SKETCH BASED IMAGE RETRIEVAL USING A SOFT COMPUTATION OF THE HISTOGRAM OF EDGE LOCAL ORIENTATIONS (S-HELO)

Jose M. Saavedra

Computer Vision Research Group
Orand S.A.
Estado 360 Of 702, Santiago, Chile

ABSTRACT

This paper introduces S-HELO (Soft-Histogram of Edge Local Orientations), an outperforming method for describing images in the context of sketch based image retrieval (SBIR). This proposal exploits the advantages provided by the HELO descriptor for describing sketches, and improves significantly its performance by using a soft computation of local orientations and taking into account spatial information. We experimentally demonstrate that a soft computation process together with a local estimation of orientations are very suitable for describing sketches in the context of image retrieval. Indeed, our results show that S-HELO significantly outperforms not only HELO but also classical orientation-based descriptors as HOG. We also show that S-HELO performs very close to the optimal when what we want to retrieve are target images. Moreover, our proposal also shows an outstanding performance for similarity search, i.e., retrieving images that belong to the same category of the query sketch.

Index Terms— Sketch based image retrieval, sketch descriptors, orientation histograms.

1. INTRODUCTION

An alternative for querying in an image retrieval system is by simply drawing what the user has in mind, which further represents an intuitive way of communication between a user and a CBIR system. This kind of query leads to the sketch based image retrival problem (SBIR), that is also supported by the emerging touch screen based technology, allowing users to make a sketch directly on the screen. According to Hu et al. [1], "a key challenge in SBIR is overcoming the ambiguity inherent in sketch". In fact, a sketch exhibits different kinds of variations based on non-rigid transformations. In addition, sometimes a sketch produced by an user will look very rough because of poor drawing skills or limited time for drawing.

One salient characteristic of a sketch is the stroke orientation. Orientation is a characteristic that has been exploited widely in the computer vision community showing outperforming results in tasks like object recognition and object categorization [2, 3, 4, 5]. In the SBIR context, this characteristic has also been exploited by Saavedra et al.[6] who proposed the HELO descriptor outperforming significantly the performance of traditional SBIR descriptors.

HOG and HELO descriptors are techniques to compute orientation histograms. However they are different w.r.t. how the orientations are estimated. HOG follows a pixel-wise strategy and HELO, a cell-wise strategy. The HELO's strategy seems to be very appropriate for representing sketch-like images since sketches are sparse by nature. Indeed, a pixel-wise strategy will produce many zeros in the final histogram, which may affect the effectiveness. However, HELO, as proposed originally [6], is computed in a hard manner and does not take into account any kind of spatial information.

Thereby, our contribution in this work is to propose S-HELO (Soft-Histogram of Edge Local Orientations) which significantly improves HELO. First, S-HELO computes cell orientations in a soft manner using bilinear and tri-linear interpolation according to two processing levels (a cell-based level and a block-based level). Second, S-HELO takes into account spatial information by dividing the image into blocks and computing a soft orientation histogram for each block. This division is inspired by the Spatial Pyramid Matching [7]. Third, S-HELO computes an orientation histogram using weighted votes from the estimated cell orientations. Although, some of these ideas are inspired on the soft computation applied in HOG, the idea of soft computing together with a local estimation of orientations is novel. We demonstrate that S-HELO outperforms significantly HOG descriptor as well HELO descriptor in the context of SBIR.

This document has been organized as follow. Section 2 describes the state of the art related to the sketch based image retrieval. In Section 3, we discuss our proposal in detail. Section 4 describes our experiment settings and discusses the achieved results using S-HELO with respect to state-of-the-art techniques. Finally, in Section 5, we present our conclusions and future work.

Partially funded by CONICYT-PAI Nro 781204025, Chile.

2. RELATED WORK

Classical methods are based on building a frequency histogram of orientations [8] or a histogram representing the distribution of edge pixels [9]. Elastic contours [10] are another alternative, where a sketch is represented by a parametric curve that is then strained or bent in order to fit the border of an object. Others approaches transform the input sketch into a regular image, with color an texture [11, 12], to finally apply a classical CBIR strategy.

In the case of local features, they are commonly aggregated using the Bag of Features (BoF) model [13, 14, 1]. In this vein, Eitz et al. [13] proposed two techniques based on *SIFT* [2] and *Shape Context* [15]. Hu et al. [14, 1] also proposed a BoF approach that transforms a sketch into a gradient field (GF) image. The GF images are used to compute HOG descriptors in different scales. After that, a BoF model is applied to form a frequency histogram. This method requires solving a sparse linear equation system to obtain the GF image. The number of variables in the equation is the order of the size of the input image.

Yan Cao et al. [16] also presented a method based on the Chamfer Distance [17]. Even though the authors present a technique to deal with large database based on the inverted index structure, they do not show how to deal with geometric variations.

The majority of SBIR methods are based on histograms of orientations either to compute a global representation or a local one. In the context of histograms of orientations, HOG [4] seems to be the favorite descriptor in the community of computer vision. However, when we deal with sketches, the result of applying HOG may be a sparse description as sketches are drawing by simple strokes. This fact, may drastically affect the retrieval effectiveness. A different approach to compute a histogram of orientations was presented by Saavedra et al. [6]. They proposed HELO (histogram of edge local orientations). In this case, the orientation histogram is formed by local orientations. These local orientations are estimated by grouping pixels in cells and determining just one representative orientation for each cell. In addition, HELO sets the number of cells as fixed, which permits to tackle size variations in a natural way. In contrast, HOG uses a fixed cell size, which requires images to be resized to a standard size.

Therefore, the contribution of this paper is to propose an improved descriptor inspired by both the HELO descriptor, to compute orientations in a local way, and the HOG descriptor, to apply a soft computation for estimating orientations and forming the histogram. We call our proposal S-HELO from Soft-Histogram of Edge Local Orientations.

3. S-HELO IN ACTION

Different from HOG, this proposal computes one orientation using nearby pixels. This idea is inspired by the HELO descriptor [6]. In addition, we propose the following improvements over HELO: (a) Compute a representative orientation for each cell, in which the image will be divided, in a soft manner using a bilinear interpolation, (b) Use weighted votes. Each orientation of a pixel is weighted by the magnitude of its corresponding gradient, and (c) Take into account spatial information dividing the image into non-overlapping blocks and computing a soft histogram of orientations for each block.

Regarding our improvements presented above, S-HELO is composed of three stages: 1) cell orientation estimation, 2) local histogram computation, and 3) S-HELO composition.

3.1. Cell Orientation Estimation

Let I be an image with M rows and N columns. We divide I using a $W \times W$ grid. Suppose that we label each cell using the pair (p, q), $p, q = 0..W-1$, where p refers to rows, and q to columns. To compute a representative orientation for each cell, we process each pixel (i, j) of I as follows:

1. We determine the four nearest cells to the pixel (i, j). These are specified by (l_pos, n_pos), (r_pos, n_pos), (l_pos, s_pos), and (r_pos, s_pos). The indices of the cells are computed as follow:

$$\begin{array}{rr} p' = (j/N) * W, & q' = (i/M) * W \\ l\_pos = \lfloor (p' - 0.5) \rfloor, & n\_pos = \lfloor (q' - 0.5) \rfloor \\ r\_pos = \lfloor (p' + 0.5) \rfloor, & s\_pos = \lfloor (q' + 0.5) \rfloor \end{array} \quad (1)$$

2. Compute a weight value for each of the four nearest cells with respect to the pixel (i, j). This weight is computed inversely w.r.t. the distance of the pixel (i, j) to the center of each cell. We compute this distance in the x-dimension and y-dimension. In the first case, we use the p' value, and in the second case, we use the q' value. Below, we show how this works using p'.

   - Compute the distance of p' to the most left side of the underlying cell:

$$dist_p = p' - \lfloor p' \rfloor \quad (2)$$

   - If $(dist_p < 0.5)$:

$$l\_weight = 0.5 - dist_p \quad (3)$$
$$r\_weight = 1 - l\_weight \quad (4)$$

   - If $(dist_p >= 0.5)$:

$$r\_weight = dist_p - 0.5 \quad (5)$$
$$l\_weight = 1 - r\_weight \quad (6)$$

   - The values for s_weight and n_weight are computed in a similar way using the q' value.

3. Compute the representative orientation and magnitude for each cell. Following the squared gradient approach used by the HELO [6], we compute:

$$\begin{bmatrix} G_{sj} \\ G_{si} \end{bmatrix} = \begin{bmatrix} G_j^2 - G_i^2 \\ 2G_jG_i \end{bmatrix}, \quad (7)$$

where $[G_j, G_i]^T$ is the Sobel gradient of the pixel (i, j) and $[G_{sj}, G_{si}]^T$ is the corresponding squared gradient. In order to produce a soft estimation, we compute the cell orientation visiting all the pixels and aggregating the corresponding gradient components. To this end, we define D as a $W \times W$ matrix that will represent the aggregation of the first components of the squared gradient for each cell. We also define A as a $W \times W$ matrix that will represent the aggregation for the second component.

Apply the following process to each pixel (i, j) in I. Let (x, y) be a cell affected by a pixel (i, j), where $(x, y) \in \{(l\_pos, n\_pos), (r\_pos, n\_pos), (l\_pos, s\_pos), (r\_pos, s\_pos)\}$ and (w_x, w_y) be the corresponding weights of x and y as computed in step 2. Using the weight $\alpha = w_x * w_y$, we increment A and D for each affected cell (x, y) as follow:

$$A_{x,y} \quad += \quad \alpha * 2 * sin(\theta) * cos(\theta) \quad (8)$$

$$D_{x,y} \quad += \quad \alpha * (cos^2(\theta) - sin^2(\theta)) \quad (9)$$

with θ being the gradient orientation of (i, j) and α being its corresponding magnitude.

4. Finally, the orientation $\beta_{p,q}$ of a cell (p, q) is computed as:

$$\beta_{p,q} = 0.5 * atan2(A_{p,q}, D_{p,q}) \quad (10)$$

where $-\pi/2 \leq \beta_{p,q} \leq \pi/2$. In addition, the magnitude of a cell is obtained by summing up all the pixel gradient magnitudes that affect that cell.

3.2. Local Histogram Computation

We take into account spatial information by dividing the image into $B \times B$ non-overlapping blocks. For each block we form a K-bin orientation histogram w.r.t cell orientations. We compute the histograms by tri-linear interpolation. We estimate the interpolation weights in a similar way as in the previous step. In addition, to form a orientation histogram, a weighted vote is applied using the cell magnitudes.

3.3. S-HELO Composition

In this last step, we normalize each local histogram to the unit and the final descriptor is formed as the concatenation of all the normalized local histograms. Therefore, the descriptor size results to be equal to $B \times B \times K$.

Method	EHD	HELO	HOG	S-HELO
MQR	208.264	24.604	22.359	4.09

Table 1. Mean Query Rank for SBIR approaches. The lower the MQR value, the better the performance.

4. EXPERIMENTAL EVALUATION

We compare our results using the dataset proposed by Saavedra et al. [6]. This dataset is composed of 1326 test and 53 hand-drawn sketches. Each sketch resembles a target image in the dataset. For similarity search evaluation, we also assess our method using the dataset proposed by Hu and Collomosse [1] that containg 14660 images and 330 query sketches.

In order to be fair in the comparative evaluation, we use the same evaluation strategy as that used by Saavedra et al. [6]. The evaluation was performed by querying each sketch for the most similar images and finding the target image rank (the position of the target image in the ranking). We called this rank *query rank*. Ideally, the target image must appear in the rank 1. For measuring the results, we use two metrics: *the mean query rank* (MQR) and *the recall ratio* Rn [6].

Our results show a noticeable improvement in the target image retrieval. Table. 1 shows the high effectiveness (MQR=4.09) of S-HELO. Moreover, from Fig. 1 we can see that S-HELO allows us to get the 96.2% of target images with only retrieving the first 9 images. This represents a significant improvement since HOG, under the same condition, retrieves only 83% of target images. In addition, if we would be interested in the first response our method achieves an accuracy of 60.4%, while the HOG approach achieves 39.6%.

To demonstrate the goodness of our proposal, we also have conducted experiments aiming to evaluate its performance for **similarity search**. This evaluation relies on the *precision-recall curve*, widely used in the area of information retrieval [18]. In this case, our method also achieves a noticeable increment in precision with respect to HOG and HELO (see Fig. 2). S-HELO achieves a MAP of 0.277 while HOG and HELO achieves 0.205 and 0.143, respectively.

We have also compared S-HELO with the GF-HOG descriptor [14, 1] using the Hu's dataset. Here, S-HELO achieves a MAP of 0.124 that is competitive with the Hu's result (MAP=0.122). Moreover, our method is simpler and does not need to compute a gradient field image.

To illustrate our results, we show in Fig. 3 and Fig. 4 examples of the S-HELO responses.

4.1. Parameters

We chose the parameters of S-HELO and HOG experimentally. We run both methods with different values of their parameters and pick up those that showed the best performance. The results of the other methods was obtained directly from the paper of Saavedra et al. [6]. In Table 2 we summarize

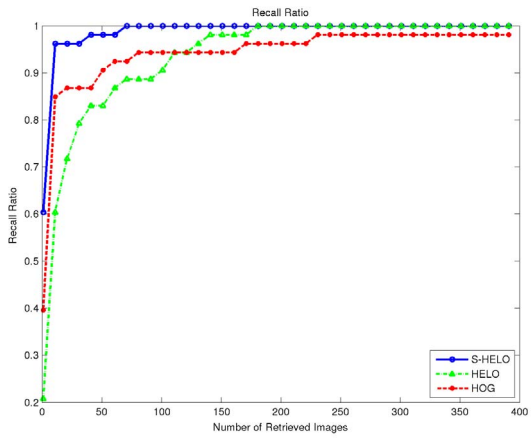

Fig. 1. Recall-Ratio graphic, comparing S-HELO, HELO and HOG approaches.

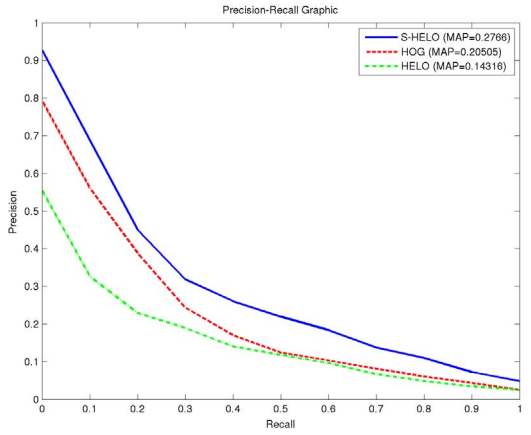

Fig. 2. Precision-recall comparing S-HELO, HOG and HELO. This graphic shows the precision in obtaining relevant images (images from the correct category).

Fig. 3. Two examples of the retrieval performance of S-HELO. In these examples we are interested in retrieving a target image. The figure shows the first three responses.

the used parameters. We can see from Table 2 that the size of HOG descriptor is lower than that of the S-HELO. This could seem unfair. We tried with similar sizes, but the performance of HOG was worst. For instance, using a cell size of 14×14 with a block size of 3×3 produces a 1296-size descriptor, the same that the S-HELO descriptor's size. However, using these parameter values, HOG achieves a MQR=35.81, which is lower than the MQR using the selected parameters for HOG.

Fig. 4. These examples show the effectiveness of our proposal for retrieving similar images w.r.t. a query sketch. The figure shows the first three responses.

Table 2. Parameters of HOG and S-HELO

HOG	S-HELO
Image size: 200×200	.
Cell size: 18×18	# of cells (W): 25×25
Block size: 5×5	# of blocks (B): 6×6
Histogram size: 9	Histogram size (K): 36
Descriptor Size: 900	1296

5. CONCLUSIONS

We have presented an outperforming method for the SBIR . Our method is based on a soft computation of cell orientations. Different from HOG we estimate only one orientation for each cell in which the image is divided. Our results show that this method seems to be the best option to compute a orientation histogram in the context of SBIR. S-HELO, achieves a mean query rank very close to the optimal value (S-HELO's MQR=4.09), which represents a very significant increment in precision when we compare it with the effectiveness achieved by the state-of-the-art methods.

In addition, we show the goodness of our method for similarity search. If we evaluate the performance of our method in retrieving images from the same class S-HELO achieves a MAP of 0.277, while HOG achieves only 0.205.

Our ongoing work is focused on evaluating S-HELO in larger datasets. We also are working in extending S-HELO to deal with sketches containing multiple objects and to be invariant to rotation.

6. REFERENCES

[1] Rui Hu and John Collomosse, "A performance evaluation of gradient field hog descriptor for sketch based image retrieval," *Computer Vision and Image Understanding*, vol. 117, no. 7, pp. 790–806, July 2013.

[2] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[3] Krystian Mikolajczyk and Cordelia Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.

[4] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*. 2005, pp. 886–893, IEEE Computer Society.

[5] Pedro Felzenszwalb, McAllester David, and Ramanan Deva, "A discriminatively trained, multiscale, deformable part model," in *International Conference on Computer Vision and Pattern Recognition*, 2008.

[6] Jose Saavedra and Benjamin Bustos, "An improved histogram of edge local orientations for sketch-based image retrieval," in *Pattern Recognition*, vol. 6376 of *Lecture Notes in Computer Science*, pp. 432–441. 2010.

[7] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, 2006, vol. 2, pp. 2169–2178.

[8] Chee Sun Won, Dong Kwon Park, and Soo-Jun Park, "Efficient use of MPEG-7 edge histogram descriptor," *Electronic and Telecomunications Research Institute Journal*, vol. 24, pp. 23–30, 2002.

[9] Abdolah Chalechale, Golshah Naghdy, and Alfred Mertins, "Sketch-based image matching using angular partitioning," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 35, no. 1, pp. 28–41, 2005.

[10] Alberto Del Bimbo and Pietro Pala, "Visual image retrieval by elastic matching of user sketches," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 121–132, 1997.

[11] Tao Chen, Ming-Ming Cheng, Ping Tan, Ariel Shamir, and Shi-Min Hu, "Sketch2photo: internet image montage," *ACM Transactions on Graphics*, vol. 28, no. 5, pp. 124:1–124:10, Dec. 2009.

[12] Mathias Eitz, Kristian Hildebrand, Tamy Boubekeur, and Marc Alexa, "Photosketch: a sketch based image query and compositing system," in *SIGGRAPH 2009: Talks*, 2009, SIGGRAPH '09, pp. 60:1–60:1.

[13] Mathias Eitz, Kristian Hildebrand, Tamy Boubekeur, and Marc Alexa, "Sketch-based image retrieval: Benchmark and bag-of-features descriptors," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 11, pp. 1624–1636, 2011.

[14] Rui Hu, M. Barnard, and J. Collomosse, "Gradient field descriptor for sketch based retrieval and localization," in *17th IEEE International Conference on Image Processing (ICIP), 2010*, 2010, pp. 1025 –1028.

[15] 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, Apr. 2002.

[16] Yang Cao, Changhu Wang, Liqing Zhang, and Lei Zhang, "Edgel index for large-scale sketch-based image search," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*. 2011, pp. 761–768, IEEE Computer Society.

[17] Gunilla Borgefors, "Hierarchical chamfer matching: A parametric edge matching algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 6, pp. 849–865, Nov. 1988.

[18] Ricardo Baeza-Yates and Berthier Ribeiro-Neto, *Modern Information Retrieval: Concepts and Technology behind Search*, Addison-Wesley Professional, USA, second edition, 2011.