Review of Image Expression of Content-based Image Retrieval Technology

Shicong Cao
No.24, Block 1, Xuefu Road, Chengdu 610225, China
595532686@qq.com

Abstract. With the continuous development of science and technology, the popularity and use of projectors, cameras and other electronic products have led to an ever-expanding image source. Due to the enormous workload of text-based image retrieval (TBIR) annotation, content-based image retrieval (CBIR) was proposed to solve this problem. The CBIR has experienced the process featured by an integration of image digital, image semantic, and other multiple image features. CBIR, however, still has shortcomings in embodying the high-level semantics of the image. As the ability of autonomously learning image features, the defects of CBIR in image expression can be effectively solved by deep learning, which thus has become a hot topic in current image expression researches. On the basis of content-based image retrieval technology, some readings and analyses are made in the article for exploring the future development of image retrieval technology in image expression field.

1. Introduction
With the popularity of multimedia equipment and technology and the fast development of internet technology, billions of images are shared and spread across the internet. The emergence of CBIR plays an irreplaceable role in meeting the needs of people to find the specific they desired from a large quantity image [1].

Since it was proposed, a large number of researches about content-based image retrieval technology have been made by many researchers and scholars at home and abroad. In the first place, element information based-image’s peripheral area were adopted by researchers and scholars as the tool to satisfy people's needs for image retrieval technology, such as adding labels and titles to the images so that people could search for those they needed according to a large amount of image annotation information[2-3]. Due to the fact that there were some differences between the image information with text descriptions and added labels and the actual visual information, however, the result of the final search was not ideal. Therefore, in order to solve the problems above, CBIR had been proposed and made great progress in the speed and image-preparation rate of the image retrieval under the efforts of a large number of researchers and scholars [4].

Since the proposal of content-based image technology, the key issue of researches had gone into how to effectively express the image. At present, the image expression in image retrieval technology was mainly processed in the following two ways: one was based on the manual extraction and fusion of features; another was based on the deep learning [5]. The researches of based-manual extraction of features were mainly conducted from two aspects: Scale-Invariant Feature Transform (SIFT), which was a kind of local feature, and the Bag of Words model (BoW). SIFI, as an image retrieval feature, had its own huge advantage due to its excellent rotatability, scale invariance and robustness on light [6]. As for another image expression way based on deep learning, it mainly focused on Convolutional
Neural Networks (CNN) and its improvement methods. The proposal of CNN made its contribution in the easement of the complexity and instability in engineering of manual extraction of features to a large extent. Therefore, it was of great significance to apply CNN to design image retrieval system [7].

No matter which technique method is used to extract picture features, it will face the problem of how to improve the retrieval efficiency and performance. For the method based on manually extracting digital image features, the problem is how to extract image features effectively and how to improve the retrieval performance in more effective ways through these image features. For the method based on deep learning, the problem it will face is how to design a reasonable network to improve the retrieval performance. Therefore, according to the above two methods, some summaries and analyses against image expression by using image retrieval technology are made in the article.

2. The Fundamental Principle of Content-based Image Retrieval
The basic framework of content-based image retrieval is shown in Figure 1. The figure shows that CBIR system mainly consists of two phases—offline and online. In the offline phase, the first is to establish the internet image database which is needed through crawler system, and then convert the image information into eigenvector of digital image and build the index database. In online phase, it mainly consists of analysis of user intention, search condition, image expression, image correlativity score, result ranking and output module.

![Figure 1. The basic framework of CBIR](image)

The core issue of the CBIR system is how to distinguish the difference of each image and effectively match the desired from a large number of images, i.e., the image expression and the image similarity. The image expression mainly consists of extraction, quantization, aggregation of feature and other issues. In this article, the focus is placed on the research status of manual extraction and automatic generation of features.

3. The Image Expression Methods of Based-manual Extraction of Digital Image Features
As it can be seen from the previous section, image expression has a very important position in image retrieval system and makes a fruitful progress under the efforts of lots of researchers and scholars. For a long time, the manual design image features have been a hot topic in the field of image expression. In the early days, the color, shape, and texture of image were often regarded as the image expression way to describe the contents contained in images [8]. Of which, Generalized Search Trees (GIST) is the most widely used global feature because of its low computation complexity. Although global features had been once a research hotspot in image feature extraction, they became the bottleneck of researches due to their low accuracy in the processing of complex images [9].

In order to solve the above problems, a comparative analysis of the commonly used image features extraction algorithms was conducted in reference [8]. The conclusion of this article was that the colour feature was the most commonly used local image feature. In reference [10], the calculation of colour feature of Hue, Saturation, Value (HSV) domain based on the histogram was performed with the pixel average value as the assessment standard. From the experimental results, it could be seen that the method had no obvious effect on CBIR system, but a certain degree of improvement on semantic-
based image retrieval (SBIR). For the poor effect caused by the irregular texture feature of Gabor-based image, a POCBS-based texture feature extraction algorithm was proposed in reference [11]. This method filled the irregular region with numbers to generate a regular shape of a square shape. The mean and variance in each sub-band were used as the final extracted texture feature. From the experimental result, the POCBS-based texture feature had a certain improvement in the accuracy of the retrieval compared with other texture features. A new kind of target region feature--colour and scale feature--was put forward in reference [12]. This was a feature which combined colour with the size and distribution information of the region. In the end, these information were integrated as extracted feature through linear weighting. In reference [13], root-SIFT feature was obtained by calculating the square root (the operation of normalization) of basic SIFT feature. The experimental result showed that root-SIFT feature had an enhancement in the accuracy and stability of the retrieval compared with the basic SIFT feature. The mid-value was re-graded as the threshold value to generate the binary image feature of SIFT in reference [14]. On the basis of such binary feature, a new index method of image retrieval was raised in this reference. As for the difficulty in matrix partition existing in the binary feature of SIFT, a new method to compare matrix through dimension was proposed in reference [15]. It could be seen in the experimental result that such index method was more convenient in the process of extracting features. The Affine-SIFT feature was put forward in reference [16]. Such method effectively handled the impact of affine transformation and achieved full affine invariance by adjusting the longitude and latitude of the camera. In reference [17], according to the distinction of the region extraction of ill structured SIFT feature, the samples were randomly selected from the 0-225 discrete random variable. Then, entropy could be used to standardize the samples and these standardized samples were regarded as the feature extracted from SIFT. From the experimental results, it could be concluded that this method greatly improved the performance of image retrieval.

The algorithm which used Difference of Gaussia (DoG) to extract image local features was proposed in reference [18]. The experimental results showed that such algorithm indexed faster than that of SIFT feature. Reference [19] came up with ORB feature extraction method, which could eliminate the noise and correlation of images. Compared with SIFT feature, it was more efficient in image retrieval.

The extracted features were often merged so that the information of the image could be described more comprehensively. There were mainly three methods for commonly used local feature fusion: Bow expression, VLAD and Fisher Vector. In reference [21], sparse coding was applied to local features, and then the VLAD hierarchy was established by aggregating coding in the way of max-pooling. Bag-of-visual-words (BoVW) of hidden layer was also introduced into such hierarchy. Residual vector was used to make the distribution more uniform so that the feature expression could be easier to distinguish. Reference [22] grouped key points on the image level and then aggregated the local features in each group to obtain a searchable image feature with high accuracy. The gradient vector, which was aggregated and normalized by Gaussian mixture model (GMM), was used to feature fusion in reference [23].

The manual method of extracting image features has attracted the attention of many scholars. From the analysis of the reference retrieval results mentioned above, it can be concluded that the feature fusion method performs better, and is also the first choice of image expression when designing an actual complex image retrieval system.

4. The Image Expression Methods of Based-deep learning

With the breakthrough and development of deep convolutional neural networks, deep learning has been used in many industries in recent years. Because deep learning can automatically extract and learn related information of the image, the use of deep convolutional neural networks to process image expression in image retrieval has become a current trend. Deep-restricted Boltzmann Machine was used to extract local image features in reference [24]. Compared to most of the manual image feature extraction methods, the extraction method using Boltzmann machine is more efficient in image retrieval according to the research result. In reference [25], the Alex-Net neural network was selected
to extract image features with the activation value of the six-layer neural network as the deep output feature of images. Furthermore, the author combined the output feature with SIFT, HSV, and GIST features as the final extracted image feature. By doing so, the advantages of local features were also utilized in the extraction process. From the final output, it could be seen that such extraction method greatly improved the speed and accuracy of image retrieval compared with the manual extraction of features. In reference [26], the author drew on the VGG-Net which was opened source by Google to extract image features and made an attempt to take the final Convolutional Layer--max pooling--as the extracted feature. The Convolutional Neural Network (CNN) was adopted to extract local image features and applied to the actual system in reference [27]. The experimental results showed that such method had a better performance in image retrieval. In reference [28], the CNN was also adopted to establish an image expression method in block-level without supervision, which was an ideal method. In reference [29], the author suggested that the image could be converted into hash binary code through CNN, which could reduce the loss caused by the ranking of image features. From the experimental result, it could be clearly seen that there is a certain increase in search efficiency and search complexity compared with other networks.

5. Conclusion and Outlook

This article sums up the researches on image expression of content-based image retrieval (CBIR) technology in recent years, focusing on the research results and status quo of based-manual extraction features as well as the automatic extraction of features of based-deep learning. Through the above analyses and summaries, a conclusion is founded that the image expression in future image retrieval will be developed toward the trend of combining automatic learning features with manual extraction features in pace with the development of deep learning and the image expression in image retrieval.

References

[1] Cheng M M, Zhang Z, Lin W Y, et al. Bing: Binarized normed gradients for abjectness estimation at 300 fps [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2014: 3286-3293
[2] Li X R, Uricchio T, Ballan L, et al. Socializing these mantic gap: A comparative survey on image tag assignment, refinement, and retrieval [J]. ACM Computing Surveys, 2016, 49(1) : 14
[3] Lin Z J, Ding G G, Hu M Q, et al. Semantics-preserving hashing for cross-view retrieval [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2015:3864-3872
[4] Lew M S, Sebe N, Djeraba C, et al. Content-based multi-media information retrieval: State of the art and challenges[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2006, 2 ( 1 ) :1-19
[5] TZEL.EPI M, TEFAS A. Relevance feedback in deep convolutional neural networks for content based image retrieval[C]// Proc. Of the Hellenic Conference on Artificial Intelligence, 2016, 27-28
[6] Zhang Leilei, Chen Lifang. A Novel Image Retrieval Method Based on Combinatorial Improvement Clustering of SIFT Feature Points [J]. Software Guide, 2017, 16 (9)
[7] Bi Xiaojun, Pan Tiewen. Image retrieval method with relevance feedback based on improved teaching-learning-based optimization algorithm [J]. Systems and Engineering and Electronics, 2017:39 (10)
[8] Liu Y, Zhang D S, Lu G, et al. A Survey of Content based Image Retrieval with High-Level Semanties [J]. Pattern Recognition, 2007, 40(1): 262-282
[9] Gao Z N, Xue J R, Zhou W G, et al. Fast democratic aggregation and query fusion for image search [C] // ACM International Conference on Multimedia Retrieval, 2016:35-42
[10] Zheng L, Wang S J, Tian L, et al. Query-adaptive late fusion for image search and person re-identification [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1741-1750
[11] Mina N, Ershad B, James T. Radon-Gabor barcodes for medical image retrieval[C]// Pro the 23th International Conference on Pattern Recognition, IEEE, 2016:1309-1314
[12] Dong Xiaoli, Zhang Lihe, Mi Xiaoli. Color image feature extraction using color index local correlations and statistics [J]. Journal of Optoelectronics Laser, 2011, 22 (4): 623-528
[13] Arandjelovic R, Zisserman A. Three things everyone should know to improve object retrieval [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2012: 911-2918
[14] Zhou W G, Lu Y J, Li H Q, et al. Scalar quantization for large-scale image search [C] // ACM International Conference on Multimedia, 2012: 169-178
[15] Liu Z, Li H Q, Zhang L Y, et al. Cross-indexing of binary SIFT codes for large-scale image search [J]. IEEE Trans-actions on Image Processing, 2014, 23( 5) : 2047-2057
[16] Heo J P, Lee Y, He J F, et al. Spherical hashing [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2012: 2957-2964
[17] Tang J H, Li C Z, Wang M, et al. Neighborhood discriminant hashing for large-scale image retrieval [J]. IEEE Transactions on Image Processing, 2015, 24 (9) : 2827-2840
[18] Avrithis Y, Tolias G, Kalantidis Y. Feature map hashing: Sub-linear indexing of appearance and global geometry [C] // International Conference on Multimedia, 2010:231-240
[19] Jiang K, Que Q C, Kulis B. Revisiting kernelized locality sensitive hashing for improved large-scale image retrieval [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2015: 4933-4941
[20] Wu Z, Ke Q F, Sun J, et al. Scalable face image retrieval with identity-based quantization and multi reference ranking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33( 10) : 1991-2001
[21] Ge T Z, Ke Q F, Sun J. Sparse-coded features for image retrieval [C] // British Machine Vision Conference, 2013: 132
[22] Duan L Y, Gao F , Chen J, et al. Compact descriptors fromobile visual search and MPEG CDVS standardization [C] // IEEE International Symposium on Circuits and Systems, 2013: 885-888
[23] Hrster E, Lienhart R. Deep networks for image retrieval on large-scale databases [C] // ACM International Conference on Multimedia, 2008: 643-646
[24] Krizhevsky A, Sutskever I, Hinton G E. Image Net classification with deep convolutional neural networks [C] // International Conference on Neural Information Processing Systems, 2012: 1097-1105
[25] Sun S Y, Zhou W G, Tian Q, et al. Scalable object retrieval with compact image representation from generic object regions [J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2015, 12 (2) : 29
[26] Paulin M, Douze M, Harchaoui Z, et al. Local convolutional features with unsupervised training for image retrieval [C] // IEEE International Conference on Computer Vision, 2015: 91-99
[27] Lai H J, Pan Y, Liu Y, et al. Simultaneous feature learning and hash coding with deep neural networks [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3270-3278.