

Product Images Retrieval Based on Smart Mobile Devices

Yuzhong Zhang and Chongwen Wang

Abstract—In respect of the particularity of the product images, this paper studies the relevant technologies of content-based image retrieval, and proposes an improved method including image segmentation, perceptual hash, feature point matching and normalization for the product image retrieval. Using recall and precision as the performance evaluation criteria of the product image retrieval system, this paper also studies the influence of each component in the improved method on the performance of product image retrieval through experiments.

Index Terms—CBIR, image segmentation, perception hash, feature point matching, normalization.

I. INTRODUCTION

According to the characteristics of the input image, the image retrieval is known as a kind of retrieval based on the similarity of images in the database. In the early stages of image retrieval, the text-based image retrieval (TBIR) technology was more widely used. Due to using text to describe the characteristics of the image, it was extremely vulnerable to the interference of vocabulary. With the shortcomings of slow update, it determined that the method couldn't adapt to the changing requirements of various types of image retrieval. In this urgent situation, researchers propose the content-based image retrieval (CBIR) technology, which is a kind of retrieval based on the semantic analysis of image content, so as to avoid the retrieval failure caused by the ambiguous of text description. It belongs to the content-based retrieval (CBR), and puts emphasis on the content features and visual similarity of images to avoid the use of textual description [1].

CBIR extracts image characteristics such as color, texture, shape *et al.* and these characteristics are stored together with the image, in order to achieve the purpose of image retrieval, which is different from the traditional image retrieval [2]. CBIR extracts information directly from the image content, avoiding the problem of text ambiguity, also frequently uses the method of human-computer interaction to extract image features and build index in order to study human perception of similarity between the images. The basic components of the image retrieval system are shown in Fig. 1. At first the image feature database can be generated by the image database through the image feature extraction. When the user provides the image to be matched, the system will extract the image features, and compare them with the features in the image feature database for image feature matching, and then similar

images from high to low sequence will be sent back to the user. The user can send feedback according to the query results, and image retrieval system will return a more accurate result of similar images based on the feedback [3].

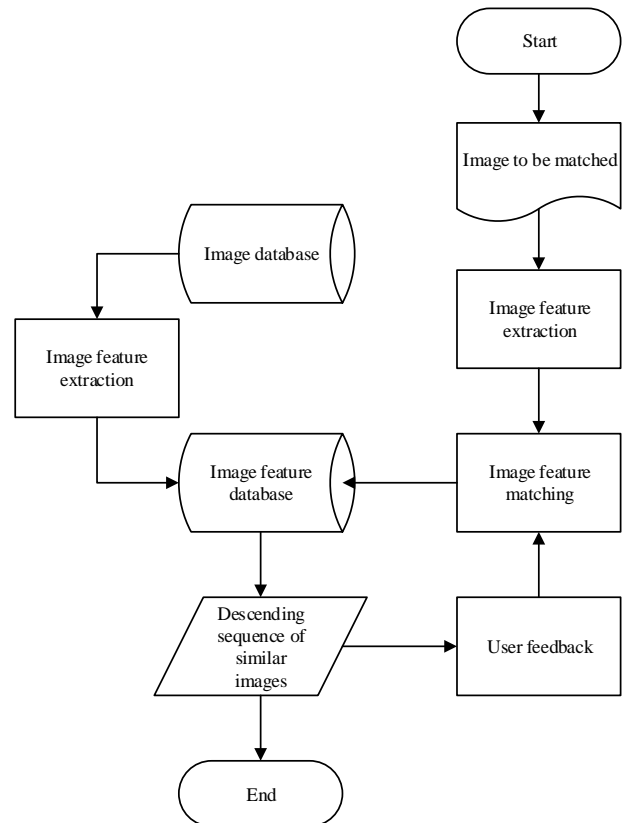


Fig. 1. The basic components of an image retrieval system.

The key point of the content-based image retrieval is the visual features extraction of an image, and to do image retrieval based on these features. Visual features of an image can be divided into low-level visual features and advanced semantic features. Low-level visual features mainly refer to the color, shape, texture and other features of the image. The advanced semantic features represent the complex semantic meaning, such as the specific objects included in the image, the occurrence of scenes and the emotional color described in the image [2]. Image feature extraction and image feature matching are two most critical parts of the content-based image retrieval. In current there are many ways to achieve the image retrieval, such as fractal-based image retrieval, color-based image retrieval and integrated multi-methods image retrieval *et al.*

II. ALGORITHM OVERVIEW

In respect of the particularity of the product images, this paper proposes an improved method to improve the efficiency

of image retrieval. The framework is shown in Fig. 2.

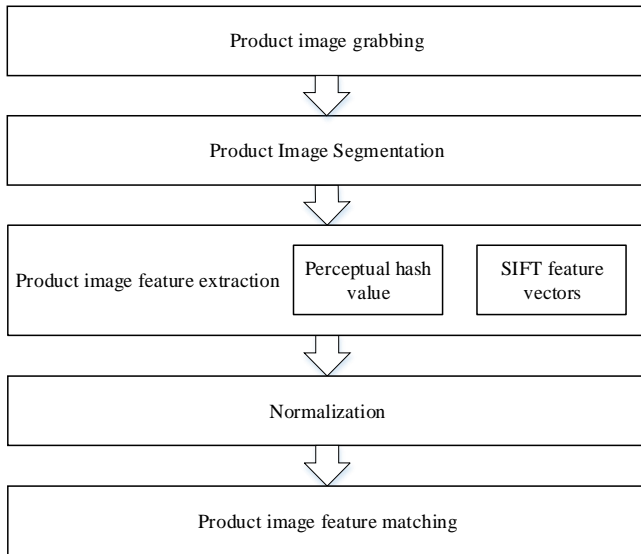


Fig. 2. The framework of improved method.

Firstly, the edge detection method is used for the product image segmentation to extract the specific goods included in the image. Followed by the product image feature extraction, including the perceptual hash value and SIFT feature vectors. When the perceptual hashing algorithm is applied to the product image, it can generate a unique "fingerprint" string for each picture. Perceptual hashing algorithm has advantages of summary, robustness, one-way *et al.* The similarity of product images can be obtained by comparing their fingerprints, the closer the fingerprints are, the more similar the product images are. The SIFT algorithm is used to extract the SIFT feature points which can be used to describe the local features of the product image, and the SIFT feature vector can maintain a certain degree of stability to routine operations, such as rotating zoom, perspective changes, affine transformation, brightness variation *et al.* Then the normalization is used for the purpose of SIFT feature vector space dimension reduction and assigning the appropriate weights for different image features. Finally, conduct image feature matching upon the perceptual hash value and SIFT feature vector.

III. PRODUCT IMAGE SEGMENTATION

During the process of the product image retrieval, users are usually more concerned about the specific goods included in the image. If the original image can be divided into two parts of the target image and the background image, and then get the target area for feature extraction, description and similarity measure, it can exclude the interference of the background image content features on the search results and greatly improves the accuracy of image retrieval [4]. Therefore, using the image segmentation methods to extract the main product has a positive meaning for the product image retrieval system.

According to the characteristics of the product image, generally it has clear target edge and simple background, so the edge detection method can be used to get the target area [5]. The first is contour detection on the picture. Based on the

requirements, the product itself should be the core part of the matching image in need, so assume that the product occupies the most regions of the whole picture. Then find the maximum closed contour of the product image, set the color of the internal area in the maximum closed contour to white, external area to black. Create a new image that is the same with the original one, then apply the "with" operation on them, which can get the internal part in the maximum closed contour of the picture. According to the previous assumption, this part is the specific goods, as shown in Fig. 3.



Fig. 3. Product image segmentation based on the edge detection.

IV. PRODUCT IMAGE FEATURE EXTRACTION AND MATCHING

For the particularity of the product images, use the perceptual hash value and SIFT feature vectors in the product image retrieval.

A. Perceptual Hash

In the generation process of perceptual hash value of product image, there usually are preprocessing, feature extraction and quantization of perception [6]. The flow chart is shown in Fig. 5. Feature extraction is the most important section of the perceptual hashing algorithm. The hash value extracted from the image has two characteristics: the first is when normal operations (such as pan, zoom, rotate *et al.*) are applied into the image, its hash value can still be extracted; the second is that the hash value remains sensitive for image tampering operation [7]. Quantization of perception turns the feature data extracted from the image into binary vector, which is the hash value of the image, so as to reduce the amount of data [8].

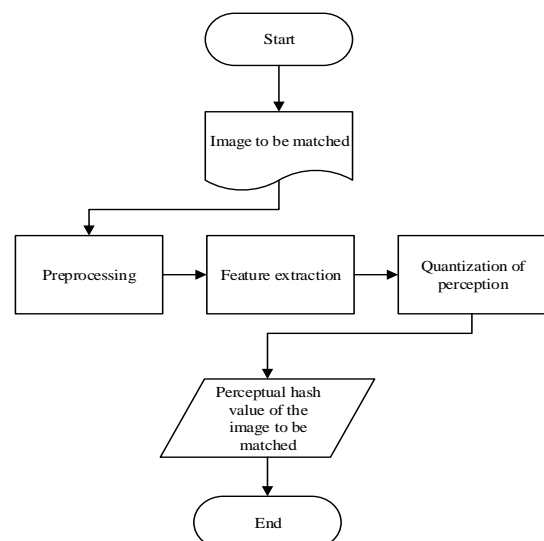


Fig. 4. The generation process of perceptual hash value.

When the perceptual hashing algorithm is applied to the

product image, it can generate a unique "fingerprint" string for each picture. The similarity of product images can be obtained by comparing their fingerprints, the closer the fingerprints are, the more similar the product images are. The flow chart of the perceptual hashing algorithm in the product image retrieval system is shown in Fig. 4.

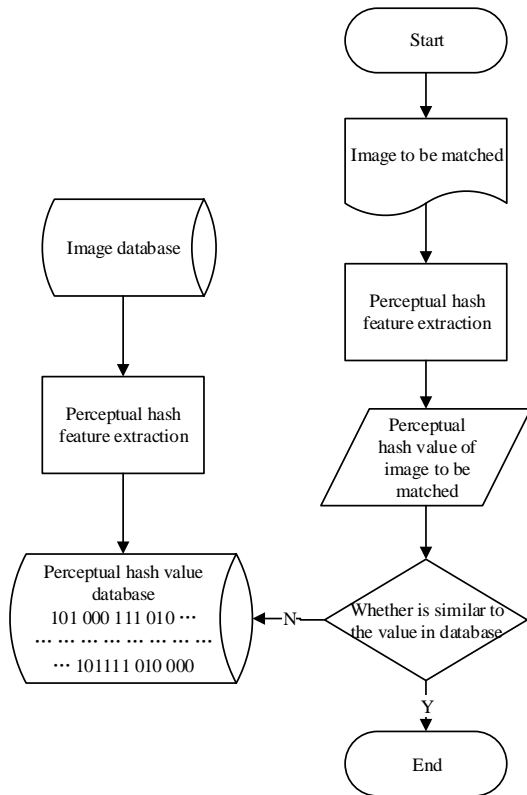


Fig. 5. Perceptual hashing in the product image retrieval system.

Perceptual hash sequence similarity is mainly measured by perceptual hash distance, which is calculated or obtained from the Hamming distance, Euclidean distance or other distance [9]. In current, most perceptual hash algorithm use Hamming distance to measure the similarity of the perceptual hash sequence, and then determine whether the image is similar, so the measure is also used as hash value metrics in the experiments in this paper.

B. SIFT Feature Matching

SIFT (Scale Invariant Feature Transform) is a local feature descriptor to describe the invariant region detection of combined scale and gradient distribution [10]. Compared with the other kinds of image descriptors, Mikolajczyk thinks that the SIFT descriptors has the best image retrieval performance [11]. In 1999, D. G. Lowe proposed the SIFT algorithm, mainly used to detect the local feature of images, then Y. Ke put forward the improvement of the method of replacing the histogram by PCA in part of the descriptors. SIFT has the characteristics of multi volume, uniqueness and extensibility, and maintain a certain degree of invariance and stability to the routine operations of product image such as rotation, scaling, perspective changes, affine transformation, illumination changes *et al.* [12] SIFT feature matching algorithm can be divided into three steps: feature extraction, feature matching and eliminating the bad matching points.

Based on some routine operations of the product image,

this paper compares the matching rate of SIFT feature points before and after the operations through experiments. Fig. 6 to Fig. 9 gives the feature points extraction of the test image after four kinds of routine operation of perspective change, gauss blur, illumination change and rotating amplification. Table I gives the matching rate of feature points after four kinds of routine operation to the original image.

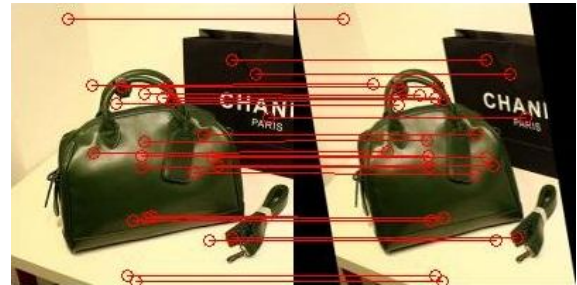


Fig. 6. Perspective change.

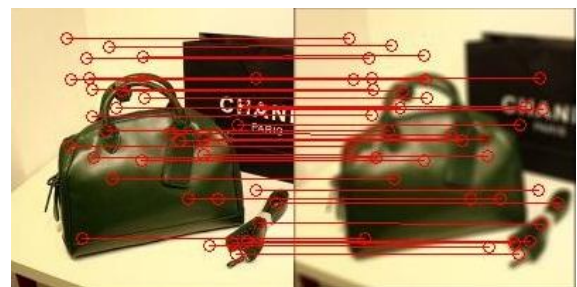


Fig. 7. Gauss blur.



Fig. 8. Illumination change.



Fig. 9. Rotating amplification.

TABLE I: MATCHING RATE OF FEATURE POINTS AFTER FOUR KINDS OF ROUTINE OPERATION

operation	perspective change	gauss blur	illumination change	rotating amplification
rate	89.6%	94.3%	93.5%	85.9%

As can be seen from the table, SIFT feature can be used to describe the product image content, and maintain a certain degree of invariance and stability to the routine operations of perspective change, gauss blur, illumination change and rotating amplification [13]. Therefore, use the SIFT features

as a character of the product image in the product image retrieval system.

After image segmentation to extract the subject of product image, use the SIFT algorithm to extract the feature points of goods which can get the feature vector of 128 dimensional. Then use the Euclidean distance as the measure of feature point descriptor similarity to match feature points, finally use the method presented by Lowe to eliminate the bad matching points, by comparing the distance to nearest neighbor and the second nearest neighbor.

V. NORMALIZATION OF PRODUCT IMAGE RETRIEVAL

In the process of product image retrieval, usually two or more features are used, and each is also characterized by vectors of two or more dimensional, so it needs to be normalized before combining these features [14]. Normalization can be divided into internal normalization and external normalization. Internal normalization is to make the individual vector extracted from one feature stand the same position when measuring similarity. External normalization is normalizing different feature vector, its purpose is to make

each feature vector in a comprehensive feature stand the same position when calculating the similarity distance [15]. When the normalization is applied to product image retrieval, it includes SIFT feature vector space dimension reduction and weight distribution for different features of the image.

SIFT feature vector space dimension reduction is to remove the affection of illumination changes on image retrieval. After using SIFT algorithm to extract the feature points of the product image, a feature vector of 128 dimensions is generated. SIFT feature vector space dimension reduction removes the dimensions that do not care about such as lighting or scale change, which is extracting the main component of SIFT feature vector.

The allocation of appropriate weights for different features of the product image makes their position similar in the distance calculation, thereby reducing the errors caused by uneven proportion on the search results [16]. In the implementation process of product image retrieval system in this paper, the image features include the perceptual hash value and SIFT feature vectors, allocating an appropriate weight for them have a positive impact on the performance of the product image retrieval system.



Fig. 10. The available product gallery.

VI. EXPERIMENTAL RESULTS

Product image retrieval need the support of massive product gallery, in this paper the product image database is generated by the image grabbing script. After running the grabbing script periodically, the available product gallery is shown in Fig. 10, the total number of samples in the experiment is 450.

In the experiment, the control variables method is used to measure the performance of different component in the improved method for product image retrieval. Suppose image segmentation, perceptual hashing, matching feature points and normalization are variables A, B, C, D respectively, controlling each variable can get the following algorithms, as shown in Eq. (1).

$$\begin{cases} F_1 = ABCD \\ F_2 = \overline{ABCD} \\ F_3 = \overline{A}BCD \\ F_4 = AB\overline{C}D \\ F_5 = ABC\overline{D} \end{cases} \quad (1)$$

F1 is a combination of four components, in theory it should have the highest performance in the product image retrieval. F2 does not use the image segmentation, so the subject identification of the product image is not performed. Perceptual hash, feature points matching and normalization are directly applied on the global image to study the performance of product image retrieval. F3 does not use the

perceptual hash, feature points matching and normalization are used to measure the similarity of images based on the commodity subject. In F4 the feature points matching is not used, at first obtained the perceptual hash value of the product subject, and then normalized to retrieve the product image. F5 does not use normalization, based on the commodity subject, obtain the perceptual hash value, then match feature points, and finally get the search results.

Comparing F1 with F2, F3, F4 and F5 respectively, the experimental results are shown in Fig. 11. The horizontal axis represents the recall rate and the vertical axis represents the precision rate, there is a PVR curve to measure the performance of the product image retrieval system.

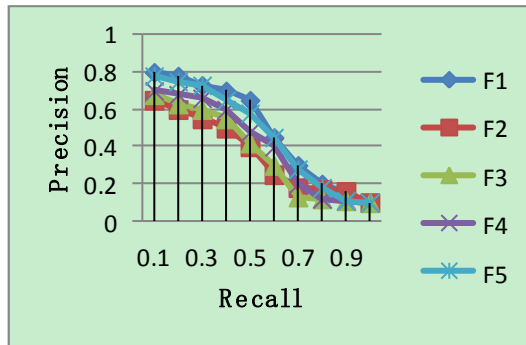


Fig. 11. PVR curve of the experimental results.

As can be seen from the chart, no matter what method based on, the precision is gradually decreasing when the recall increases from the overall trend. It conforms to the basic rules of PVR curve, which is the inverse of the relationship between recall and precision.

In view of the specific data after image segmentation, the overall performance of F2, image retrieval based on the global image is higher than F1, image retrieval based on the commodity subject, and with the increase of recall distance of precision is even wider. When the recall is 0.1, the precision based on commodity subject is 0.8, and the precision based on the global image is 0.65, the initial gap is 0.15. When the recall reaches 0.5, the gap is up to 0.25. Thus, the image segmentation can eliminate the interference of background image feature in the retrieval results, so as to improve the accuracy of product image retrieval.

Using the perceptual hash algorithm to generate "fingerprint" for images has positive effect on the performance of image retrieval, as a result of the precision of F1 using the perceptual hash is much higher than F3 without the use of perceptual hash. When the recall is 0.1, the gap of precision is 0.12. And when the recall reaches 0.6, the precision without using the perceptual hash is close to 0.1, while the precision using the perceptual hash can still maintain a certain height.

As can be seen from the graph, the precision of F4 without using the feature points matching and precision of F5 without using the normalization has no substantial landing, which explains the influence of feature points matching and normalization is smaller than image segmentation and perceptual hash for product image retrieval. Overall, the image retrieval using feature point matching and normalization still has higher precision, but the gap has been maintained in a certain range.

VII. CONCLUSION

Based on the smart mobile device, this paper proposes an improved method including image segmentation, perceptual hash, feature point matching and normalization for the product image retrieval.

Firstly, the edge detection method is used for the product image segmentation to extract the specific goods included in the image. Followed by the product image feature extraction, including the perceptual hash value and SIFT feature vectors. When the perceptual hashing algorithm is applied to the product image, it can generate a unique "fingerprint" string for each image. Perceptual hash algorithm has advantages of summary, robustness, one-way *et al.* The similarity of product images can be obtained by comparing their fingerprints, the closer the fingerprints are, the more similar the product images are. The SIFT algorithm is used to extract the SIFT feature points which can be used to describe the local features of the product image, and the SIFT feature vector can maintain a certain degree of stability to routine operations, such as rotating zoom, perspective changes, affine transformation, brightness variation *et al.* Then the normalization is used for the purpose of SIFT feature vector space dimension reduction and assigning the appropriate weights for different image features. Finally, conduct image feature matching upon the perceptual hash value and SIFT feature vector.

Using recall and precision as the performance evaluation criteria of the product image retrieval system, this paper also studies the influence of each component in the improved method on the performance of product image retrieval through experiments. Experimental results show that the improved method increases the efficiency of the product image retrieval system in a certain extent.

REFERENCES

- [1] CBIR. Wikipedia. [Online]. Available: <http://zh.wikipedia.org/wiki/CBIR>
- [2] Y. Hong, "Research on content based image retrieval technology," *China New Technologies and Products*, vol. 8, p. 20, April 2010.
- [3] F. J. Meng and B. L. Guo, "Research on content-based image retrieval technology," *Application Research of Computers*, vol. 21, no. 7, pp. 21-24, July 2004.
- [4] B. Xu and C. Y. Wei, "Review on the new methods of image segmentation," *Public Communication of Science & Technology*, no. 8, p. 112, June 2010.
- [5] R. B. Huang, G. B. Liang, and M. Y. Du, "Object region based color image retrieval integrating multi-features," *2007 Doctoral Forum of China*, no. 9, 2007.
- [6] X. M. Niu and Y. H. Jiao, "An overview of perceptual hashing," *Acta Electronica Sinica*, vol. 36, no. 7, September 2008.
- [7] T. Kalker, J. Haitma, and J. C. Oostveen, "Issues with digital watermarking and perceptual hashing," in *Proc. SPIE Conf. on Multimedia Sys. & Appl.*, August 2001, pp. 189-197.
- [8] Z. Q. Liu, "Research on key technologies of image perception of hash," Ph.D thesis, Dept. Comput., Harbin Institute of Technology, Harbin, 2013.
- [9] Y. Zeng, "Image perceptual hashing research and application," M.S. thesis, Dept. Comput., Zhejiang SCI-TECH Univ, Zhejiang, 2012.
- [10] G. D. Lowe "Object recognition from local scale-invariant features," *IEEE Trans. Computer Vision*, vol. 2, pp. 1150-1157, September 1999.
- [11] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. on PAMI*, vol. 27, pp. 1615-1630, October 2005.

- [12] C. L. Lu, L. H. Ma, and H. Chen, "An improved algorithm for SIFT feature matching," *Journal of Air Force Engineering University*, vol. 15, no. 1, February 2014.
- [13] T. Z. Bai and X. B. Hou, "An improved image matching algorithm based on SIFT," *Transactions of Beijing Institute of Technology*, vol. 33, no. 6, 2013.
- [14] D. Y. Han, "A method of integrated features normalization in content based image retrieval," *Journal of Weifang University*, no. 2, June 2006.
- [15] R. Xu and D. J. Huang, "Research on image retrieval based on color and texture," *NEI JIANG KE JI*, vol. 28, no. 11, pp. 106-119, November 2007.
- [16] L. Chen, "The Research of target region based on image retrieval system," M.S. thesis, Dept. Computer., Zhejiang University of Technology, Zhejiang, 2009.



Yuzhong Zhang was born on April 12, 1990 in Tianjin, China. He received the bachelor degree in computer science from Tianjin University of Technology, China in 2012. Currently he is doing his master degree in software engineering at Beijing Institute of Technology, China. His research interests include multimedia processing for image, software technologies.



Chongwen Wang was born on May 25, 1974 in Henan Province, China. He received PhD degree in instrument science from Chongqing University, China in 2002. Currently he is an associate professor at Beijing Institute of Technology, China. His research interests include digital media technology, software engineering.