Theory and Practice of Science and Technology
2024, VOL. 5, NO. 6, 63-68
DOI: 10.47297/taposatWSP2633-456909.20240506

Research on Image Content Retrieval Technology Based on Visual Feature Extraction

Chengwei Bai

Hohhot No.1 High School computer 33 Huanhe Street, Huimin District, Hohhot, Inner Mongolia, China

ABSTRACT

Image content retrieval technology can quickly and accurately locate similar images in large-scale image databases by analyzing the visual features of images, and has become one of the core technologies for multimedia data management. This paper systematically reviews image content retrieval technology from the perspective of visual feature extraction, and deeply analyzes the application status and performance of low-level features (such as color, texture, shape), mid-level features (such as edges and points of interest), and high-level semantic feature extraction methods based on deep learning. Low-level and mid-level feature extraction methods can effectively describe the basic information of images and are suitable for image scenes with relatively simple structures. Deep learning models improve the retrieval accuracy in complex scenes through multi-layer feature expression capabilities and show strong generalization capabilities. This paper further discusses the main challenges faced by image content retrieval in visual feature extraction, including feature calculation overhead, adaptability to complex scenes, etc., and points out the importance of improving feature extraction efficiency and retrieval system accuracy, providing a theoretical basis and practical reference for the development and application of related technologies.

KEYWORDS

image content retrieval; visual feature extraction; deep learning; low-level features; mid-level features

1 Introduction

1.1 Research background and importance

With the explosive growth of image data, image content retrieval technology has become crucial in the modern information society. Traditional text-based retrieval methods rely on manual labels, which not only face challenges in label accuracy and consistency, but also have difficulty in effectively dealing with large-scale image databases. Content-based image retrieval (CBIR) methods use visual features of images for retrieval, so they have significant advantages in efficiency and accuracy and have become an important direction for solving image retrieval problems[1-2]. CBIR systems extract low-level visual features of images such as color, texture, and shape for retrieval, allowing the system to more accurately express the content information of images[3].

In recent years, with the advancement of deep learning technology, image content retrieval systems based on deep learning have been widely used and have significantly improved retrieval performance [4]. For example, convolutional neural networks (CNNs) can effectively process complex image content through multi-level feature extraction, making automated retrieval of large-scale image databases possible[5]. However, despite the superior performance of these methods, visual feature extraction still faces challenges, especially how to accurately extract and express image content with complex semantics[6].

1.2 Research Objectives

First, the basic framework of current image content retrieval technology will be analyzed, and the core steps and key technologies will be explored. Secondly, through in-depth research on visual feature extraction methods, the application and performance of low-level visual features (such as color, texture, shape) and high-level semantic features in image retrieval will be summarized. Based on this, this study will also evaluate the impact of deep learning on image content retrieval, especially the application potential of convolutional neural networks in visual feature extraction.

2 Basic technologies of image content retrieval

2.1 Overview and classification of image content retrieval

Content-Based Image Retrieval (CBIR) technology realizes image retrieval by analyzing the content features of images (such as color, texture, and shape). CBIR systems have been widely developed since the 1990s and have gradually become one of the key technologies in multimedia database management[7]. Traditional image retrieval methods are mainly divided into two categories: those based on global features and those based on local features[8]. In recent years, with the application of deep learning, the technology of extracting deep semantic features based on convolutional neural networks (CNN) has gradually developed, which can more effectively capture complex semantic information in images, making CBIR more advantageous in dealing with large-scale data[9]. Figure1 represents the Content-Based Image Retrieval (CBIR) process, incorporating a feedback loop that allows users to iteratively refine search results. The system adjusts based on user feedback until the desired images are returned, enhancing retrieval precision through iterative interaction.

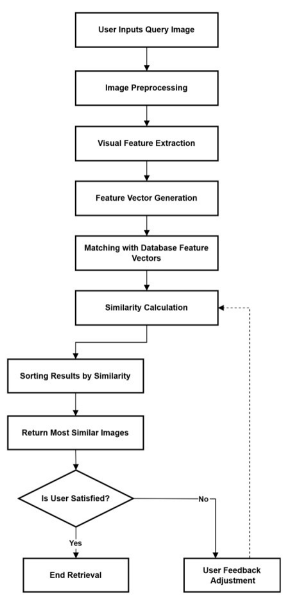

Figure1 Content-Based Image Retrieval (CBIR) Process

2.2 Image feature extraction methods

Image feature extraction is the core step of the CBIR system, which determines the system's ability to express image content [10]. Common feature extraction methods include color, texture, and shape features (see Figure 2 for the extraction process). Color features are often quantified using color histograms, while texture features can be expressed through grayscale co-occurrence matrices or Gabor filters. Shape feature extraction focuses on edge detection and contour analysis, usually using algorithms such as SIFT or SURF. In recent years, deep learning models, especially convolutional neural networks (CNNs), have been particularly prominent in feature extraction, improving retrieval results through their multi-level feature representation capabilities [11]. In addition, multi-feature fusion methods have also been applied in practice, integrating color, texture, and shape features to improve the robustness and accuracy of retrieval systems [12].

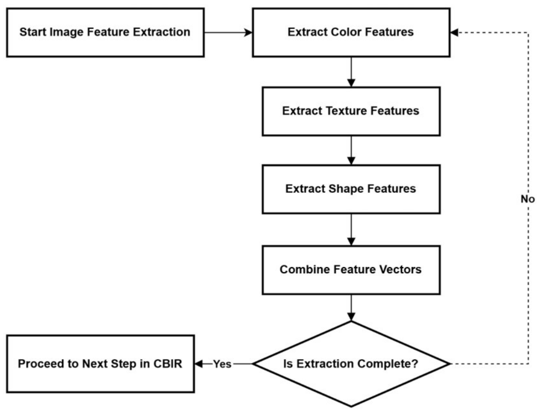

Figure 2 The extraction process

2.3 Evaluation criteria of retrieval system

The performance evaluation of image content retrieval system is usually measured by precision and recall. Precision indicates the proportion of retrieved results that are truly relevant to the query image, while recall measures the proportion of all relevant images that are successfully retrieved [13]. In recent years, with the application of deep learning, system response speed has also become one of the evaluation criteria. Retrieval efficiency has been significantly improved by optimizing the neural network structure and reducing the dimension of feature vectors. In addition, some CBIR systems introduce user feedback mechanisms to continuously optimize retrieval results through interactive feedback, so as to be closer to the actual needs of users. In the performance evaluation process, experiments are usually conducted based on standard data sets to ensure objective comparison of different algorithms[14].

3 Application of visual feature extraction in image retrieval

3.1 Low-level feature extraction

Low-level feature extraction mainly starts from the basic visual attributes of the image, including features such as color, texture and shape. Color features, as the most intuitive information of the image, are usually expressed in the form of color histograms, color moments, etc., and retrieval is achieved by analyzing the color distribution pattern of the image. Texture features characterize image content by capturing properties such as the repetitive pattern of the image surface structure or the coarseness of the texture. Common methods include gray-level co-occurrence matrix and Gabor filter. Shape features focus on the description of the outline or boundary of the object, such as extracting shape information through edge detection or Hough transform. These low-level features can effectively characterize the basic content information of the image and play an important role in image content retrieval, especially for images with relatively simple colors and structures.

3.2 Mid-level feature extraction

Mid-level feature extraction usually focuses on more prominent local areas or points of interest in the image to capture more recognizable structural information. Edge features mainly reflect the outline and boundary of objects in the image, and are an important part of shape and structure. They are often extracted through techniques such as Canny edge detection or Sobel operator. Interest point detection uses algorithms such as SIFT and SURF to identify key points in the image that are rotationally and scale-invariant. This type of feature not only has good contrast between different images, but also remains consistent when the image is rotated or scaled, making it highly robust in image retrieval with scene changes. Figure 3 is an example of mid-level feature extraction.

Figure 3 Example of Middle-Level Feature Extraction

3.3 Application of deep learning in feature extraction

With the development of deep learning, the application of convolutional neural network (CNN) in image feature extraction has gradually become popular. Different from the traditional manually designed feature extraction method, CNN automatically extracts hierarchical features from images through multi-layer convolution, pooling and full connection operations, and can effectively capture the high-level semantic information of the image. Deep learning models are particularly suitable for processing complex image data. By training on large-scale data sets, they can identify specific targets such as faces and objects. This type of feature extraction method has shown excellent retrieval effect in image content retrieval, which not only greatly improves the accuracy, but also can adapt to a variety of application scenarios, laying the foundation for the intelligent development of image retrieval.

4 Research Challenges and Future Prospects

4.1 Existing Problems of Image Retrieval Technology

Although the current image content retrieval technology has made significant progress, there are still many challenges. First, the retrieval system is easily disturbed when dealing with complex backgrounds or similar objects, resulting in a high false detection rate. In addition, existing methods mostly rely on low-level features or local features, lack a comprehensive description of images in diverse scenes, and it is difficult to effectively distinguish different objects with similar colors or textures. Although deep learning models have improved the accuracy of retrieval, they require a large number of annotated data sets for training, which places high demands on data storage and computing resources. In addition, image retrieval systems still need to be optimized in terms of response speed and real-time performance to meet the needs of fast retrieval in large-scale image libraries.

4.2 Key Challenges in Visual Feature Extraction

Visual feature extraction plays a core role in image retrieval, but it also faces key challenges. Low-level features such as color and texture often lack sufficient discrimination, and it is difficult to accurately extract target information in complex scenes or backgrounds. The detection of mid-level features places high demands on the robustness of the algorithm, especially under conditions of image deformation, occlusion, and illumination changes, which are prone to feature loss or deviation. In addition, the extraction of high-level semantic features requires deep learning models, which have high computational overhead and require a large amount of annotated data, limiting their application in resource-limited environments. Therefore, how to improve the robustness and efficiency of feature extraction remains an important research direction.

4.3 Future Research Directions and Application Prospects

In the future, image retrieval technology will develop in the direction of intelligence and efficiency. The technology based on multi-feature fusion is expected to improve the accuracy of retrieval. By combining color, texture, shape and deep semantic features, the system has higher adaptability in multiple scenarios and complex environments. In addition, the application of few-shot learning and unsupervised learning will be gradually promoted, which will promote the popularization of image retrieval by reducing the dependence on a large amount of annotated data. With the optimization of computing resources, lightweight deep learning models and distributed computing are expected to improve the real-time performance of retrieval systems. In general, image retrieval technology will have broad application prospects in the fields of intelligent monitoring, medical image analysis, e-commerce, etc.

5 Conclusion

Image content retrieval technology can quickly and accurately locate relevant images from large-scale data sets by analyzing the visual features of images, and has become a key tool for multimedia information management and retrieval. Based on the systematic analysis of image content retrieval technology, this paper conducts a detailed review of various methods of visual feature extraction, including low-level features (such as color, texture, shape), mid-level features (such as edges and points of interest), and deep learning applications of high-level semantic features. These feature extraction methods show their own advantages and disadvantages in different scenarios: low-level and mid-level feature extraction are suitable for retrieval tasks with simple image structures and bright colors, while deep learning, with its powerful feature expression ability, shows high accuracy and robustness in complex scenarios and diverse data. However, existing image retrieval technologies still face challenges in dealing with complex backgrounds, interference from similar objects, and feature calculation overhead, especially the labeling requirements of large-scale data sets impose strict requirements on resources and costs. Therefore, improving the accuracy, efficiency, and robustness of feature extraction remains an important direction for future research.

In the future, image content retrieval technology will continue to develop in the direction of intelligence and diversification, especially driven by emerging technologies such as multi-feature fusion, few-sample learning, and unsupervised learning. It is expected to break through the limitations of traditional methods and achieve efficient processing of complex image content. With the improvement of hardware computing power and the innovation of lightweight deep learning models, the real-time and universality of image retrieval systems will be further improved, and the application prospects will be broad. In the future, this technology has great potential in the fields of intelligent monitoring, medical image analysis, e-commerce, etc., and will provide more efficient solutions for multimedia data management.

References

[1] Jigisha M. Patel, Nikunj Gamit.: A review on feature extraction techniques in Content Based Image Retrieval. 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2259-2263 (2016).

[2] N. Vassilieva.: Content-based image retrieval methods. Programming and Computer Software, 35, 158-180(2009).

[3] R. Saritha, V. Paul, P. G. Kumar.: Content based image retrieval using deep learning process. Cluster Computing, 22, 4187-4200(2019).

[4] Aasia Ali, Sanjay Sharma.: Content based image retrieval using feature extraction with machine learning. 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 1048-1053(2017).

[5] Hiroki Tanioka.: A Fast Content-Based Image Retrieval Method Using Deep Visual Features. 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW), 20-23(2019).

[6] Zahid Mehmood, Fakhar Abbas, Toqeer Mahmood, M. Javid, A. Rehman, Tabassam Nawaz.: Content-Based Image Retrieval Based on Visual Words Fusion Versus Features Fusion of Local and Global Features. Arabian Journal for Science and Engineering, 43, 7265-7284(2018).

[7] Sagarmay Deb, Yanchun Zhang.: An overview of content-based image retrieval techniques. 18th International Conference on Advanced Information Networking and Applications (AINA 2004), 59-64(2004).

[8] D. Cerra, M. Datcu.: Image retrieval using compression-based techniques. 2010 International ITG Conference on Source and Channel Coding (SCC), 1-6(2010).

[9] Zhao Zhi-we.: Content-Based Image Retrieval Techniques. Shandong Electronics, (2005).

[10] Gholamreza Rafiee, S. Dlay, W. L. Woo.: A REVIEW ON CONTENT BASED IMAGE RETRIEVAL. 2010 7th International Symposium on Communication Systems, Networks & Digital Signal Processing (CSNDSP 2010), 775-779(2010).

[11] Mona Mahrous Mohammed, A. Badr, M. B. Abdelhalim.: Image classification and retrieval using optimized Pulse-Coupled Neural Network. Expert Systems with Applications, 42, 4927-4936(2015).

[12] Ms. Jyoti D. Gavade, Mrs. Gyankamal J. Chhajed, Ms. Kshitija A. Upadhyay.: Review on Image Retrieval Systems. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Energy, 2, 945-949(2013).

[13] Effat Naaz, T. Kumar.: Enhanced content based image retrieval using machine learning techniques. 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), 1-12 (2017).

[14] P. Gosselin, M. Cord.: A comparison of active classification methods for content-based image retrieval. ACM International Conference on Multimedia, 51-58(2004).