

# CONTENT BASED IMAGE RETRIEVAL USING SALIENT ORIENTATION HISTOGRAMS

*Andrea Manno-Kovacs*

Institute for Computer Science and Control, MTA SZTAKI, Budapest, Hungary

## ABSTRACT

Content-aware image retrieval is a very important topic nowadays, when the amount of digital image data is highly increasing. Existing sketch based image retrieval (SBIR) systems perform at a reduced level on real life images, where background data may distort image descriptors and retrieval results. To avoid this, a preprocessing step is introduced in this paper to distinguish between foreground and background, using integrated saliency detection. To build the descriptor only on the most relevant pixels, orientation feature is extracted at salient Modified Harris for Edges and Corners (MHEC) keypoints using an improved edge map, resulting in a Salient Orientation Histogram (SOH). The proposed SBIR system is also augmented with a segmentation step for object detection. The method is tested on the THUR15000 database, containing random internet images. Image retrieval and object detection both give promising results compared to other state-of-the-art methods.

**Index Terms**— Direction selectivity, modified Harris for edges and corners, saliency detection, SBIR

## 1. INTRODUCTION

Content-aware image retrieval is a very important topic nowadays with the constantly increasing amount of digital image data. Outline sketches have recently been shown to be more comfortable for retrieval than a complete image, as sketch based image retrieval (SBIR) expects simpler descriptors resulting in faster comparison and retrieval.

Descriptors can be grouped into global and local types. While the former includes information of the whole image [1], the latter concentrates only on a small image part [2]. Recently published SBIR systems [3, 4] employs local features, as global ones are not handling affine variations well, and the fact that fine details of the drawing are often missing.

Existing SBIR systems are mainly tested on image databases without significant background information. However, randomly selected internet images often contain a lot of background data with varying texture and color, which can influence the image descriptors and make the comparisons more

challenging. To avoid this, a preprocessing step can help to distinguish between foreground and background, which increases the importance of saliency detection [4, 5, 6, 7]. However, the dimension of a salient area description can still be very high, thus further reduction is needed. Interest point detectors, like Harris [8] emphasize relevant structures in the image. Thus, if the local descriptors are calculated at interest point locations, the extracted salient region information can be reduced while retaining their relevance. Modified Harris for Edges and Corners (MHEC) was proposed earlier by the author [9] for efficient image segmentation, and the method's strong ability for object detection was also shown previously [10], supporting its capability of holding efficient structure and content information for image comparisons and retrieval.

Orientation as a descriptor has already been introduced in earlier SBIR systems [1], moreover many improvements of the Histogram of Oriented Gradients (HoG) [11] were published over the past years [12, 3]. The original HoG calculated the histogram for the whole image. Improved adaptations of HoG for SBIR systems are mostly using Canny edge maps [12, 3] with orientation histograms calculated on pixels of the Canny edge map [12] or randomized pixels [3]. Following this technique, the background texture may create false edges in the Canny edge map and the keypoint selection could include background hits. Both of them may cause the distortion of the orientation histogram and reduced retrieval accuracy.

In this paper, an improved SBIR with object detection is presented for natural images. First, the salient region is located by a texture distinctiveness map [5], followed by an MHEC feature point detection step to extract features from relevant areas. Moreover, instead of the Canny edge map, an improved edge map is introduced for orientation calculation and object detection. A salient orientation histogram (SOH) is built using the salient keypoints and the improved edge map. The proposed method is tested on the THUR15000 database [13] for image retrieval and detection purposes.

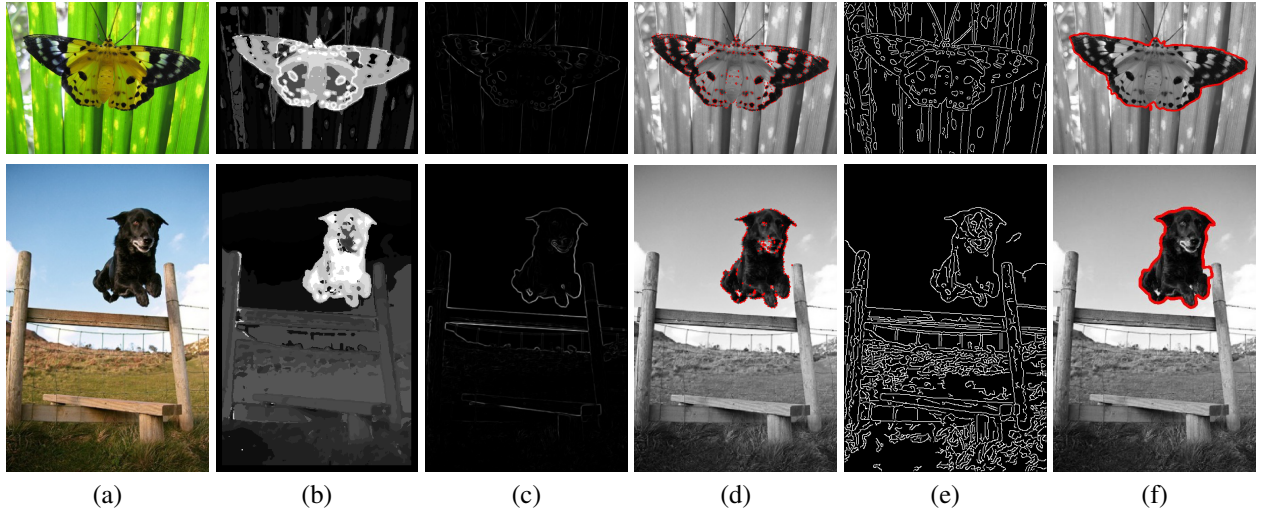
## 2. SALIENT KEYPOINT DETECTION

### 2.1. Texture distinctiveness

Statistical texture distinctiveness model was introduced in [5] and it used a rotational-invariant neighborhood-based textural representation to learn representative texture atoms for sparse

---

This work was partially funded by the Government of Hungary through a European Space Agency (ESA) Contract under the Plan for European Co-operating States (PECS) and the Hungarian Industry Incentive Scheme. E-mail: andrea.manno-kovacs@sztaki.mta.hu



**Fig. 1.** Sample images for illustrating the proposed SBIR system: (a) the original image; (b)  $S$  texture distinctiveness map; (c)  $R_{\text{mod}}$  improved edge map; (d)  $\text{MHEC}_S$  point set; (e) Canny edge map and (f) the detection result achieved by DHVFC.

texture model. Statistical texture distinctiveness measures the uniqueness of the atoms and the relationships between them, by constructing a weighted graph model. The  $S(x, y)$  texture distinctiveness map (see Fig. 1(b)) quantifies the expected relative distinctiveness of each texture, incorporating high-level aspects, such as spatial location of regions relative to the image center. For detailed description, see [5].

The  $S$  map is thresholded with Otsu’s method [14] to extract highly distinctive image parts. By selecting the region of maximal area, an automatic  $S_{\text{ROI}}$  is initialized. In Figure 1(b) the  $S_{\text{ROI}}$  areas can be well identified in both samples.

## 2.2. MHEC interest point set and improved edge map

If the aim is to emphasize object contours, the Modified Harris for Edges and Corners (MHEC) detector is shown to be an efficient tool [15, 10]. The method adapts the Harris matrix and uses its  $\lambda_1$  and  $\lambda_2$  eigenvalues in the following modification of the  $R$  characteristic function [15]:

$$R_{\text{mod}} = \max(\lambda_1, \lambda_2). \quad (1)$$

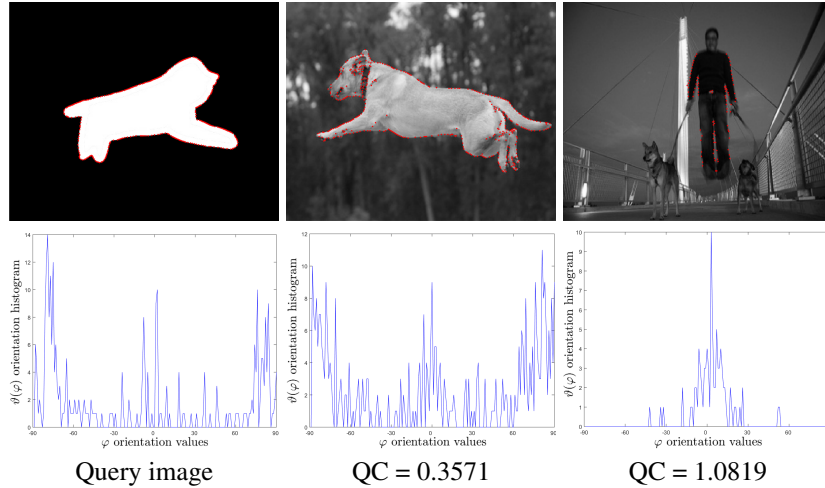
The introduced modification has the ability to emphasize edge and corner regions in a balanced manner. MHEC interest points ( $p_i$ ) are selected as the local maxima of the  $R_{\text{mod}}$ . The point set is able to represent points-of-interest inside the salient  $S_{\text{ROI}}$  region, therefore, the importance of the extracted pixels is two-fold: while the  $S$  ensures the distinctiveness of the selected texture; MHEC points represent potential object contours inside the  $S_{\text{ROI}}$  area. Features extracted for the MHEC points are able to describe the corresponding object more efficiently. The point subset inside the  $S_{\text{ROI}}$  area is marked by  $\text{MHEC}_S$ , samples are given in Fig. 1(d).

Besides the MHEC point set, the calculated  $R_{\text{mod}}$  map emphasizes the object contours, therefore it can be applied for gradient calculation and for the extraction of the orientation histogram. By using the  $R_{\text{mod}}$  map, a more specific contour map is obtained, than other traditional edge maps (e.g. Canny) used in previous SBIR systems [3].

The examples of Figure 1 illustrate the main contribution of the paper: while the Canny edge maps in earlier SBIR systems often include false edges which can severely distort the orientation histograms (Fig. 1(e)), the improved edge map (Fig. 1(c)) together with the salient point set (Fig. 1(d)) is able to sample the most relevant pixels of the image and extract essential orientation information. For earlier methods, the parallel edges of the background in the butterfly image and the presence of other objects in the dog image may influence the orientation statistics, leading to distorted histograms.

## 3. IMAGE RETRIEVAL AND SEGMENTATION

The members of the  $\text{MHEC}_S$  point set include the relevant feature points of the most distinctive image area. Therefore, the extracted salient orientation histogram (SOH) for the  $\text{MHEC}_S$  point set can give an improved object description. To obtain the SOH, the local gradient orientation density [16] is analyzed around a feature point to find the main direction. Instead of histogram binning [11], the exact main direction is calculated for the members of the point set and each count is also weighted by the gradient magnitude at the pixel. Let the gradient vector be denoted by  $\nabla g_i^R$  with  $\|\nabla g_i^R\|$  magnitude and  $\varphi_i^\nabla$  orientation for the  $i^{\text{th}}$  point, where  $R$  upper index refers to the gradient calculation on the  $R_{\text{mod}}$  improved edge map. Then, if  $W_n(i)$  defines the  $n \times n$  neighborhood of the point (where  $n$  depends on the resolution), the weighted



**Fig. 2.** Image comparison based on  $\vartheta(\varphi)$  SOH function (Eq. 3) calculated for keypoints marked by red; and the QC distance metric (Eq. 4). Smaller QC value indicate more similar image to the query image.

density of  $\varphi_i^\nabla$  is as follows:

$$\lambda_i(\varphi) = \frac{1}{N_i} \sum_{r \in W_n(i)} \frac{1}{h} \cdot \|\nabla g_r^R\| \cdot k\left(\frac{\varphi - \varphi_r^\nabla}{h}\right), \quad (2)$$

with  $N_i = \sum_{r \in W_n(i)} \|\nabla g_r^R\|$  and  $k(\cdot)$  kernel function with  $h$  bandwidth parameter. From this orientation histogram, the main orientation for  $i^{\text{th}}$  feature point is defined as:  $\varphi_i = \operatorname{argmax}_{\varphi \in [-90, +90]} \{\lambda_i\}$ .

After calculating the dominant direction for all  $K$  feature points, the salient orientation histogram (SOH) function  $\vartheta$  is:

$$\vartheta(\varphi) = \frac{1}{K} \sum_{i=1}^K H_i(\varphi), \quad (3)$$

where  $H_i(\varphi)$  is a logical function: 1, if  $\varphi_i = \varphi$ ; 0 otherwise.

If the SOH is calculated for both the query object and the members of the test database (see Fig. 2), the distance of the computed SOHs has to be measured. The detailed study of [12] showed that for orientation histograms, the  $\chi^2$  metric gives the best results. However, since then the Quadratic-Chi (QC) histogram distance family was published [17], introducing an improved technique for histogram comparison. The QC histogram distance for  $P$  and  $Q$  histograms is defined as:

$$QC_m^A(P, Q) = \sqrt{\sum_{ij} \left( \frac{P_i - Q_i}{(\sum_c (P_c + Q_c) A_{ci})^m} \right) \left( \frac{P_j - Q_j}{(\sum_c (P_c + Q_c) A_{cj})^m} \right) A_{ij}}, \quad (4)$$

with an  $A$  non-negative symmetric bounded bin-similarity matrix and an  $m$  normalization factor.

The cross-bin  $\chi^2$ -like normalization of the QC method reduces the effect of large bins and it is less sensitive to light changes and shape deformations than  $\chi^2$ . The retrieval tests

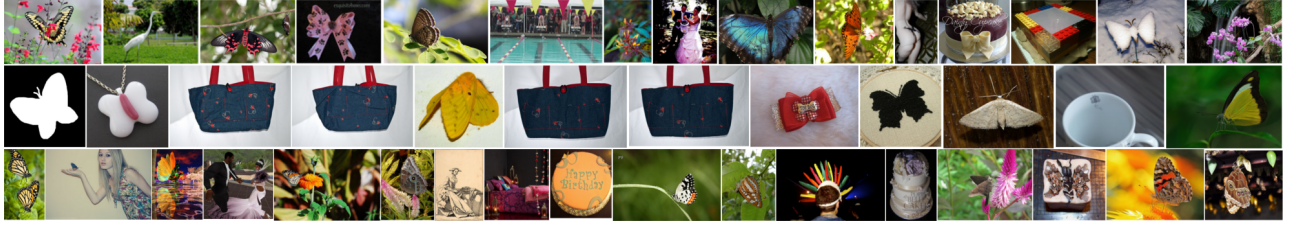
of the author have also confirmed QC's superiority over  $\chi^2$  metric, thus QC was applied in the evaluation process. After calculating the QC distance metric for each member of the database, the closer the distance, the more relevant image in the database, which is shown in Figure 2.

Another advantage of the improved  $R_{\text{mod}}$  map (Eq. 1) is its applicability in the external force of parametric active contours for object contour detection, shown in the previous work of the author [10]. It can be applied for segmenting the retrieved images in an automatic system. As a novelty, this paper proposes to integrate the texture distinctiveness information ( $S_{ROI}$ ) through the MHEC<sub>S</sub> point set to achieve an automatic initialization step. As a next step, the active contour is initialized as the convex hull of the MHEC<sub>S</sub> point set. Finally, applying the Directional Harris based Vector Field Convolution (DHVFC) [10] parametric active contour method, an iterative detection is performed (see Fig. 1(f)).

## 4. EXPERIMENTS

The THUR15000 database was introduced in [13] as the largest dataset for content based image retrieval using randomly selected internet images, downloaded from Flickr for 5 keywords: "butterfly", "coffee mug", "dog jump", "giraffe" and "plane". The database also contains salient regions marked at pixel accuracy, for images where such region exists. Therefore, it can also be used for evaluating the performance of segmentation methods. In the first part of the evaluation the database is used to evaluate the SOH descriptor, then the shape based image segmentation is tested.

In the first part, the proposed SOH based SBIR system is compared to Flickr and the SHOG based SBIR [3] (retrieval results are given in [13]). Table 1 shows the retrieval results



**Fig. 3.** Best retrieval results of different methods for keyword "butterfly" (Flickr) and query image in the second row (SHOG [3] and proposed SOH). First row: Flickr; second row: query image and SHOG; third row: the proposed SOH.

TPR(%)	Butterfly		Coffee mug		Dog jump		Giraffe		Plane		Average	
	T50	T100	T50	T100	T50	T100	T50	T100	T50	T100	T50	T100
Flickr	28	28	58	51	56	55	30	25	44	48	<b>41.4</b>	<b>43.2</b>
SHoG [3]	36	40	82	78	74	73	18	18	90	91	<b>60.0</b>	<b>60.0</b>
Proposed	48	50	82	82	84	87	54	59	54	64	<b>64.4</b>	<b>68.4</b>

**Table 1.** True positive retrievals for THUR15000 keywords.

for the best 50 and the best 100 retrieved images for each category. An image is considered as true positive if it contains a target object specified by the keywords. According to the results, the proposed SOH-based retrieval looks promising, the achieved true positive ratio is the highest in almost all categories and the method reached the highest average relevance on the whole THUR15000. The best retrieval images for all the three methods are shown in Figure 3.

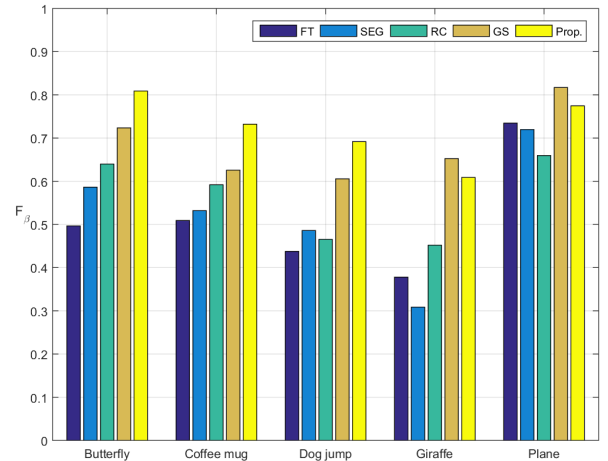
In the second part, the images with pixel accurate ground truth are segmented, by applying the proposed automatic saliency-based initialization. The segmentation method is compared to other state-of-the-art techniques: FT [7], SEG [6], RC [4] and GS [13]. To weigh precision more than recall (advised in [7]), the  $F_\beta$  is calculated with  $\beta^2 = 0.3$ :

$$F_\beta = \frac{(1 + \beta^2)Precision \times Recall}{\beta^2 \times Precision + Recall}. \quad (5)$$

Figure 4 shows that the proposed automatic detection technique is able to perform at high accuracy, achieving the highest performance in the majority of the keywords and for the average  $F_\beta$  (Table 2) on the whole database. (Segmentation samples are presented in Figure 1(f)).

## 5. CONCLUSION

In this paper, a novel SBIR system is introduced, using a salient keypoint based orientation histogram (SOH). The proposed method first extracts the salient image region based on texture distinctiveness, followed by a Modified Harris for Edges and Corners (MHEC) interest point detection. This way the most relevant pixels of the image are selected to build an orientation histogram on an improved edge map, instead of applying Canny edge map like earlier SBIR systems.



**Fig. 4.**  $F_\beta$  evaluation results for THUR15K database for different segmentation methods: FT [7], SEG [6], RC [4], GS [13] and the proposed method.

	FT	SEG	RC	GS	Prop.
Aver. $F_\beta$	0.51	0.53	0.56	0.68	<b>0.72</b>

**Table 2.** Average  $F_\beta$  values of segmentation methods.

The edge map is also adapted for segmentation. Overall, the proposed descriptor achieves high performance on the THUR15000 dataset, and it also provides an efficient object detection method. Future work will investigate the improved integration of saliency in SBIR systems.

## 6. REFERENCES

- [1] Abdolrah Chalechale, Golshah Naghdy, and Alfred Mertins, "Sketch-based image matching using angular partitioning," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 35, no. 1, pp. 28–41, 2005.
- [2] David G. Lowe, "Object recognition from local scale-invariant features," in *Proc. of International Conf. on Computer Vision*, 1999, pp. 1150–1157.
- [3] Mathias Eitz, Kristian Hildebrand, Tamy Boubekeur, and Marc Alexa, "Sketch-based image retrieval: Benchmark and bag-of-features descriptors," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 17, no. 11, pp. 1624–1636, 2011.
- [4] Ming Cheng, Niloy J Mitra, Xumin Huang, Philip HS Torr, and Song Hu, "Global contrast based salient region detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 37, no. 3, pp. 569–582, 2015.
- [5] Christian Scharfenberger, Alexander Wong, Khalil Fergani, John S Zelek, David Clausi, et al., "Statistical textural distinctiveness for salient region detection in natural images," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 979–986.
- [6] Esa Rahtu, Juho Kannala, Mikko Salo, and Janne Heikkilä, "Segmenting salient objects from images and videos," in *European Conference on Computer Vision*, pp. 366–379. Springer, 2010.
- [7] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Ssstrunk, "Frequency-tuned Salient Region Detection," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, 2009, pp. 1597 – 1604.
- [8] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. of the 4th Alvey Vision Conf.*, 1988, pp. 147–151.
- [9] A. Kovács and T. Szirányi, "Improved Harris feature point set for orientation sensitive urban area detection in aerial images," *IEEE Geoscience and Remote Sensing Letters (in press)*, vol. 10, no. 4, pp. 796–800, 2013.
- [10] A. Manno-Kovacs, "Direction selective vector field convolution for contour detection," in *IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 4722–4726, "Top10%" award.
- [11] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, vol. 1, pp. 886–893.
- [12] Rui Hu, Mark Barnard, and John Collomosse, "Gradient field descriptor for sketch based retrieval and localization," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE, 2010, pp. 1025–1028.
- [13] Ming-Ming Cheng, NiloyJ. Mitra, Xiaolei Huang, and Shi-Min Hu, "Salientshape: group saliency in image collections," *The Visual Computer*, vol. 30, no. 4, pp. 443–453, 2014.
- [14] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
- [15] Andrea Kovacs and Tamas Sziranyi, "Harris function based active contour external force for image segmentation," *Pattern Recognition Letters*, vol. 33, no. 9, pp. 1180–1187, 2012.
- [16] Sanjiv Kumar and Martial Hebert, "Man-made structure detection in natural images using a causal multi-scale random field," in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, 2003, pp. 119–126.
- [17] Ofir Pele and Michael Werman, "The quadratic-chi histogram distance family," in *Proceedings of the 11th European Conference on Computer Vision: Part II*. 2010, pp. 749–762, Springer-Verlag.