ISSN: 2710-0170

Visual Analysis on the Research of Color-based Image Retrieval -- Based on Bibliometric Analysis

Yizhou Su, Kangjian Xu*

Nanjing University of Finance and Economics, Nanjing, Jiangsu, 210000, China

*Email: 1652815583@qq.com

Abstract

This study retrieved relevant literature on color-based image retrieval in the Web of Science core database, and used bibliometrics and CiteSpace 6.1.R2 software to visually analyze the authors, institutions, and keywords of literature published in this field from 2005 to 2023. This study summarized the current status and development trend of color-based image retrieval and combined information extraction to select and include 716 articles for analysis. The results showed that: (1) In the last 20 years, the number of publications in this field has gone through a slow budding phase, a fluctuating growth phase, and an explosive growth phase; research in this area has been highly active in recent three years. (2) The authors of this field have a relatively even distribution of publications and some authors have formed close collaborative groups. (3) The research institutions mainly include research institutes and comprehensive universities, which have close collaborations with each other. (4) Co-occurrence and clustering analysis of keywords showed that research in this field is mainly focused on the main paths, research methods, and technological tools of color-based image retrieval, as well as key objects and potential influencing factors in the technological process. (5) Emerging keywords analysis predicted that more precise color description and extraction technology will be the focus of future research. In conclusion, color-based image retrieval is still in a hot research and technology explosion phase, and cooperation between countries, institutions, and authors should be strengthened to conduct more in-depth research.

Keywords

Image Retrieval; Color; Bibliometric Analysis; Citespace.

1. Introduction

Research on content-based image retrieval (CBIR) began as early as the 1970s, which uses textual annotation to describe the content of images and formulates keywords for each image, such as objects and scenes. It can be done manually or semi-automatically through image recognition technology[1]. This text-based approach to image retrieval has quickly gained popularity in the market due to its advantages of speed, ease of implementation, and high accuracy, and has been widely applied in various industries such as Google, Taobao, Amazon, and social platforms like Pinterest. However, its limitation lies in the restricted annotation process, and users sometimes find it challenging to obtain the desired images[2]. Thus, color-based image retrieval gradually enters the research field based on this foundation. In 2008, Chan, YK, and Liu,YT began to focus on color differences among the edges of objects in an image. They constructed an image retrieval system based on CDESSO feature to capture the feature of color differences among the edges of objects, and experimental results showed that this system can effectively and quickly display desired images to users[3]. Subsequently, more and more

ISSN: 2710-0170

scholars began to pay attention to this field. Xing-Yuan, W proposed an effective two-level color image retrieval method based on the RGB color model in 2011, which can improve the retrieval accuracy measured in terms of recall and precision[4]. In 2013, Park, Y proposed an image retrieval method using color and shape features, believing that using both color and shape information could make image retrieval more accurate[5]. In 2022, Liu, H proposed to first apply the Grab-Cut automatic segmentation algorithm, and then use the CCV and main color methods to extract color features, resulting in more accurate image retrieval results[6].

Due to the cutting-edge nature of the technology, existing literature has a generally recent publication year distribution, mainly focusing on the last few years. Although many studies and review analyses have been conducted on the main paths, research methods, and technical tools for color-based image retrieval, as well as key objects and possible influencing factors in the technical process[7][8][9][10], no relevant literature has reported it from the perspective of bibliometrics.

CiteSpace is an information visualization software developed by Professor Chaomei Chen's team for measuring and analyzing scientific literature data[11]. It can visually display the development trend of a certain discipline or knowledge field during a certain period and analyze the evolution of the research frontier areas[12]. The software can provide valuable knowledge map overviews in a timely and systematic manner, analyze research progress, and predict future research directions, providing a reference for subsequent studies. This article aims to use CiteSpace 6.1.R2 software to conduct quantitative analysis of literature published from January 2005 to September 2023, explore the research dynamics and hotspots of color-based image retrieval at home and abroad, summarize and generalize the research status and trends, and provide theoretical references for relevant researchers in the field.

2. Data and Methods

2.1. Data Source

In the Web of Sience (WoS) database, the search formula was set to: Topic = computer image retrieval OR color OR recognition, document type was set to article or review, and the time span was from January 2005 to March 2023, with the database being set to the Web of Science Core Collection.

2.2. Data Processing

The focus of the search was on the title of literature, rather than the content of literature, as there are many other fields in which content related to image retrieval appears, but it is not our research focus. The search results were de-duplicated, and conference calls, comments, news, and corrections were removed. Then, the results were manually screened, and finally 716 literature were screened out in the Web of Science.

The literature in Web of Science was exported in plain text format, and the content recorded included full records and cited references. The files were named "Download-**.txt" and imported into the CiteSpace 6.1.R2 software for de-duplication and format conversion, then data analysis was conducted using a combination of quantitative and qualitative methods.

Specifically, data analysis followed the usage method introduced by Professor Chaomei Chen regarding the CiteSpace software. The Refworks file was imported into CiteSpace, and the data was converted into a format that the software can analyze, imported into the final data, and the time slice, node type, data selection mode, and network trimming path were set. A co-occurring keywords knowledge map was generated through data operation and, based on this, relevant analysis of keyword clustering and keyword emergence was conducted.

Time slicing was set from January 2005 to March 2023, with two-year intervals per slice. Node types were set to author, institution, and keyword, with top N = 50. Pathfinder and pruning

ISSN: 2710-0170

sliced networks were used as the graph pruning algorithm to process the obtained graph, and the remaining settings used the default values of the CiteSpace 6.1.R2 software system.

2.3. Data Visualization

Based on the above settings, visual analysis was conducted on the 716 literature included, including annual publication volume, author collaboration network, institution collaboration network, co-occurrence analysis, clustering analysis, and emergence analysis of keywords related to color-based image retrieval technology and research, creating a knowledge map. The general information of the literature was plotted using Origin 2021 software.

2.4. Observation Indicators

The results of visual analysis were displayed in the form of a knowledge map, with a clustering coefficient (Q) of 0.5755, which is greater than 0.3, indicating significant clustering results, and an average silhouette coefficient (S) of 0.7823, which is greater than 0.5, indicating reasonable clustering.

3. Results

3.1. Publication Volume Trend Analysis

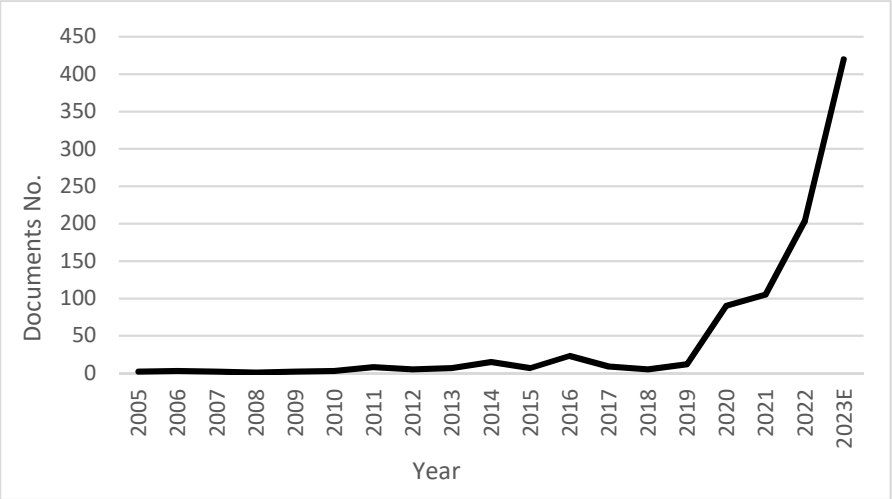

Fig.1 2005-2023 Documents numbers published on color-based image retrieval research

The publication status of literature is usually considered a key indicator for measuring the development level and research achievements of a discipline, and annual publication volume can also reflect the future development trend to some extent in related fields of color-based image retrieval. After statistically analyzing the annual publication volume of literature in this field, it was found that the overall level of publication volume has increased significantly from 2005 to 2023, mainly divided into three stages: from 2005 to 2013 was a slow germination period, with an average annual publication volume of only 3.6 literature. From 2014 to 2019, it was a fluctuating growth period, with an obvious increase in publication volume compared to the previous stage, reaching a peak of 23 literature in 2016, and an average annual publication volume of 11.8 literature. From 2020 to 2023, it was an explosive growth period, where AI technology and computer algorithms led to an explosive growth in related research on color-based image retrieval technology during this period. The cumulative publication volume in this stage reached 503 literature, and the average annual publication volume reached 125.8 literature, which was 34.94 times and 10.66 times the first and second stages, respectively. About 83% of articles were published in the last three years, and from January to March 2023, 105 articles had already been published. According to a year-on-year comparison, it is

ISSN: 2710-0170

estimated that 420 literature will be published in 2023, showing an explosive growth trend. This indicates that color-based image retrieval technology has a high research interest and excellent research prospects now.

3.2. High Citations Frequency Journal Analysis

Through analyzing the publications included in the literature sample, we can better understand the influential and in-depth journals in the field, and focus on current research directions. The publication journals of the 716 literature samples were analyzed, as shown in Table 1, which displays the top 10 journals ranked by h-index for published papers among all literature samples. The h-index is a mixed quantitative indicator, defined as follows: if a person has N papers in his/her academic literature and each of them is cited at least N times, then his or her h-index is N. The h-index is commonly used to evaluate the quantity and quality of academic output of researchers[13].

In the field of color-based image retrieval, the top three journals ranked by h-index are MULTIMEDIA TOOLS AND APPLICATIONS, MULTIMEDIA TOOLS AND APPLICATIONS, and INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS, respectively. MULTIMEDIA TOOLS AND APPLICATIONS was ranked first, with an h-index of 9, which means that at least 9 papers from this journal have been cited at least 9 times, and its impact factor of 2.577 is higher than that of other journals, indicating that this journal is playing a leading role in the field and its academic level is advanced. Moreover, the total citation frequencies of the top three journals ranked by h-index have all exceeded 100 times, indicating that the research output quality is high.

Table 1. Frequency analysis of highly cited papers.

No.	Journal Title	h_index	total citations
1	MULTIMEDIA TOOLS AND APPLICATIONS	9	215
2	IEEE ACCESS	7	137
3	INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS	6	103
4	INTERNATIONAL JOURNAL OF COMPUTATIONAL SCIENCE AND ENGINEERING	4	38
5	REMOTE SENSING	4	37
6	3D RESEARCH	3	56
7	COMPUTER JOURNAL	3	18
8	COMPUTERS AND ELECTRONICS IN AGRICULTURE	3	53
9	IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY	3	88
10	IET IMAGE PROCESSING	3	12

3.3. Keywords Co-occurrence Analysis

Keywords are highly summarized representations of the main themes of literature, and co-occurrence analysis is a visualization of knowledge maps formed by cutting the nodes based on keywords. From the changes in hot fields and research methods in different time periods, this analysis can reveal the intrinsic connections between research topics in a given field. Among the 716 articles related to color-based image retrieval, there are 183 keywords, of which 10 appear more than 20 times. The high-frequency keywords related to color-based image retrieval include "color," "retrieval," "classification," "feature," "algorithm," "feature extraction," etc., as shown in Table 5. Based on the results analysis, the main paths of image retrieval based on color are retrieval, classification, feature extraction, image color analysis, and

ISSN: 2710-0170

recognition. Research methods that can be integrated include deep learning and machine learning. Color, feature, and texture are important factors that play a role in the image retrieval process, while algorithm and model are key tools used in image retrieval technology[14].

Table 2. Top 20 keywords of color-based image retrieval research

No.	Keywords	Occurrence frequency
1	color	54
2	retrieval	49
3	classification	31
4	feature	27
5	algorithm	25
6	feature extraction	24
7	deep learning Xiamen Univ	23
8	recognition	21
9	model	21
10	image color analysis	20
11	image retrieval	17
12	convolutional neural network	17
13	machine learning	17
14	atmospheric correction	16
15	content-based image retrieval	15
16	neural network	15
17	seawif	14
18	image	14
19	coastal	13
20	system	12

The co-occurrence analysis of keywords in the field of color-based image retrieval is visualized in Fig.5. The results show that the keyword co-occurrence network has obtained 183 keywords and 532 connections. Color, as a more refined and restored derivative in content-based image retrieval, revolves around the center of color and texture. Texture is usually automatically extracted and stored in the data graph as one of the basic image features, along with color, shape, and positioning. In the process of color-based image retrieval, texture elements are extracted from the image to express its texture information, and then combined with color for image presentation[15]. Meanwhile, the new form of image retrieval technology revolves around key steps such as algorithm, analysis, extraction, and classification. The algorithms used in the technical process range from the basic MATLAB algorithm used in content-based image retrieval to the histogram algorithm which has defects in local color recognition, and then to more advanced retrieval methods such as the retrieval algorithm based on color features which can fully utilize chromaticity in the HSV space, as well as more refined scanning line projection methods[16]. As the technology in this field becomes increasingly advanced, the corresponding research is also becoming more popular.

The co-occurrence results also show the integration of new technologies from multiple fields, such as deep learning and support vector machines. By using support vector machines (SVM) to classify data in a supervised learning manner, SVM algorithms can achieve more accurate

ISSN: 2710-0170

results when processing small regions, even though they are still facing issues such as slow machine algorithm speed, complex algorithm implementation paths, and high computational requirements[17]. Through DL, combined with the feature data and deep belief networks (DBN) of images, color image recognition can be improved, thereby enhancing the recognition and processing capabilities of computer languages and improving the accuracy of image recognition in response to the changing nature of images[18].

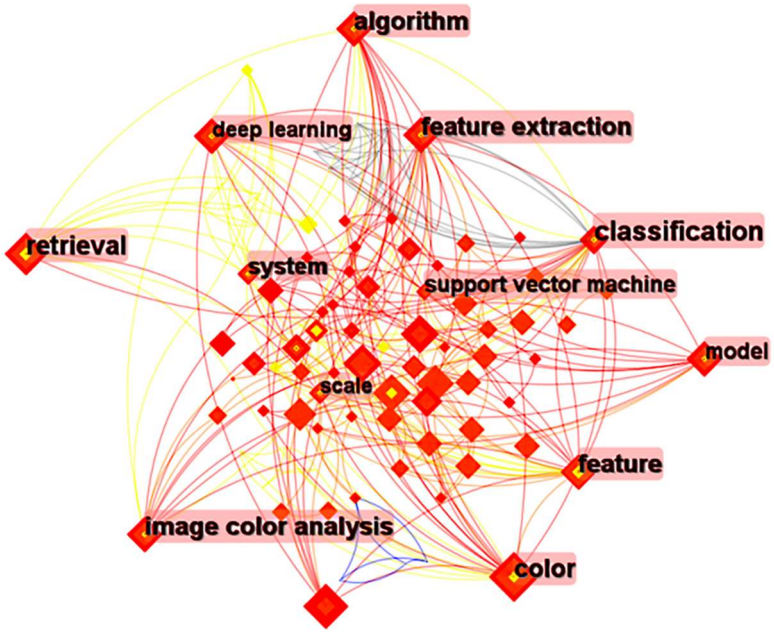

Fig.2 Keyword co-occurrence collaborative network display of color-based image retrieval research literature

4. Conclusion and Discussion

This study analyzed 716 literature related to color-based image retrieval published from January 2005 to March 2023 using CiteSpace 6.1.R2 software, aimed to clarify the current research hotspots and future trends in this field of image retrieval and provide direction for better research in the future.

The results showed that: (1) The number of literature in this field showed a continuous increase, especially in the past three years. The academic level of the journals in this field is relatively high. (2) The number of authors publishing literature presents a small concentration and large dispersion structure, the volume of literature is relatively average, and some authors have formed close cooperation teams. (3) The research institutions are mainly research institutes and comprehensive universities with strong strength, and the relationship between institutions is close, and their cooperation has a certain regional nature. (4) The current research literature mainly focuses on the main paths and technical tools for color-based image retrieval, key objects and possible influencing factors in the technical process, and cross-disciplinary extension applications of the technology. (5) Through trend forecasting and analysis of related research in this field, it is found that more accurate color description and extraction technologies such as the use of mixture of noval angle patterns and colors, dominant color descriptor, and specific practices that integrate color and image retrieval into environmental monitoring are the future research hotspots and trends.

ISSN: 2710-0170

References

[1] Ammatmanee, C. , & Lu, G. . (2021). A ten-year literature review of content-based image retrieval (cbir) studies in the tourism industry. The Electronic Library.

[2] Manjula, K. , Monisha, A. , Reshma, K. , Swetha, P. , & Vijayarekha, K. . (2016). Content based image retrieval systems: a review. Research Journal of Pharmaceutical, Biological and Chemical Sciences, 7(6), 1915-1921.

[3] Chan, Y. K. , & Liu, Y. T. . (2008). An image retrieval system based on the image feature of color differences on edges in spiral scan order. International Journal of Pattern Recognition & Artificial Intelligence, 17(08), 1417-1429.

[4] Wang, X. Y. , Chen, Z. F. , & Yun, J. J. . (2011). A novel two-level color image retrieval method. International Journal of Image and Graphics, 11(03), 1100418-.

[5] Park, Y. , Park, K . , & Kim, G. . (2013). Content-based image retrieval using colour and shape features. International Journal of Computer Applications in Technology, 48(2), 155-161.

[6] Liu, H., Wang, Y., Chen, D., Lv, J., & Alshalabi, R. (2022). Garment image retrieval based on grab cut auto segmentation and dominate color method. Applied Mathematics and Nonlinear Sciences, 6, 1–11.

[7] Konstantinidis, K. , & Andreadis, I. . (2005). Performance and computational burden of histogram based color image retrieval techniques. Journal of Computational Methods in Sciences & Engineering, 5(2), p.141-147.

[8] Hiremath, P. S. , Shivashankar, S. , & Pujari, J. . (2006). Wavelet based features for color texture classification with application to cbir. International Journal of Computer ence & Network Security(9A), 124-133.

[9] HK Bhuravarjula, & VNSV Kumar. (2012). A novel content based image retrieval using variance color moment. International Journal of Computer & Electronics Research, 1(3).

[10] Verma, P. , & Mahajan, M. . (2022). Content based image retrieval by comparing color and shape features.

[11] Chen, Y., Chen, C. M., Liu, Z. Y., Hu, Z. G., & Wang, X. W. (2015). Methodological features of CiteSpace knowledge graph. Scientific Research, 33(2), 242–253.

[12] Chen, C. M. (2014). Science frontier atlas: An exploration of knowledge visualization. Science Press.

[13] Bar-Ilan, J. . (2008). Which h-index? - a comparison of wos, scopus and google scholar. Scientometrics, 74(2), 257-271.

[14] Sun, X., Xue, Y., Xu, Q., Shen, J., & Xu, Z. (2022). Research on colored yarns for a full color phase mixing model and the Stearns–Noechel color prediction approach. Textile Research Journal, 93(1-2), 206–220.

[15] Guo, E., Wei, Y., Zhu, S., Bai, L., & Han, J. (2023). Imaging of color targets through scattering media based on mixed speckle pattern separation. Optics and Lasers in Engineering, 161, Article 107324.

[16] Pahlevan, N. , Smith, B. , Binding, C. , Gurlin, D. , & Giardino, C. . (2020). Hyperspectral retrievals of phytoplankton absorption and chlorophyll-a in inland and nearshore coastal waters. Remote Sensing of Environment.

[17] Ahmed, K. T. , Jaffar, S. , Hussain, M. G. , Fareed, S. , & Choi, G. S. . (2021). Maximum response deep learning using markov, retinal & primitive patch binding with googlenet & vgg-19 for large image retrieval. IEEE Access, PP(99), 1-1.

[18] Mehta, T. , & Bhensdadia, C. K. . (2020). Near duplicate image retrieval using multilevel local and global convolutional neural network features. International Journal of Advanced Computer Science and Applications(4).

[19] Bai, C. , Li, H. , Zhang, J. , Huang, L. , & Zhang, L. . (2021). Unsupervised adversarial instance-level image retrieval. IEEE Transactions on Multimedia, PP(99), 1-1.

[20] Wang, X. Y. , Chen, Z. F. , & Yun, J. J. . (2011). A novel two-level color image retrieval method. International Journal of Image and Graphics, 11(03), 1100418-.

ISSN: 2710-0170

[21] Alnihoud, J. Q. . (2018). Image retrieval system based on color global and local features combined with glcm for texture features. International Journal of Advanced Computer Science and Applications(9).

[22] Abter, S. O. , & Abdullah, N. A. Z. . (2017). Colored image retrieval based on most used colors. International Journal of Advanced Computer Science and Applications, 8(9), 2017.

[23] Prasad, R. M. . (2019). Color and Edge Oriented Histogram for Real-Time Costume Image Retrieval.

[24] Bhardwaj, S. , Pandove, G. , & Dahiya, P. K. . (2021). An analysis of two novel and efficient deep learning models for fast and accurate image retrieval. Journal of information science and engineering: JISE(1), 37.

[25] Nazeer, M. , & Nichol, J. E. . (2016). Improved water quality retrieval by identifying optically unique water classes. Journal of Hydrology, 541, 1119-1132.

[26] Xu, G. . (2020). Multiple geometry atmospheric correction for image spectroscopy using deep learning. Journal of Applied Remote Sensing, 14(2).

[27] Jeyanthi, P. , Rubini, K. , & Vinitha, S. . (2016). Medical image retrieval using content based image retrieval. Research Journal of Pharmaceutical, Biological and Chemical Sciences, 7(4), 3014-3021.

[28] Dewan, J. H., & Thepade, S. D. (2021). Feature fusion approach for image retrieval with ordered color means based description of keypoints extracted using local detectors. Journal of Engineering Science and Technology, 16(1), 482–509.

[29] Singh, A. , Pannu, H. , & Malhi, A. . (2021). Explainable information retrieval using deep learning for medical images.

[30] Divya, M. O. , & Vimina, E. R. . (2020). Content based image retrieval with multi-channel LBP and colour features. International Journal of Applied Pattern Recognition, 6(2).

[31] Tavora, J. , Boss, E. , Doxaran, D. , & Hill, P. . (2020). An algorithm to estimate suspended particulate matter concentrations and associated uncertainties from remote sensing reflectance in coastal environments. Remote Sensing, 12(13), 2172.

[32] Hamroun, M. , Tamine, K. , Claux, F. , & Zribi, M. . (2021). A new content-based image retrieval system using deep visual features. International Journal of Image and Graphics, 2150055.