Robust Similar Image Retrieval Based on Extracted Object Features

Hiroto Nishiki and Shigeo Wada

Graduate School of Engineering, Tokyo Denki University
5 Senju-Asahi-cho, Adachi-ku, Tokyo 120-8551, Japan
E-mail: 16kmh11@ms.dendai.ac.jp, wada@cck.dendai.ac.jp

Abstract

In this paper, a robust similar image retrieval method using extracted object features is proposed. A local texture feature as well as global contour and color features are extracted from an object image to generate unified robust feature vectors. To show the effectiveness of our method, experimental noisy image retrieval was executed using standard object image databases. The recall rate, precision rate and F-measure obtained by cross-validation were calculated to evaluate the performance of object image retrieval. High-performance image retrieval was achieved compared with the conventional methods without using combined robust features of extracted objects.

1. Introduction

Recently, owing to the development of smart devices with high-performance cameras, we have captured images in daily life. Numerous image data can instantly be stored and exchanged by a social network system. To handle such massive data, automatic image retrieval and classification techniques are indispensable.

Content-based image retrieval (CBIR) using a query image is a useful methodology to retrieve a huge number of images quickly because the indexing of images is not required [1]. Global image features are extracted from various formats of images by image processing. The features are combined for use in CBIR [2], [3]. CBIR systems for scene images, biomedical images, product images, and so on, have been investigated for applications in fields such as security systems (biometrics), medical management systems, online e-commerce and computer vision (see [1]-[10]).

On the other hand, when a user would like to find an interested region of similar images, a region-based image retrieval (RBIR) system that searches for images containing a specific region, such as medical cellular images, can be used [8]. In this case, a local texture feature is extracted from an ambiguous cell region [8]. Furthermore, when a user searches for target objects in similar images, an object-based image retrieval (OBIR) system is applied. Although objects equivalent to the query input with various sizes, positions and angles can be retrieved, the background image is restricted to monotone scenes [9]. Furthermore, conceptually equivalent objects such as traffic signs and cars in images can be searched for by OBIR. In this case, because the retrieval is based on the category of the object, physically consistent similar objects are occasionally not retrieved. In addition, when noise and disturbance are superposed on objects, the retrieval accuracy becomes low. In such a noisy environment, a Bayesian retrieval system can be used [4]. However, sample images are required to increase the efficiency of learning processes.

In this paper, we propose a robust similar object image retrieval method using extracted object features. Our method retrieves similar objects such faces from a complicated background containing noise from an image database. To improve the retrieval accuracy, we extract several robust features from local and global regions of an object and combine them.

To characterize an image and object, the surface texture, color, contour and shape features are important elements. A local binary pattern (LBP) and its variants are used to describe the surface texture. A LBP and local mesh pattern (LMeP) are noise-robust texture features and are robust to illumination and geometrical variation. Furthermore, a global contour feature described by a directional edge histogram based on the Canny edge detector and a color feature described by a Hue histogram are used. These are scale-, shift- and rotation-invariant features. In our method, to extract several robust features, an object is extracted by \( k \)-means-based clustering, which does not require a learning process with a supervisor. Then, individual features of an object are combined to obtain a unified robust feature vector using weight functions.

2. Similar Image Retrieval Using Object Features

2.1 System configuration

Figure 1 shows the configuration of our object image
retrieval system. When a query image is input by a user, similar object images can be retrieved from an image database. To retrieve similar images precisely based on users’ requirements, features of the query as above image and those of the image database are extracted. Both local and global histogram features are extracted from the object. The local feature is calculated from a texture feature (texture of object), whereas the global feature is calculated from an edge feature (contour of object) and color feature (color of object). They are combined using weight functions.

The similarity between images is measured by a statistically normalized norm using the average and variance. Retrieved images are sorted in order. Finally, the user obtains the desired object images.

3. Methods

In this section, we explain our feature extraction and unification processes.

3.1 Feature extraction

Figure 3 shows the feature extraction and unification processes. An image represented by RGB is transformed to a grayscale image (Y component) that represents the brightness of the image. The image is divided into a number of small blocks, and local texture features such as the LBP and LMeP are calculated. Each texture feature represented by a binary number is converted to a decimal number and its histogram is obtained. Figures 4(a) and 4(b) show examples of texture feature images based on the LBP and LMeP features, respectively.

Furthermore, a global feature is extracted using a grayscale image and Hue image (H component). The H component represents the color feature and its histogram is calculated. The contour of an object is represented as an edge image that is calculated by a Canny edge detector. Then, the distribution of the edge direction is obtained as an edge histogram. Figures 4(c) and 4(d) show examples of contour feature images obtained by the edge detector and the color based on H component, respectively.

Finally, the local and global histogram features are combined to obtain a unified feature vector using weight functions to represent the object image.
3.2 User oriented feature unification

When a user would like to retrieve similar images, there are various possibilities. If a user considers that both the object and the background are important, features associated with not only the extracted object but also the separated background must be used for retrieval. In such a situation, the unification of weighted feature vectors of the object and background is required. Furthermore, when a user considers the contour of an object to be more important than the color, a feature-weighting function is used. When the surface texture of the object is important, the weight of the texture feature will be emphasized.

4. Simulations

First, the effectiveness of combining feature vectors is shown. Next, the effectiveness of object extraction from a complicated background and noisy environment is shown.

4.1 Simulation conditions

Three object image databases were prepared. Table 1 shows their number of images, image size, format and the number of individual objects. ‘Coil-100’ (http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php) is an object image database that contains 100 color objects such as toys and cups with various viewpoints and sizes (D1, Fig.5(a)). ‘Essex Face’ (http://cswww.essex.ac.uk/mv/allfaces/index.html) is a face object image database that contains thirty persons’ faces with various expressions (D2, Fig.5(b)). These images have a uniform background. ‘Caltech’ (http://www.vision.caltech.edu/html-files/archive.html) is a face object image database that contains fifteen persons’ faces with complicated backgrounds (D3, Fig.5(c)). To evaluate the tolerance to disturbance, several types of noise and distortion were added to query images. Examples of noisy images and PSNR [dB] values are shown in Fig. 6.

Table 1: Image databases

|        | D1     | D2     | D3     |
|--------|--------|--------|--------|
| Num. of images | 1200   | 300    | 150    |
| Image size | 128x128| 180x200| 300x198|
| Image format | png, RGB | jpg, RGB | jpg, RGB |
| Num. of objects | 12    | 10     | 10     |

Figure 5: Examples of database images

(a) I: Gaussian 23.84 [dB]  (b) II: Poisson 31.41 [dB]  (c) III: Salt & Pepper 21.64 [dB]  
(d) IV: Speckle 24.32 [dB]  (e) V: Illumination (+) 23.54 [dB]  (f) VI: Illumination (-) 24.10 [dB]

Figure 6: Examples of noisy images

4.2 Performance analysis

Cross-validation was executed. The recall rate, precision rate and F-measure (harmonic mean of recall and precision rates) were used to evaluate the performance of object image retrieval. The average rate and standard deviation were also calculated. Noted that when the number of retrieved images is equal to that of object images, the precision and recall rates can be unified by the F-measure.

Tables 2(a) and 2(b) show the average F-measures and standard deviations of object-extracted images without distortion or noise using Euclidean and statistically normalized similarity norms, respectively. It was found that 588, 210 and 26 objects from databases D1 - D3 were almost correctly extracted by k-means clustering, respectively. These images were used in Table 2. In the tables, ‘Comb.’ represents our combined feature vector with uniform weights. Our unified feature vectors using the statistically normalized norm showed high values compared with single-feature methods.

Table 3 shows the average F-measures and standard deviations of original images (without object extraction) using the statistically normalized norm. When the background is uniform, such as in D1 and D2, the F-measure becomes high. However, when the retrieved image contains a complicated background, such as in D3 or for various viewpoints of scaled-object images in D1, the accuracy becomes low. In these cases, the global feature is not effective. Therefore, object extraction is effectively used to reduce the noise and disturbance of object images. If an object is precisely extracted, the F-measure is improved.
Table 2: Average F-measures and standard deviations
(extracted object features when noise-free)

|       | LBP    | LMeP   | Edge  | Color | Comb.  |
|-------|--------|--------|-------|-------|--------|
|       | D1     |        |       |       |        |
|       | 18.8±24.7 | 27.0±25.6 | 28.0±21.7 | 64.7±25.8 | 66.1±26.4 |
|       | D2     |        |       |       |        |
|       | 23.6±36.3 | 14.2±27.7 | 96.6±6.53 | 96.0±4.71 | 99.9±0.47 |
|       | D3     |        |       |       |        |
|       | 15.9±28.4 | 22.5±28.6 | 60.1±33.9 | 72.6±34.3 | 65.6±33.9 |
|       |        |        |       |       |        |
| (a)   | Euclidean norm |        |       |       |        |
|       | D1     |        |       |       |        |
|       | 32.6±26.8 | 37.5±27.7 | 27.9±21.7 | 64.8±25.7 | 69.3±25.2 |
|       | D2     |        |       |       |        |
|       | 88.8±18.3 | 89.0±18.5 | 85.2±20.3 | 80.6±19.4 | 91.3±18.2 |
|       | D3     |        |       |       |        |
|       | 60.7±31.2 | 64.7±30.9 | 60.7±33.7 | 70.7±34.2 | 73.5±32.5 |
|       |        |        |       |       |        |
| (b)   | Statistically normalized norm |        |       |       |        |

Table 3: Average F-measures and standard deviations
(combined features when noise free)

|       | LBP    | LMeP   | Edge  | Color | Comb.  |
|-------|--------|--------|-------|-------|--------|
|       | D1     |        |       |       |        |
|       | 25.8±23.1 | 45.3±28.2 | 31.6±27.9 | 76.8±21.3 | 75.8±23.1 |
|       | D2     |        |       |       |        |
|       | 85.1±18.6 | 90.9±15.6 | 97.3±9.48 | 93.0±10.7 | 99.3±1.98 |
|       | D3     |        |       |       |        |
|       | 17.5±6.40 | 17.9±5.24 | 18.2±4.13 | 25.9±12.6 | 27.7±12.7 |

4.3 Discussion

Next, the robustness to various types of noise and disturbance was examined. The relationship between the PSNR and the average F-measure of noisy and distorted object-extracted images of D3 is shown in Fig. 7. Figures 7(a)-7(e) represent the noise types I, III, IV, V and VI, respectively. Our method showed robustness to noise and disturbance compared with other single-feature methods.

Figure 7(f) shows the average F-measure for our method for all types of noise and disturbance. The decrease in the F-measure for our method in the case of varying illumination was low.

![Figure 7](image)

Figure 7: Relationship between PSNR and F-measure of noisy and distorted object-extracted images:
(a) I, (b) III, (c) IV, (d) V, (e) VI, (f) Only Comb.

5. Conclusions

In this paper, a noise-robust object image retrieval method was proposed. Robust local texture, global contour and color histogram features were extracted from an object in an image. In the simulations, the effectiveness of our method was examined using noisy and complicated background images. High performance was achieved compared with methods without using object extraction. The accuracy of retrieval in a noisy environment depended on the performance of object extraction. To increase accuracy rates, improving correctness of object extraction and estimation of the optimum weight function are left as future works.

References

[1] T. Deselaers, D. Keysers and H. Ney: Features for image retrieval: An experimental comparison, Information Retrieval, Vol. 11, pp. 77-107, 2008.
[2] Y. D. Chun, N. C. Kim and I. H. Jang: Content-based image retrieval using multi-resolution color and texture features, IEEE Trans. Multimedia, Vol. 10, No. 6, pp. 1073-1084, 2008.
[3] J. Yu, Z. Qin, T. Wan and X. Zhang: Feature integration analysis of bag-of-features model for image retrieval, Neurocomputing, Vol. 120, No. 23, pp. 355-364, 2013.
[4] D. Hoiem, R. Sukthankar, H. Schneiderman and L. Huston: Object-based image retrieval using the statistical structure of images, Proc. IEEE CVPR, 2004.
[5] G. H. Liu, Z. Y. Li, L. Zhang and Y. Xu: Image retrieval based on macro-structure descriptor, Pattern Recognition., Vol. 44, No. 9, pp. 2123-2133, 2011.
[6] A. Surulianti, K. Meena and R. Reena: Local binary pattern and its derivatives for face recognition, IET Computer Vision, Vol. 6, No. 5, pp. 480-488, 2012.
[7] S. Murala and Q. M. J. Wu: Local mesh patterns versus local binary patterns, IEEE J. of Biomed. and Health Inf., Vol. 18, No. 3, pp. 929-938, 2014.
[8] G. W. Ji, P. S. J. D. Raj: Content-based image retrieval in dermatology using intelligent technique, IET Image Process., Vol. 9, No. 4, pp. 306-317, 2015.
[9] A. M. Patel and R. Thakkar: Object based image retrieval from database using local and global features, International Journal of Innovative Research in Technology, Vol. 1, No. 12, pp. 1309-1314, 2015.
[10] S. R. Dubey, S. K. Singh and R. K. Singh: Multichannel decoded local binary patterns for content-based image retrieval, IEEE Trans. Image Process., Vol. 25, No. 9, pp. 4018-4032, 2016.