

Tartalom alapú képi visszakeresés kiemelt irány hisztogram használatával *

Manno-Kovács Andrea

Elosztott Események Elemzése Kutatólaboratórium
MTA SZTAKI, Budapest
`andrea.manno-kovacs@sztaki.mta.hu`

Abstract. A tartalom alapú képi visszakeresés napjaink kiemelt kutatási területe, mivel a rendelkezésre álló digitális képi adat mennyisége ugrásszerűen nő. A létező vázlat („sketch”) alapú képi visszakereső (SBIR) rendszerek a valós felvételeken kevésbé hatékonyak, mivel az összetett háttér információ miatt a képi leírók eltorzulnak. A torzulás csökkentésére ez a cikk első lépésként egy előfeldolgozást javasol, ami képes automatikusan kiemelni a hangsúlyos („salient”) képrészleteket. Ezután a kiemelt részekben a Modified Harris for Edges and Corners (MHEC) kulcspont kereső által visszaadott ponthalmazt alapul véve, a kép leírására a kulcspontok körüli irányinformációt használjuk fel, felépítve a képet jellemző Salient Orientation Histogram-ot (SOH). A javasolt SBIR rendszer a visszakeresés mellett képes a visszaadott kép szemgentálására is, a képen levő salient objektum körvonalának megkeresésével. A módszer a random internetes képeket tartalmazó THUR15000 adatbázison került tesztelésre, a javasolt eljárást az egyéb szakirodalmi módszerekkel összehasonlítva a visszakeresési és az objektum detektálási eredmények is igen biztatóak.

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

* Eredeti publikáció: A. Manno-Kovács: ”Content based Image Retrieval using Salient Orientation Histogram”, in Proc. of IEEE International Conference on Image Processing, Phoenix, Arizona, USA, 2016. A munkát a European Space Agency (ESA) Hungary Industry Incentive Scheme támogatta az OWETIS projekt keretében.

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 [4–7]. 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 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

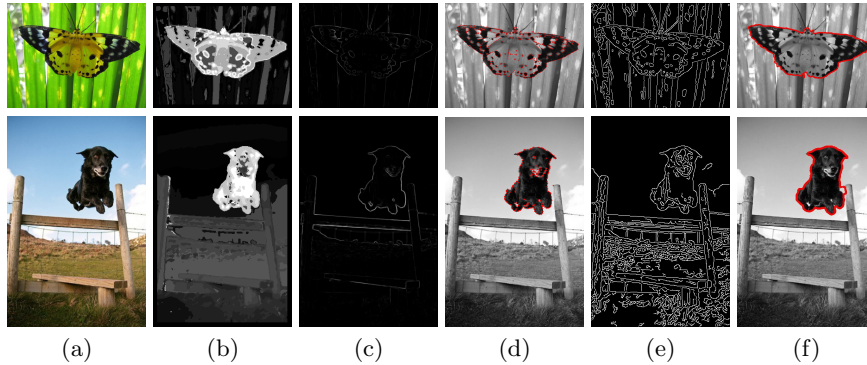


Fig. 1. Sample images for illustrating the proposed SBIR system: (a) the original image; (b) S texture distinctiveness map; (c) R_{mod} improved edge map; (d) MHEC_S point set; (e) Canny edge map and (f) the detection result achieved by DHVFC.

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_{ROI} is initialized. In Figure 1(b) the S_{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 λ_1 and λ_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_{mod} . The point set is able to represent points-of-interest inside the salient S_{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_{ROI} area. Features extracted for the MHEC points are able to describe the corresponding object more efficiently. The point subset inside the S_{ROI} area is marked by MHEC_S , samples are given in Fig. 1(d).

Besides the MHEC point set, the calculated R_{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_{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 $MHEC_S$ point set include the relevant feature points of the most distinctive image area. Therefore, the extracted salient orientation histogram (**SOH**) for the $MHEC_S$ 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 ∇g_i^R with $\|\nabla g_i^R\|$ magnitude and φ_i^∇ orientation for the i^{th} point, where R upper index refers to the gradient calculation on the R_{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 density of φ_i^∇ 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^{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 ϑ 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 χ^2 metric gives the best results. However, since then the Quadratic-Chi (QC) histogram distance family was published [17], introducing an improved technique

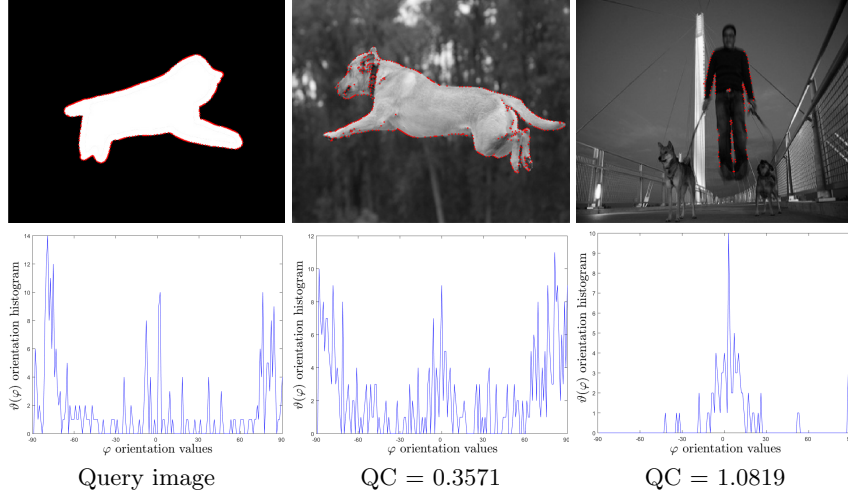


Fig. 2. Image comparison based on $\vartheta(\varphi)$ SOH function (Eq. 3) calculated for keypoints marked by red; and the QC distance metric (Eq. 4). Smaller QC value indicate more similar image to the query image.

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 χ^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 χ^2 . The retrieval tests of the author have also confirmed QC's superiority over χ^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_{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_S point set to achieve an automatic initialization step. As a next step, the active contour is initialized as the convex hull of the MHEC_S 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)).



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	41.4	43.2
SHoG [3]	36	40	82	78	74	73	18	18	90	91	60.0	60.0
Proposed	48	50	82	82	84	87	54	59	54	64	64.4	68.4

Table 1. True positive retrievals for THUR15000 keywords.

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 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 4.

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_β is calculated with $\beta^2 = 0.3$:

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

	FT	SEG	RC	GS	Prop.
Aver. F_β	0.51	0.53	0.56	0.68	0.72

Table 2. Average F_β values of segmentation methods.

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_β (Table 2) on the whole database. (Segmentation samples are presented in Figure 1(f)).

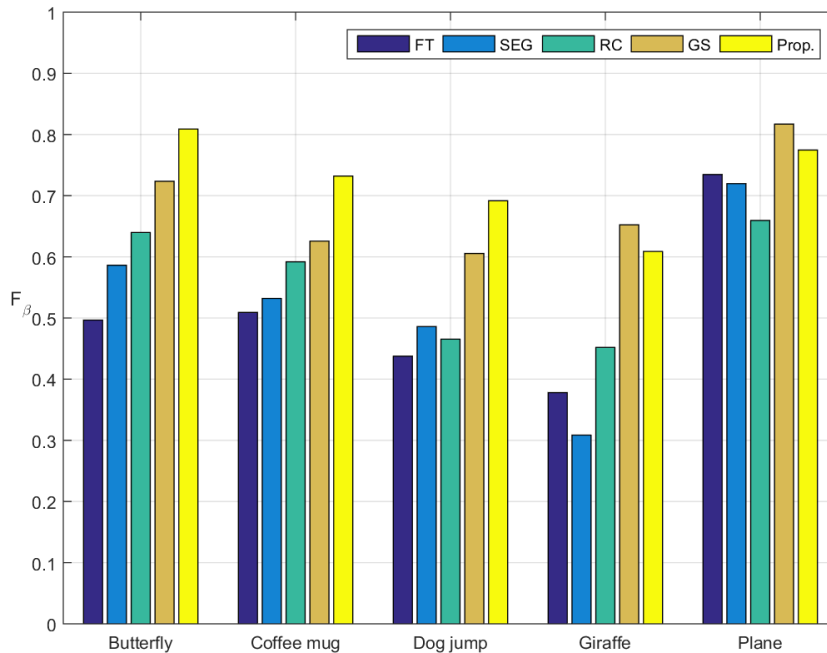


Fig. 4. F_β evaluation results for THUR15K database for different segmentation methods: FT [7], SEG [6], RC [4], GS [13] and the proposed method.

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. 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.

References

1. Chalechale, A., Naghdy, G., Mertins, A.: Sketch-based image matching using angular partitioning. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* **35** (2005) 28–41
2. Lowe, D.G.: Object recognition from local scale-invariant features. In: *Proc. of International Conf. on Computer Vision*. (1999) 1150–1157
3. Eitz, M., Hildebrand, K., Boubekeur, T., Alexa, M.: Sketch-based image retrieval: Benchmark and bag-of-features descriptors. *Visualization and Computer Graphics, IEEE Transactions on* **17** (2011) 1624–1636
4. Cheng, M., Mitra, N.J., Huang, X., Torr, P.H., Hu, S.: Global contrast based salient region detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **37** (2015) 569–582
5. Scharfenberger, C., Wong, A., Fergani, K., Zelek, J.S., Clausi, D., et al.: Statistical textural distinctiveness for salient region detection in natural images. In: *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. (2013) 979–986
6. Rahtu, E., Kannala, J., Salo, M., Heikkilä, J.: Segmenting salient objects from images and videos. In: *European Conference on Computer Vision*. Springer (2010) 366–379
7. Achanta, R., Hemami, S., Estrada, F., Süsstrunk, S.: Frequency-tuned Salient Region Detection. In: *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2009)*. (2009) 1597 – 1604
8. Harris, C., Stephens, M.: A combined corner and edge detector. In: *Proc. of the 4th Alvey Vision Conf.* (1988) 147–151
9. Kovács, A., Szirányi, T.: Improved Harris feature point set for orientation sensitive urban area detection in aerial images. *IEEE Geoscience and Remote Sensing Letters* (in press) **10** (2013) 796–800
10. Manno-Kovacs, A.: Direction selective vector field convolution for contour detection. In: *IEEE International Conference on Image Processing (ICIP)*. (2014) 4722–4726 ”Top10%” award.
11. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Volume 1. (2005) 886–893
12. Hu, R., Barnard, M., Collomosse, J.: Gradient field descriptor for sketch based retrieval and localization. In: *Image Processing (ICIP), 2010 17th IEEE International Conference on, IEEE* (2010) 1025–1028
13. Cheng, M.M., Mitra, N., Huang, X., Hu, S.M.: Salienshape: group saliency in image collections. *The Visual Computer* **30** (2014) 443–453
14. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics* **9** (1979) 62–66

15. Kovacs, A., Sziranyi, T.: Harris function based active contour external force for image segmentation. *Pattern Recognition Letters* **33** (2012) 1180–1187
16. Kumar, S., Hebert, M.: Man-made structure detection in natural images using a causal multiscale random field. In: *Proc. IEEE Conf. Comput. Vision Pattern Recogn.* (2003) 119–126
17. Pele, O., Werman, M.: The quadratic-chi histogram distance family. In: *Proceedings of the 11th European Conference on Computer Vision: Part II*, Springer-Verlag (2010) 749–762