ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Content Based Image Retrieval Using Salient Orientation Histograms

M.Jayasri
M.Tech Scholar
NRI Institute of Technology.

R.Sunitha
Associate Professor
NRI Institute of Technology.

Prof R.Suman
Professor
KL University.

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.

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, the latter concentrates only on a small image part. Recently published SBIR systems 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.

However, the dimension of a salient area description can still be very high, thus further reduction is needed. Interest point detectors, like Harris 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 for efficient image segmentation, and the method's strong ability for object detection was also shown previously, 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; moreover many improvements of the Histogram of Oriented Gradients (HoG) were published over the past years. The original HoG calculated the histogram for the whole image. Improved adaptations of HoG for SBIR systems are mostly using canny edge maps with orientation histograms calculated on pixels of the Canny edge map or randomized pixels. Following this technique, the background texture may

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

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.

2. SALIENT KEYPOINT DETECTION

2.1. Texture distinctiveness

Statistical texture distinctiveness model was introduced in 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 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. The *S* map is thresholded with Otsu's method to extract highly distinctive image parts. By selecting the region of maximal area, an automatic *SROI* is initialized. In Figure 1(b) the *SROI* areas can be well identified in both samples.

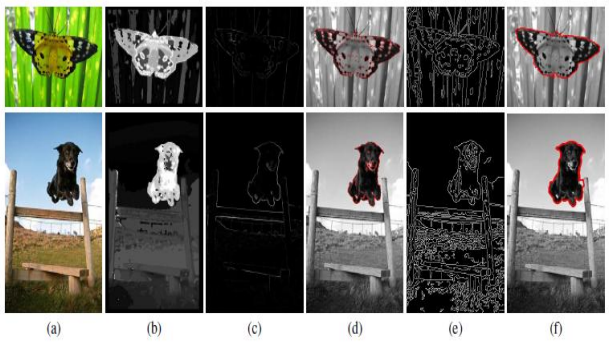

Figure1: Sample images for illustrating the proposed SBIR system: (a) the original image; (b) *S* texture distinctiveness map; (c) *R*mod improved edge map; (d) MHEC*S* point set; (e) canny edge map and (f) the detection result achieved by DHVFC.

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. The method adapts the Harris matrix and uses its $\lambda 1$ and $\lambda 2$ eigenvalues in the following modification of the *R* characteristic function.

$$R\text{mod} = \max(\lambda 1, \lambda 2). \quad (1)$$

The introduced modification has the ability to emphasize edge and corner regions in a balanced manner. MHEC interest points (*pi*) are selected as the local maxima of the *R*mod. The point set is able to represent points-of-interest inside the salient *SROI* 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 *SROI* area. Features extracted for the MHEC points are able to describe the corresponding object more efficiently. The point subset inside the *SROI* 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.

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. PROPOSED METHOD

Image Retrieval and Segmentation

Here for image retrieval we are using Scale Invariant Feature Transform.

3.1. Scale-invariant feature transform (SIFT):

SIFT is an algorithm in computer vision to detect and describe local features in images. For any object in an

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

image, interesting points on the object can be extracted to provide a "feature description" of the object. This description, extracted from a training image, can then be used to identify the object when attempting to locate the object in a test image containing many other objects. To perform reliable recognition, it is important that the features extracted from the training image be detectable even under changes in image scale, noise and illumination. Such points usually lie on high-contrast regions of the image, such as object edges.

Another important characteristic of these features is that the relative positions between them in the original scene shouldn't change from one image to another.

A. Scale-space extrema detection:

We begin by detecting points of interest, which are termed keypoints in the SIFT framework. The image is convolved with Gaussian filters at different scales, and then the difference of successive Gaussian-blurred images are taken. Key points are then taken as maxima/minima of the Difference of Gaussian(DoG) that occur at multiple scales.

Specifically, a DoG image is given by where is the convolution of the original image with the Gaussian blur at scale i.e. Hence a DoG image between scales and is just the difference of the Gaussian-blurred images at scales and For scale space extrema detection in the SIFT algorithm, the image is first convolved with Gaussian-blurs at different scales. The convolved images are grouped by octave (an octave corresponds to doubling the value of), and the value of is selected so that we obtain a fixed number of convolved images per octave. Then the Difference-of-Gaussian images are taken from adjacent Gaussian-blurred images per octave.

Once DoG images have been obtained, keypoints are identified as local minima/maxima of the DoG images across scales. This is done by comparing each pixel in the DoG images to its eight neighbors at the same scale and nine corresponding neighboring pixels in each of the neighboring scales. If the pixel value is the maximum or minimum among all compared pixels, it is selected as a candidate keypoint.

This keypoint detection step is a variation of one of the blob detection methods developed by Lindeberg by detecting scale-space extrema of the scale normalized Laplacian, that is detecting points that are local extrema with respect to both space and scale, in the discrete case by comparisons with the nearest 26 neighbours in a discretized scale-space volume. The difference of Gaussians operator can be seen as an approximation to the Laplacian, with the implicit normalization in the pyramid also constituting a discrete approximation of the scale-normalized Laplacian.

B.Keypoint localization:

Scale-space extrema detection produces too many keypoint candidates, some of which are unstable. The next step in the algorithm is to perform a detailed fit to the nearby data for accurate location, scale, and ratio of principal curvatures. This information allows points to be rejected that have low contrast (and are therefore sensitive to noise) or are poorly localized along an edge.

First, for each candidate keypoint, interpolation of nearby data is used to accurately determine its position. The initial approach was to just locate each keypoint at the location and scale of the candidate keypoint. The new approach calculates the interpolated location of the extremum, which substantially improves matching and stability. The interpolation is done using the quadratic Taylor expansion of the Difference-of-Gaussian scale-space function, $D(x, y, \sigma)$ with the candidate keypoint as the origin. This Taylor expansion is given by:

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x}$$

Where D and its derivatives are evaluated at the candidate keypoint and $\mathbf{x} = (x, y, \sigma)^T$ is the offset from this point. The location of the extremum is determined by taking the derivative of this function with

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

respect to x and setting it to zero. If the offset x^ is larger than 0.5 in any dimension, then that's an indication that the extremum lies closer to another candidate keypoint. In this case, the candidate keypoint is changed and the interpolation performed instead about that point. Otherwise the offset is added to its candidate keypoint to get the interpolated estimate for the location of the extremum.

C. Get rid of Low Contrast points:

Key points generated in the previous step produce a lot of key points. Some of them lie along an edge, or they don't have enough contrast. In both cases, they are not useful as features. So we get rid of them. The approach is similar to the one used in the Harris Corner Detector for removing edge features.

This is simple. If the magnitude of the intensity (i.e., without sign) at the current pixel in the DoG image (that is being checked for minima/maxima) is less than a certain value, it is rejected.

Because we have subpixel keypoints (we used the Taylor expansion to refine keypoints), we again need to use the taylor expansion to get the intensity value at subpixel locations. If it's magnitude is less than a certain value, we reject the keypoint.

D. Salient Oriented Histogram:

After step 3, we have legitimate key points. They've been tested to be stable. We already know the scale at which the keypoint was detected (it's the same as the scale of the blurred image). So we have scale invariance. The next thing is to assign an orientation to each keypoint. This orientation provides rotation invariance.

The idea is to collect gradient directions and magnitudes around each keypoint. Then we figure out the most prominent orientation(s) in that region. And we assign this orientation(s) to the keypoint.

Any later calculations are done relative to this orientation. This ensures rotation invariance.

The size of the "orientation collection region" around the keypoint depends on it's scale. The bigger the scale, the bigger the collection region.

First, the Gaussian-smoothed image $L(x, y, \sigma)$ at the keypoint's scale σ is taken so that all computations are performed in a scale-invariant manner. For an image sample L(x,y) at scale σ, the gradient magnitude, m(x,y), and orientation, $\theta(x, y)$, are precomputed using pixel differences:

Gradient magnitudes and orientations are calculated using these formulae:

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}$$

$$\theta(x,y) = \tan^{-1}((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y)))$$

The magnitude and orientation is calculated for all pixels around the keypoint. Then, A histogram is created for this.

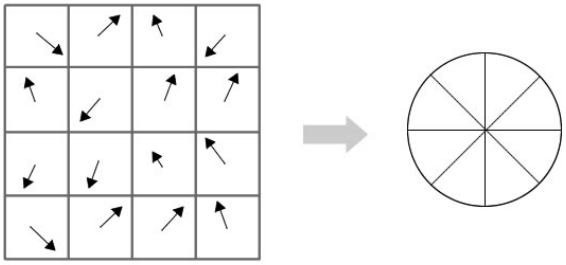

Figure2: Orientation assignment

In this histogram, the 360 degrees of orientation are broken into 36 bins (each 10 degrees). Lets say the gradient direction at a certain point (in the "orientation collection region") is 18.759 degrees, then it will go into the 10-19 degree bin. And the "amount" that is added to the bin is proportional to the magnitude of gradient at that point.

Once you've done this for all pixels around the keypoint, the histogram will have a peak at some point. Above, you see the histogram peaks at 20-29 degrees. So, the

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

keypoint is assigned orientation 3 (the third bin). Also, any peaks above 80% of the highest peak are converted into a new keypoint. This new keypoint has the same location and scale as the original. But it's orientation is equal to the other peak.

So, orientation can split up one keypoint into multiple keypoints.

E.Keypoint descriptor

Previous steps found keypoint locations at particular scales and assigned orientations to them. This ensured invariance to image location, scale and rotation. Now we want to compute a descriptor vector for each keypoint such that the descriptor is highly distinctive and partially invariant to the remaining variations such as illumination, 3D viewpoint, etc. This step is performed on the image closest in scale to the keypoint's scale.

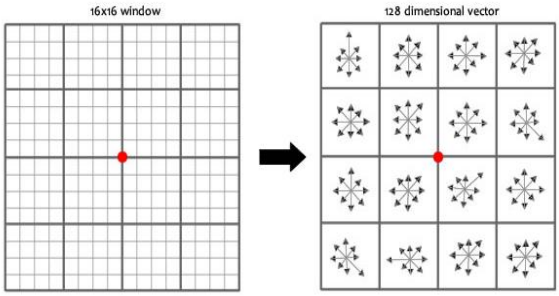

Figure3: locating the descriptor points.

First a set of orientation histograms is created on 4x4 pixel neighborhoods with 8 bins each. These histograms are computed from magnitude and orientation values of samples in a 16 x 16 region around the keypoint such that each histogram contains samples from a 4 x 4 subregion of the original neighborhood region. The magnitudes are further weighted by a Gaussian function with equal to one half the width of the descriptor window. The descriptor then becomes a vector of all the values of these histograms. Since there are 4 x 4 = 16 histograms each with 8 bins the vector has 128 elements. This vector is then normalized to unit length in order to enhance invariance to affine changes in illumination. To reduce the effects of non-linear illumination a threshold of 0.2 is applied and the vector is again normalized.

Although the dimension of the descriptor, i.e. 128, seems high, descriptors with lower dimension than this don't perform as well across the range of matching tasks and the computational cost remains low due to the approximate BBF method used for finding the nearest-neighbor. Longer descriptors continue to do better but not by much and there is an additional danger of increased sensitivity to distortion and occlusion. It is also shown that feature matching accuracy is above 50% for viewpoint changes of up to 50 degrees. Therefore, SIFT descriptors are invariant to minor affine changes. To test the distinctiveness of the SIFT descriptors, matching accuracy is also measured against varying number of keypoints in the testing database, and it is shown that matching accuracy decreases only very slightly for very large database sizes, thus indicating that SIFT features are highly distinctive.

4. Experimental Results:

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 database.

The best retrieval images for all the three methods are shown in Figure. 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 on the whole database.

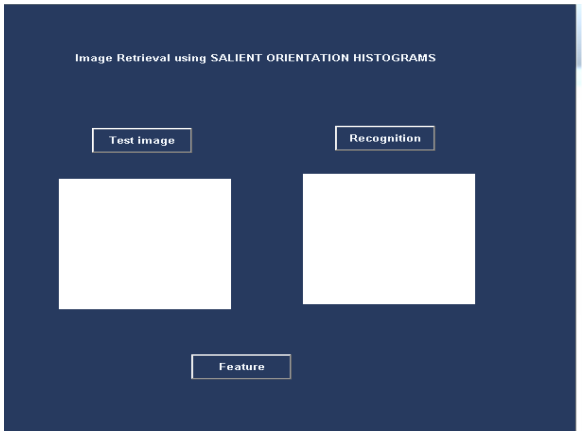

Figure4: GUI for content based image retrieval.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

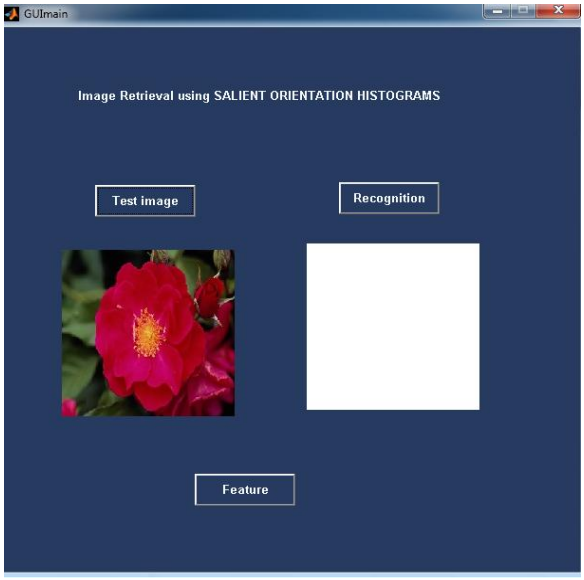

Figure5: Input image taken from database.

Figure6: Retrieved image.

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 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] Abdolah 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 Heikkil¨a, "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

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

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 an edge detector," in Proc. of the 4th Alvey Vision Conf., 1988, pp. 147–151.

[9] A. Kov´acs and T. Szir´anyi, "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 multiscale 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.