An Improved Histogram of Edge Local Orientations for Sketch-Based Image Retrieval

Jose M. Saavedra* and Benjamin Bustos

PRISMA Research Group
Department of Computer Science, University of Chile
{jsaavedr,bebustos}@dcc.uchile.cl

Abstract. Content-based image retrieval requires a natural image (e.g, a photo) as query, but the absence of such a query image is usually the reason for a search. An easy way to express the user query is using a line-based hand-drawing, a sketch, leading to the sketch-based image retrieval. Few authors have addressed image retrieval based on a sketch as query, and the current approaches still keep low performance under scale, translation, and rotation transformations. In this paper, we describe a method based on computing efficiently a histogram of edge local orientations that we call HELO. Our method is based on a strategy applied in the context of fingerprint processing. This descriptor is invariant to scale and translation transformations. To tackle the rotation problem, we apply two normalization processes, one using principal component analysis and the other using polar coordinates. Finally, we linearly combine two distance measures. Our results show that HELO significantly increases the retrieval effectiveness in comparison with the state of the art.

1 Introduction

Due to the progress in digital imaging technology, image retrieval has become a very relevant discipline in computer science. In a content-based image retrieval system (CBIR), an image is required as input. This image should express what the user is looking for. But, frequently the user does not have an appropriate image for that purpose. Furthermore, the absence of such a query image is usually the reason for the search [1]. An easy way to express the user query is using a line-based hand-drawing, a *sketch*, leading to the *sketch-based image retrieval* (SBIR). In fact, a sketch is the natural way to make a query in applications like CAD or 3D model retrieval [2].

Although there are many publications on CBIR, a few authors have addressed image retrieval based on sketches. Some of these works are *Query by Visual Example*(QVE) [3], *Edge Histogram Descriptor* (EHD) [4], *Image Retrieval by Elastic Matching* [5], *Angular partitioning of Abstract Images* [6], and *Structure Tensor*[1], that will be briefly discussed in the next section. Although these methods are applied to SBIR, they still show poor effectiveness under scale, translation, and rotation issues.

* Partially funded by CONICYT(Chile) through the Doctoral Scholarship.

M. Goesele et al. (Eds.): DAGM 2010, LNCS 6376, pp. 442–451, 2010.

The main contribution of this work is to propose a novel method based on edge orientations that gets a global representation of both the sketch and the test image. We improve the effectiveness of SBIR by estimating local orientations in a more precise way, obtaining a *histogram of edge local orientations*. The local orientations are computed using a strategy applied for computing directional fields of fingerprints, in the context of biometric processing [7]. Our proposed approach is invariant to scale and translation transformations. To tackle the rotation problem, we apply two different normalization processes, one using principal component analysis and the other using polar coordinates. Finally, we use a combined distance as similarity measure. We experimentally show that our proposed method significantly outperforms the state of the art.

The rest of this paper is organized as following. Section 2 describes the current methods for SBIR. Section 3 describes in detail the proposed method. Section 4 presents the experimental evaluation. Finally, Section 5 presents conclusions.

2 Related Work

There are a few works on sketch-based image retrieval. One of the first proposals is QVE [3]. The test image and the query are transformed into abstract representations based on edge maps. To measure similarity between two abstract representations, this method uses a correlation process based on bitwise operations. To get translation invariance, the correlation is carried out under horizontal and vertical shifts. This method is expensive and not rotation invariant. In addition, this approach does not permit indexing [6].

Another approach was presented by Del Bimbo and Pala [5]. This approach is based on elastic deformation of a sketch to match a test image. The necessary effort to adjust the query to the test image is represented by five parameters that are the input to a multi-layer neural network. This method is also expensive and not rotation invariant, and to get a good performance the query and the test image need to have similar aspect ratios, narrowing its scope.

Other methods use edge information such as edge orientation or density. One of this methods is that proposed by Jain and Vailaya [8]. They proposed a shape descriptor using a histogram of edge directions (HED) among their work on combining shape and color descriptors for CBIR. The idea is to quantize the edge orientation and to form a B-bins histogram. Although this approach may be scale and translation invariant, it is not robust to rotation changes.

Another edge-based approach is the Edge Histogram Descriptor (EHD) that was proposed in the visual part of the MPEG-7 [9]. An improved version of EHD was proposed by Sun Won et al. [4]. The idea is to get a local distribution of five types of edges from local regions of the image. The juxtaposition of local distributions composes the final descriptor. Although this approach is invariant to scale and translation transformations, it is not rotation invariant.

The histogram of distance distribution (HDD) is another descriptor that could also be applied for SBIR. HDD consists in selecting a sample of points from an edge map and then computing distances between random pairs of points. This descriptor has been used for 3D model retrieval [10] and for shape matching like in *Shape Context* [11]. Although this descriptor is invariant to translation, scale, and rotation, it is strongly dependent on the size of the sample.

An important work on SBIR was presented by Chalechale et al. [6]. This approach is based on angular partitioning of abstract images (APAI). The angular spatial distributions of pixels in the abstract image is the key concept for feature extraction. The method divides the abstract image into K angular partitions or slices. Then, it uses the number of edge points falling into each slice to make up a feature vector. To get rotation invariance, the method applies Fourier transform to the resulting feature vector. Although the method is partially invariant to translation, scaling, and rotation, it requires to recover almost 13% from the database to retrieve the correct one, so its effectiveness is low.

Recently, Eitz et al. [1] presented a new approach for SBIR. In this approach, the test image and the query are decomposed into $a \times b$ cells. Then, this method computes gradients at each edge point. To represent a unique orientation in each cell, the method computes the *structure tensor* (ST) over the local gradients. Similarity between a test image and a query is computed comparing corresponding local structure tensors. This approach is not rotation invariant.

We observe that the sketch-based image retrieval is still an open problem, because the current methods show poor effectiveness under scale, translation, and rotation changes. Thus, the main contribution of this work is to improve the effectiveness of image retrieval having as query a line-based hand-drawing.

3 Proposed Method

Our method is based on estimating local edge orientations and forming a global descriptor named HELO (Histogram of Edge Local Orientations). Since noise affects adversely the edge orientation computation [12], its presence in an image may cause descriptors to have low performance for image retrieval. So, we use a local method, which is robust to noise, to estimate edge orientations. In addition, using a local estimation, the sketches do not need to be drawn with continuous strokes.

3.1 HELO Descriptor

Our method works in two stages. The first one performs preprocessing tasks to get an abstract representation of both the sketch and the test image, while the second one make up the histogram. A detailed description is shown below:

- **Preprocessing**: In this stage, the test images are preprocessed off-line. First, the method uses the Canny algorithm [13] to get an edge map from each test image. For the Canny algorithm, we use a 9×9-size gaussian mask and a $\sigma = 1.5$. Then, the method applies a cropping operation to the result using horizontal and vertical projections in a similar way to that applied in the context of text recognition [14].

  The sketch is preprocessed on-line. First, the method uses a simple thresholding to get a binary representation of the sketch. Then, the method applies a cropping operation to the result in a similar way as in the previous case.

– **Histogram Computation**: Here, our approach computes a K-bin histogram based on local edge orientations. We propose to use a method applied for estimating directional fields of fingerprints [7], that allows us to minimize the noise sensitivity by a orientation local estimation. The main idea is to double the gradient angle and to square the gradient length. This has the effect that strong orientations have a higher vote in the local average orientation than weaker orientations [7]. This improves the retrieval effectiveness on the SBIR. The local orientation estimation works as follows:
  - Divide the image into $W \times W$ blocks. We regard each block as a local area where we will estimate the local orientation. In this approach the block size is dependent on the image size to deal with scale changes.
  - Compute gradient respect to x and to y for each pixel in a block, which will be called G_x and G_y, respectively. Here, we use Sobel masks [15].
  - Compute local orientations as follows:
    * Let b_{ij} be a block and α_{ij} its corresponding orientation ($i, j = 1..W$).
    * Let L_x and L_y be the set of local gradients of an image respect to x and y, computed on each block b_{ij} as follows:

$$L_y^{ij} = \sum_{(r,s)\in b_{ij}} 2G_x(r,s)G_y(r,s) \tag{1}$$

$$L_x^{ij} = \sum_{(r,s)\in b_{ij}} (G_x(r,s)^2 - G_y(r,s)^2) \tag{2}$$

      here, L_β^{ij} is the gradient on b_{ij} in the direction β.
    * Apply a gaussian filter on L_x and L_y to smooth the components. We use a gaussian filter with $\sigma = 0.5$ and a 3×3-size mask.
    * Calculate the local orientation α_{ij} as follows:

$$\alpha_{ij} = \frac{1}{2}tan^{-1}\left(\frac{L_y^{ij}}{L_x^{ij}}\right) - \frac{\pi}{2} \tag{3}$$

      At this point, we normalize α_{ij} to the range between 0 and π.
  - Create a K-bin histogram to represent the distribution of the local orientation in the image.
  - Map each local orientation α_{ij} to the corresponding histogram bin to increase it by one. Blocks with a few edge points are neglected. We use a threshold th_{edge} to filter those blocks. We call the resulting histogram the histogram of edge local orientation (HELO). Fig. 1 shows an orientation field of a test image, computed by HELO.

HELO is invariant to translation because the orientation is independent of edge positions. In addition, since the block size depends on the image size, HELO is also invariant to scale changes. Moreover, we measure similarity between two HELO descriptors using the L_1 distance (Manhattan distance).

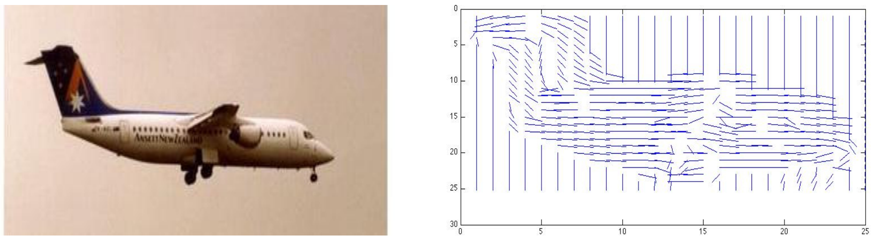

Fig. 1. An image with its corresponding orientation field. Here, $W = 25$.

3.2 HELO under Rotation Invariance

To get rotation invariance, we need to normalize both the sketch and the test image before computing HELO descriptor. We use two different normalization processes and then we compute two HELO descriptors, one for each normalization process. After that, we measure similarity by combining linearly partial distances. For normalization, we use principal component analysis (PCA) and polar coordinates (PC). We present a detailed description of this approach below:

- **Preprocessing**: This stage is similar to the previous one (Section 3.1), except that in this case the cropping operation is performed after the normalization process.
- **Orientation normalization**:
  - **Using PCA**: We compute a 2-d eigenvector v representing the axis with higher variance of the pixel (with value 1) distribution using PCA. We normalize both the sketch and the test image abstract representation rotating them $-\alpha$ degrees around their center of mass. Here, $\alpha = tan^{-1}(v_y/v_x)$.
  - **Using PC**: We transform both the test image and the sketch abstract representation into polar coordinates. In this case, two rotated images containing the same object become similar images only affected by an horizontal shifting.
- **Histogram Computation**: Exactly similar to the previous one (Sect. 3.1).

We compute similarity between a test image I and a sketch S combining PCA-based HELO and PC-based HELO. Let I_{PCA} and S_{PCA} be the PCA-based HELO descriptors computed over I and S, respectively. Let I_{PC} and S_{PC} be the correspondent PC-based HELO descriptors. The similarity measure $sm(I, S)$ is:

$$sm(I, S) = w_{PCA}L_1(I_{PCA}, S_{PCA}) + w_{PC}L_1(I_{PC}, S_{PC}) \qquad (4)$$

where, $w_{PCA}+w_{PC} = 1, w_{PCA} \geq 0, w_{PC} \geq 0$ and L_1 is the Manhattan distance.

Our proposal is configurable for working with or without rotation invariance. This is an advantage, because the rotation invariance requirement commonly depends on the application. For example, in the context of handwriting recognition rotation invariance may result in confusing the digit 6 with digit 9.

4 Experimental Evaluation

Considering that there is no standard benchmark for SBIR, we have developed one to evaluate different approaches, and to compare them with our proposal. For the test database, we have randomly selected 1326 images. We selected 1285 color images from Caltech101 [16], and we added 46 images containing castles and palaces. Many of the test images consist of more than one object or are cluttered images. For the query database we have chosen 53 images from the database. For each query image, a line-based sketch was hand-drawn. Thus, we have 53 sketches [1]. An example sketch and its corresponding target image appear in Fig. 2.

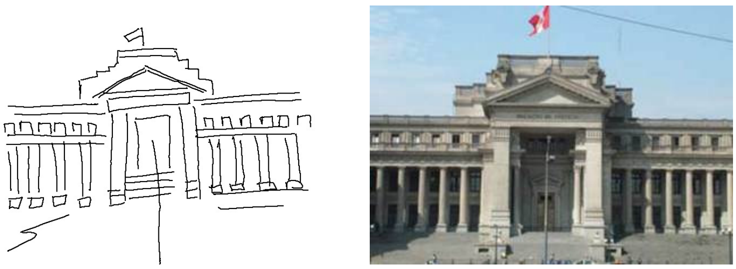

Fig. 2. A sketch on the left and its corresponding target image on the right

We compare our method with five others methods according to the state of the art. Four of these methods are: APAI [6], ST [1], HED [8], and EHD [4], which were implemented following the specification described in the corresponding papers. Additionally, we compare our method with the *histogram of distance distribution* (HDD) as explained in Section 2.

The evaluation of the methods was performed by querying each sketch for the most similar images and finding the target image rank. We called this rank *query rank*. For measuring our results, we use two metrics. First, we use *Mean of Query Rank* (MQR), for which the average of all *query ranks* is computed. Second, we use the *recall ratio* R_n, which shows the ratio to retrieve the target image in the best n-candidates. R_n is defined as follows:

$$R_n = \frac{\text{target images among first } n \text{ responses}}{\text{total number of queries}} \times 100 \qquad (5)$$

To evaluate translation, scale, and rotation robustness, we divide the experiments in two parts. First, we evaluate our method with sketches having different scale and position from the corresponding target images. Second, we evaluate our method with sketches that have been rotated by three different angles (30°, 60°, and 90°), having 212 sketches.

[1] http://prisma.dcc.uchile.cl/archivos_publicos/Sketch_DB.zip

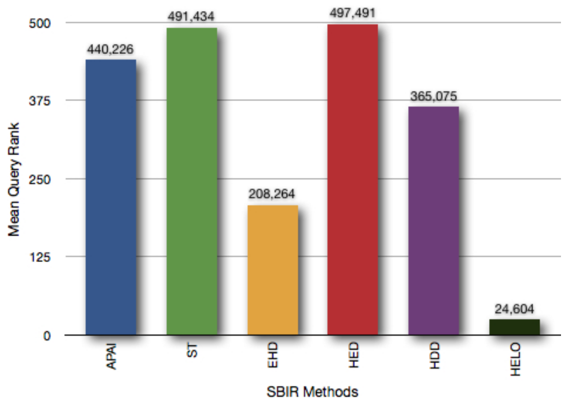

Fig. 3. Mean Query Rank of the evaluated methods

Our method needs three parameters to be specified, the histogram length (K), the number of horizontal and vertical blocks (W), and a threshold (th_{edge}). We fix $K = 72$, $W = 25$, and we set th_{edge} as 0.5 times the maximum image dimension. These parameter values were chosen experimentaly.

4.1 Translation and Scale Invariance Comparison

Fig. 3 shows the MQR for each evaluated method. We observe that our method is more robust than the other methods when scale and position changes exist. We achieve 24.60 as MQR value. This indicates that HELO needs to retrieve less than 25 images from the database to recover the target image. In comparison with the other methods, EHD is the closest to the ours with a significance difference. EHD achieves 208.26 as a MQR, i.e, EHD would require to retrieve almost 208 images to find the target one. Thus, our method improves effectiveness on recall over 8.4 times respect to the state of the art.

Fig. 4 shows the recall ratio graphic. This graphic shows that HELO outperforms the state of the art methods for any value of n. In addition, an example of image retrieval using HELO is shown in Fig. 5.

4.2 Rotation Invariance Comparison

First of all, we evaluate HELO descriptor using separately PCA and PC. Using PCA, we obtain a MQR value of 197.04. The principal axis is computed over the edge point distribution. However, a sketch is a simple rough hand-made drawing without details as those appearing in the target image. Due to these facts, the input sketch and the target image may have very different principal axis affecting the retrieval effectiveness. Using PC, sketches affected by different angular shifts have similar representations in polar coordinates. This is the reason for what PC gives a better MQR value (156.09) than that given by PCA. However, PC changes drastically the edge point distribution decreasing the discriminative power.

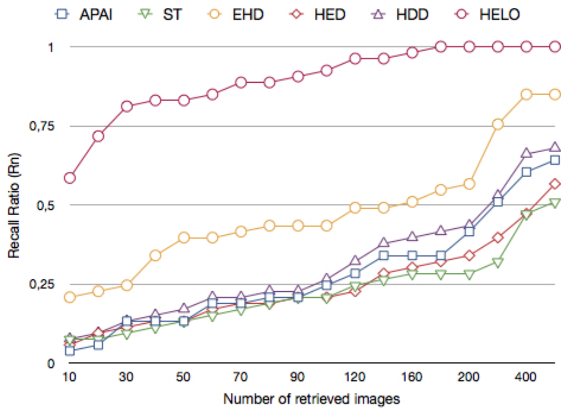

Fig. 4. Recall ratio graphic for the evaluated methods

Fig. 5. Example of the first six retrieved image using HELO. The first is the query

Therefore, to take advantage of each orientation normalization method we propose a linear combination of PC-based HELO and PCA-based HELO, that allows us to improve the retrieval effectiveness. Using our approach we get MQR value to 101.09. We described this method in the Section 3.2. We will refer to the combined-based HELO descriptor as HELO_R. To implement the HELO_R descriptor, we use $w_{PCA} = 0.3$ and $w_{PC} = 0.7$.

Fig. 6 shows the MQR for the evaluated methods under rotation distortions. Clearly, under this kind of changes, our proposal improves the effectiveness on recall over 2.6 times respect to the state of the art. To visualize how many images are needed to retrieve the target image, Fig. 7 shows the recall ratio graphics comparing the six evaluated methods.

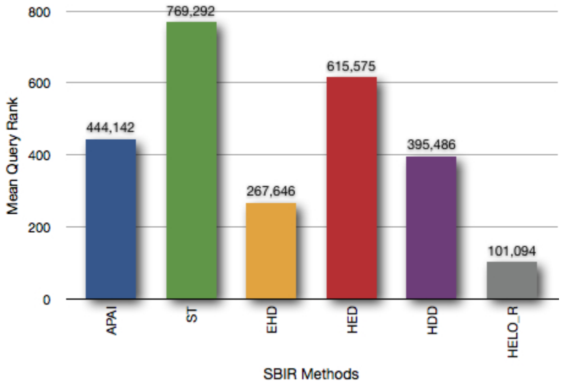

Fig. 6. Mean Query Rank of the evaluated methods under rotation invariance

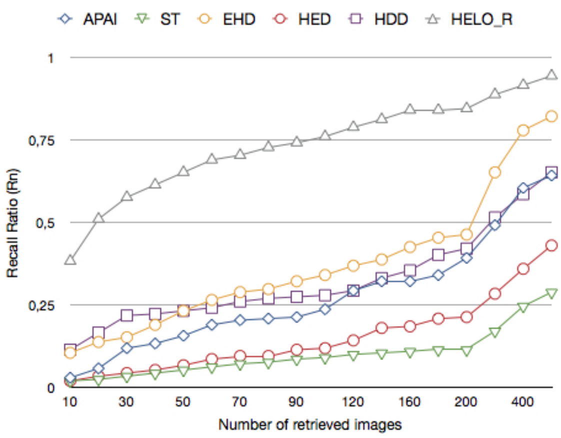

Fig. 7. Recall ratio graphic for the evaluated methods under rotation invariance

5 Conclusions

In this work, we have observed that SBIR is still an open problem, and that the current methods for SBIR do not work well enough. We have presented a novel method for SBIR that uses an efficient algorithm to compute a histogram of edge local orientations. First, we focused on SBIR under scale and translation transformations. Then, we extended our proposed approach to work under rotation invariance. We applied principal component analysis and polar coordinates to get orientation normalization.

Our achieved results show that HELO outperforms significantly the state of the art, improving recall over 8.4 times under scale and translation distortions, and over 2.6 times under rotation distortions. Furthermore, the query sketches do not need to be drawn with continuous strokes.

In our ongoing work, we are analyzing the performance of HELO under different values of its parameters. In addition, SBIR under orientation invariance must be studied in depth. For the future work, we will focus both on the rotation invariance problem and on multi-object sketch queries.

References

1. Eitz, M., Hildebrand, K., Boubekeur, T., Alexa, M.: A descriptor for large scale image retrieval based on sketched feature lines. In: Proc. of the 6th Eurographics Symposium on Sketch-Based Interfaces and Modeling, pp. 29–36 (2009)
2. Funkhouser, T., Min, P., Kazhdan, M., Chen, J., Halderman, A., Dobkin, D., Jacobs, D.: A search engine for 3d models. ACM Transactions on Graphics 22(1), 83–105 (2003)
3. Kato, T., Kurita, T., Otsu, N., Hirata, K.: A sketch retrieval method for full color image database-query by visual example. In: Proc. of the 11th IAPR International Conf. on Computer Vision and Applications, Conf. A: Pattern Recognition, pp. 530–533 (1992)
4. Sun Won, C., Kwon Park, D., Park, S.J.: Efficient use of MPEG-7 edge histogram descriptor. Electronic and Telecomunications Research Institute Journal 24, 23–30 (2002)
5. Del Bimbo, A., Pala, P.: Visual image retrieval by elastic matching of user sketches. IEEE Trans. on Pattern Analysis and Machine Intelligence 19(2), 121–132 (1997)
6. Chalechale, A., Naghdy, G., Mertins, A.: Sketch-based image matching using angular partitioning. IEEE Trans. on Systems, Man and Cybernetics, Part A: Systems and Humans 35(1), 28–41 (2005)
7. Bazen, A.M., Gerez, S.H.: Systematic methods for the computation of the directional fields and singular points of fingerprints. IEEE Trans. on Pattern Analysis and Machine Intelligence 24(7), 905–919 (2002)
8. Jain, A., Vailaya, A.: Image retrieval using color and shape. Pattern Recognition 29, 1233–1244 (1996)
9. Martínez, J.M.: MPEG-7: Overview of MPEG-7 description tools, Part 2. IEEE MultiMedia 9(3), 83–93 (2002)
10. Pu, J., Ramani, K.: A 3d model retrieval method using 2d freehand sketches. In: Sunderam, V.S., van Albada, G.D., Sloot, P.M.A., Dongarra, J. (eds.) ICCS 2005. LNCS, vol. 3515, pp. 343–346. Springer, Heidelberg (2005)
11. Belongie, S., Malik, J., Puzicha, J.: Shape context: A new descriptor for shape matching and object recognition. In: Proc. of the 2000 Neural Information Processing Systems Conference, pp. 831–837 (2000)
12. Davies, E.R.: The effect of noise on edge orientation computations. Pattern Recognition Letters 6(5), 315–322 (1987)
13. Canny, J.: A computational approach to edge detection. IEEE Trans. on Pattern Analysis and Machine Intelligence 8(6), 679–698 (1986)
14. Gatos, B., Pratikakis, I., Perantonis, S.: Hybrid off-line cursive handwriting word recognition. In: Proc. of the 8th International Conference on Pattern Recognition, pp. 998–1002 (2006)
15. Gonzales, R., Woods, R.: Digital Image Processing, 3rd edn. Prentice Hall, Englewood Cliffs (2008)
16. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. In: Proc. of the 2004 Conference on Computer Vision and Pattern Recognition, Workshop on Generative-Model Based Vision, p. 178 (2004)