proposal in two public datasets: Saavedra's and *Flickr15k*. The performance are compared under the mean average precision metric (mAP) (Table 1) as well as the precision-recall graphic (Figure 4).

|  | HOG    | GF-HOG[2]  | SHELO[4] | LKS    | gain  |
|--|--------|------------|----------|--------|-------|
| Saavedra's   | 0.2355 | unreported | 0.2766   | 0.3251 | 17.5% |
| Flickr15K  | 0.0771 | 0.1222     | 0.1236   | 0.2450 | 98.2% |
| Table 1: Mean Average Precision comparing our proposals LKS with |        |            |          |        |       |
| state-of-the-art methods   |        |            |          |        |       |



Figure 4: Precision-Recall graphic showing the performance of LKS (blue curve) and SHELO (red curve) on the Saavedra's dataset (on the left) and Flickr15k dataset (on the right).



Figure 5: Examples of results using LKS on the Flickr15K dataset. Each row shows a query sketch together with the five first responses.

- Mathias Eitz, James Hays, and Marc Alexa. How do humans sketch objects? ACM Trans. Graph., 31(4):44:1–44:10, July 2012.
- [2] Rui Hu and John Collomosse. A performance evaluation of gradient field hog descriptor for sketch based image retrieval. *Computer Vision and Image Understanding*, 117(7):790–806, July 2013.
- [3] J.J. Lim, C.L. Zitnick, and P. Dollar. Sketch tokens: A learned mid-level representation for contour and object detection. In *Computer Vision and Pattern Recognition (CVPR)*, 2013 IEEE Conference on, pages 3158–3165, June 2013.
- [4] Jose M. Saavedra. Sketch based image retrieval using a soft computation of the histogram of edge local orientations (s-helo). In *International Conference* on Image Processing, ICIP'2014 (To appear), 2014.