Query-by-Sketch Image Retrieval Using Edge Relation Histogram

Yoshiki KUMAGAI¹, Nonmember and Gosuke OHASHI¹ a, Member

SUMMARY There has recently been much research on content-based image retrieval (CBIR) that uses image features including color, shape, and texture. In CBIR, feature extraction is important because the retrieval result depends on the image feature. Query-by-sketch image retrieval is one of CBIR and query-by-sketch image retrieval is efficient because users simply have to draw a sketch to retrieve the desired images. In this type of retrieval, selecting the optimum feature extraction method is important because the retrieval result depends on the image feature. We have developed a query-by-sketch image retrieval method that uses an edge relation histogram (ERH) as a global and local feature intended for binary line images. This histogram is based on the patterns of distribution of other line pixels centered on each line pixel that have been obtained by global and local processing. ERH, which is a shift- and scale-invariant feature, focuses on the relation among the edge pixels. It is fairly simple to describe rotation- and symmetry-invariant features, and query-by-sketch image retrieval using ERH makes it possible to perform retrievals that are not affected by position, size, rotation, or mirroring. We applied the proposed method to 20,000 images in the Corel Photo Gallery. Experimental results showed that it was an effective means of retrieving images.

key words: image retrieval, feature extraction, edge image, sketch image

1. Introduction

Large image databases are becoming increasingly common with the growing popularization of digital consumer electronic products and high-speed networks. As the number of images in a database increases, it is difficult for users to quickly find the desired images. There is therefore an increasing demand for a method that can efficiently retrieve desired images [1]. The most popular image retrieval method in current use is text-based image retrieval (TBIR). TBIR is effective to retrieve desired images when users know keywords, but it is not effective to retrieve desired images when users do not know keywords or users do not query keywords. In addition, TBIR has issues: first, keyword annotation is laborious work, and second, users are sometimes not satisfied with retrieval results because keyword annotation is very subjective. In response, there has been much research on content-based image retrieval (CBIR), often with great success, because CBIR makes possible to retrieve desired images without keywords [2]–[31]. In CBIR, image features including color, shape, and texture are used along with queries in the form of example and sketch images. Query-by-example image retrieval makes possible to retrieve desired images effectively, but query-by-example image retrieval makes impossible to retrieve desired images without example images. In contrast, query-by-sketch image retrieval makes possible to retrieve without example images. For example, first query-by-sketch is used to obtain example images, and then query-by-example is to use for more effective retrieval. Query-by-sketch image retrieval is efficient because users simply have to draw a sketch to retrieve the target images. The edge-based features of shape and texture are very useful in this type of retrieval. Edge-based features including the edge histogram descriptor (EHD) in MPEG7 [3], angular radial partitioning (ARP) [17], the tensor descriptor [18], the Edgel Index [19] and the shape context [20] have been proposed. The EHD represents local edge distribution in an image, ARP extracts the angular- and radial-spatial distribution of edge pixels in an image by using the Fourier transform, the tensor descriptor uses structure tensors that encode the main gradient orientation in a certain image area, the Edgel Index describes a visual word using the position and orientation of the edge pixels, and the shape context represents the distribution over relative positions of other edge points at each edge point. QBIC [6], standing for Query By Image Content, is very famous system and was the first commercial CBIR system by IBM. QBIC allows the user to select and combine the features of color, texture and shape. The shape features of QBIC are simple global features and are based on the binary edge map (area, circularity, the major axis, and a set of associated algebraic moments up to degree 8). The authors have previously proposed ERH as an edge-based feature for query-by-sketch image retrieval. Query-by-sketch image retrieval using ERH enables shift-, scale-, rotation- and symmetry-invariant retrieval. In our previous work, ERH was focused on the global relations among edge pixels, with no consideration given to the local features [21]. The purpose of this study is to improve the retrieval accuracy of ERH by using local as well as global features.

2. Query-by-Sketch Image Retrieval

2.1 Overview

An overview of the proposed query-by-sketch image retrieval method is shown in Fig. 1. The proposed feature extraction is applied to binary line images and edge images when the images are sketch images and original images in the database, respectively. The image features of the original
images are stored in a feature database before the sketches are drawn by the user. The features of the sketch image and the database images are compared by calculating the similarity. The images are then ordered by decreasing similarity.

2.2 Edge Detection

In the proposed method, effective edge detection is required because the retrieval result depends on the edge image. Canny edge detection [33] is an effective edge detection method that includes three important parameters: standard deviation of Gaussian filter, high threshold, and low threshold. However, it is difficult to adjust the parameters for each image. Using multiple resolution images enables the usage of optimum parameters in Canny edge detection. Therefore, applying Canny edge detection using fixed parameters to multiple resolution images is equally effective in applying Canny edge detection using various parameters to the image, as shown in Fig. 2 [22]. First multiple resolution images are arranged as shown in Fig. 2 (a). Here an original image is referred to as the image at Level 0, the image, which is a half resolution of the original image, is referred to as the image at level 1 and the image, which is a half resolution of the image at level 1, is referred to as the image at level 2. Canny edge detection is applied to the multiple resolution images.

As explained above, the edge images for multiple resolution images are detected as shown in Fig. 2 (b). The edge image that is the most similar to a sketch image (among the edge images obtained by applying Canny edge detection to the multiple resolution images) is used.

2.3 Feature Extraction

The features are extracted by applying ERH as global and local feature, as shown in Fig. 3. ERH is extracted by defining each edge pixel as the pixel of interest. On the pixel of interest, the number of edge pixels in each of eight regions (determined as shown in Fig. 3 (a)) is counted. These regions include the entire image (Fig. 3 (e)) and the \(3 \times 3\) region of interest (Fig. 3 (i)) in the global process and the local process, respectively. The number of edge pixels in each region is denoted as \(c_x\), as shown in Fig. 3 (b). Figures 3 (f) and (j) show examples of \(c_x\), obtained by the global and local processes, respectively. \(s_x\) is the normalized value obtained by Eq. (1), as shown in Fig. 3 (c). Figures 3 (g) and (k) show examples of \(s_x\), obtained by the global and local processes, respectively. The value of the binary pattern \(b_x\) is calculated by the thresholding technique using Eq. (2), as shown...
The values of $b_x$ in Fig. 3 (d). Figures 3 (h) and (l) show examples of $b_x$ obtained by the global and local processes, respectively. The values of $C$ and $Th$ denote the total number of edge pixels in each of the eight regions and the threshold value, respectively. The threshold value is 0.15 because the value is near 1/8 and a little larger than 1/8 for reducing noise.

$$s_x = \begin{cases} 
C - 1 & (C \neq 0) \\
0 & (C = 0) 
\end{cases} \quad (1)$$

$$b_x = \begin{cases} 
0 & (s_x \leq Th) \\
1 & (s_x > Th) 
\end{cases} \quad (2)$$

The values of $b_x$ (0 or 1) in each direction are arranged counterclockwise from $b_0$ to express an eight-bit binary number. The binary number $(b_7, b_6, b_5, b_4, b_3, b_2, b_1, b_0)$ is transformed into a decimal number $d$ ($0 \leq d \leq 255$). We then vote for the corresponding bin $(d_g, d_l)$ in the histogram shown in Fig. 3 (m). $d_g$ and $d_l$ are the decimal numbers obtained by the global and local processes, respectively. A 256 × 256-dimensional (65,536-dimensional) image feature is obtained by applying the above process to all edge pixels and normalizing the histogram by the total number of edge pixels. This process makes shift-invariance and scale-invariance possible. The rotation- and symmetry-invariant features are easily described by shifting the binary number.

2.4 Similarity Measure

The similarity is calculated by using the histogram intersection between the feature of a query image and the features in the feature database. The images are ordered by decreasing similarity. The histogram intersection $S_i$ is given by Eq. (3).

$$S_i = \max_{j=0,1,2} \left[ \sum_{g=0}^{255} \sum_{l=0}^{255} \min(F_{g,lmn}, F_{ig,jl}) \right]$$

(3)

where $F_s$ is the feature of a query sketch, $F_i$ is the feature of a database image, $j$ is the level of multiple resolution, $g$ is a 256-dimensional global feature, $l$ is a 256-dimensional local feature, $m$ is the feature pattern for rotation-invariant retrieval, and $n$ is the feature pattern for symmetry-invariant retrieval.

3. Experimental Results

We performed retrieval experiments on 20,000 images in the Corel Photo Gallery to determine the effect of the proposed method. The retrieval targets were “hammer”, “duck”, “car”, “wineglass”, “balloon”, “card”, “fish”, “dog”, “bonsai”, and “flag”. The relevant images included 9 for “hammer”, 100 for “duck”, 748 for “car”, 94 for “wineglass”, 109 for “balloon”, 89 for “card”, 98 for “fish”, 100 for “dog”, 78 for “bonsai” and 100 for “flag”. An example of the shift-, scale-, rotation-, and symmetry-invariant retrieval result for “hammer” is shown in Fig. 4. The images ranked 1st, 2nd, 6th, 7th, and 8th demonstrate the effect of the shift- and scale-invariant feature, the image ranked 6th demonstrates the effect of the symmetry-invariant feature, and the images ranked 1st, 2nd, 7th, and 8th demonstrate the effect of the rotation-invariant feature. This demonstrates that the proposed method is able to achieve shift-, scale-, rotation-, and symmetry-invariance. Next, we compared the proposed method with ERH (global), ERH (local), the shape context, ARP, ARP without Fourier transform, the tensor descriptor, and the Edgel Index. In the shape context, the number of sampled edge points was set to 100 and the number of log-polar histogram bins for radial distance and angle were 5 and 12, respectively. For ARP, the edge and sketch images were normalized so that the long side was 384 pixels and the number of angular and radial partitions were 8 and 3, respectively. For the tensor descriptor, the number of image cells was set to 24 × 24 to a circumscribed square of an edge image. For the Edgel Index, the edge and sketch image were normalized so that the long side was 200 pixels; the tolerance radius used in a similarity measurement was set to 3. Figures 5–12 show examples of the shift-, scale-, and symmetry-invariant retrieval results for the retrieval targets of “duck”, “car”, “wineglass”, “balloon”, “card”, “fish”, “dog”, and “bonsai”, respectively. Figures 13 and 14 show examples of the shift- and scale-invariant retrieval results by different sketches for the same retrieval target of “flag”. Figures 5–14 (a) show the retrieval results of the proposed method, Figs. 5–14 (b) show the retrieval results of ERH (global), Figs. 5–14 (c) show the retrieval results of ERH (local), Figs. 5–14 (d) show the retrieval results of the shape context, Figs. 5–14 (e) show the retrieval results of ARP, Figs. 5–14 (f) show the retrieval results of ARP without Fourier transform, Figs. 5–14 (g) show the retrieval results of the tensor descriptor, and Figs. 5–14 (h) show the retrieval results of the Edgel Index. The results obtained using the proposed method were more accurate than those obtained using the other methods.

We quantitatively evaluated the results using a precision-recall graph, where precision is the proportion of retrieved relevant images and recall is the proportion of retrieved relevant images to relevant images in the database. Precision $p(T)$ and recall $r(T)$ are given by Eqs. (4).

$$p(T) = \frac{n(T)}{T}, \quad r(T) = \frac{n(T)}{N}$$

(4)
Fig. 5 Example of retrieval results for “duck”.

Fig. 6 Example of retrieval results for “car”.

Fig. 7 Example of retrieval results for “wineglass”.

(a) ERH (global and local)  (b) ERH (global)  (c) ERH (local)  (d) Shape Context

(e) ARP  (f) ARP without Fourier transform  (g) Tensor Descriptor  (h) Edgel Index
Fig. 8  Example of retrieval results for “balloon”.

Fig. 9  Example of retrieval results for “card”.

Fig. 10  Example of retrieval results for “fish”.
Fig. 11 Example of retrieval results for “dog”.

Fig. 12 Example of retrieval results for “bonsai”.

Fig. 13 Example of retrieval results for “flag” by query 1.
Fig. 14 Example of retrieval results for “flag” by query 2.

Fig. 15 Precision-Recall graph. “A” is ERH (global and local). “B” is ERH (global). “C” is ERH (local). “D” is shape context. “E” is ARP. “F” is ARP without Fourier transform. “G” is tensor descriptor. “H” is Edgel Index.
where $T$ is the shortlist of the most similar images, $n(T)$ is the number of retrieved relevant images, and $N$ is the total number of relevant images in the database. We constructed a precision-recall graph with recall on the horizontal axis and precision on the vertical axis. Figures 15 (a)–(j) show the precision-recall graph corresponding to Figs. 5–14, respectively. From Figs. 5–14, we see that the retrieval results using the proposed global and local features of ERH (Figs. 5–14 (a)) is superior to the retrieval results using the only global feature of ERH (Figs. 5–14 (b)). Although the only local feature of ERH is not effective (Figs. 5–14 (c)), the global feature of ERH with local feature taken into account for local edge connection is effective because the global feature of ERH is important. From Figs. 5–14, we see that the retrieval results using the proposed global and local features of ERH (Figs. 5–14 (a)) is the best among the retrieval results (Figs. 5–14 (a)–(h)). Except for (a), the retrieval results using the shape context (Figs. 5–14 (d)) is better than the other results, especially for “duck”, “wineglass”, and “card”. The shape context is sometimes too strict to retrieve by sketch, because the shape context represents the distribution over relative positions of other edge points at each edge point. The proposed method is simpler and more robust than the shape context and the proposed method is suitable for query-by-sketch image retrieval.

Results show that the proposed method can provide accurate query-by-sketch image retrieval. The proposed edge feature is suitable for object images easy to extract edges as shown in Figs. 4–14, but is not suitable for natural scene images hard to extract edges, because it is necessary for edge features to extract edges reliably. The proposed method is effective for query consisting of one object in order to retrieve an image consisting of one object as shown in Figs. 4–14, however the proposed method is not effective for query consisting of one object in order to retrieve an image consisting of plural objects as shown in Fig. 16 (a), because one feature is extracted from one image. In order to retrieve an image consisting of plural objects as shown in Fig. 16 (b), it is necessary to input a query consisting of two objects as shown in Fig. 16 (b) or it is necessary to extract the feature from each object segmented from images.

The average retrieval time was 1451 ms for shift-, scale-, rotation-, and symmetry-invariant retrieval (Fig. 4), 414 ms for shift-, scale-, and symmetry-invariant retrieval (Figs. 5 and 6), and 301 ms for shift- and scale-invariant retrieval (Figs. 13 and 14) on an Intel Core i7-2860QM (2.50 GHz, 16.0 GB RAM). This means it is possible to retrieve images from 20,000 images in real-time. However, if the proposed method retrieves from more than 100,000,000 images, the proposed method cannot retrieve in real-time. It would be possible to reduce retrieving time by parallel processing of similarity measure because similarity measure, which is the majority of retrieving time, is calculated for each image. In addition, It would be possible to reduce retrieving time using hierarchical structure [32].

4. Conclusion

We proposed a query-by-sketch image retrieval method using ERH as a global and local feature to improve retrieval accuracy. This method was applied to 20,000 images in the Corel Photo Gallery and the results were quantitatively evaluated using a precision-recall graph. Results showed that the retrieval accuracy was improved by using ERH as a global and local feature. This demonstrates that the proposed method is an effective means of image retrieval.

This work was supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan under grant number 21700155. The authors wish to express their gratitude to Mr. Yusuke Amano and Mr. Yuki Kuno of Shizuoka University for numerous enlightening conversations and much invaluable advice.

References


Yoshiki Kumagai received his B.E. and M.E. degrees from Shizuoka University in Hamamatsu, Japan in 2010 and 2012, respectively. His research interests include image retrieval.

Gosuke Ohashi received his B.E., M.E. and D.E. degrees from Keio University in Yokohama, Japan in 1992, 1994 and 1997, respectively. He had been an Assistant Professor since 1997 and he is currently an Associate Professor in Electrical and Electronic Engineering at Shizuoka University. He was a visiting researcher at University of California, Santa Barbara from 2003 to 2004. His research interests include image processing and image retrieval.