

Content Based Image and Video Retrieval: A Compressive Review

Sheetal Deepak Patil



Abstract: *Content-based image retrieval is quickly becoming the most common method of searching vast databases for images, giving researchers a lot of room to develop new techniques and systems. Likewise, another common application in the field of computer vision is content-based visual information retrieval. For image and video retrieval, text-based search and Web-based image reranking have been the most common methods. Though Content Based Video Systems have improved in accuracy over time, they still fall short in interactive search. The use of these approaches has exposed shortcomings such as noisy data and inaccuracy, which often result in the showing of irrelevant images or videos. The authors of the proposed study integrate image and visual data to improve the precision of the retrieved results for both photographs and videos. In response to a user's query, this study investigates alternative ways for fetching high-quality photos and related videos.*

Keywords: *Image and Video Retrieval, Content-Based, Similar Match Etc.*

I. INTRODUCTION

Recently, visual search has gained a lot of traction. This technology, in its most basic form, provides for text-based query search within an image database. This methodology is quickly utilized for product recognition [1], [2], and location recognition [3], [4], as well as commercial applications [5], [6]. A methodology for searching user-defined query photos is text-based picture categorization. The absence of correlation between the genuine significance of a picture and the assumed relevance of the related textual data limits the accuracy of the search result. The search results obtained from the text-based image search can be further refined using a visual re-ranking process, improving the accuracy of the text-based image search ranking. Popular search engines such as Google, Yahoo, and Bing are now built on the concept of retrieving photos in response to user requests utilizing annotated labels added to photos. Image retrieval has piqued the interest of academics and industry alike, because to the expanding library of online images [7], [8]. As a result, a new visual categorization [9], [10] has been proposed to optimize textual search results by modifying visual information included in images.

For example, for brand publicity, a firm or company might want to find all types of symbols or product in various sectors. A better solution to this issue is to store image independently, considering the database image as database of text, where the text could match the image. Our system, on the other hand, offers a new architecture in which the text query is immediately database compared to photos and video clips in the first stage, which improves the accuracy and scalability of our system. The main goal is to find the most relevant photographs and video clips from a big database by employing a query text for photos and a content-based approach for videos. For example, the color qualities will be appropriate for difficulties involving color dispersion, such as sunsets, sunrises, and beaches. Edge and texture features will be more efficient for some searches, such as "build" and "street." As a result, it's easy to see how semantic attributes could be viewed as descriptors or an image data mode. The most important properties of the various inquiries may be retrieved via multimodal function analysis. These encourage the use of semantic features to re-rank web image searches more effectively. The visual features and attribute characteristics are extracted from the photographs obtained. The output of different binary classifiers in all of the attributes makes up the characteristic attribute of each image. In a unified model termed hypergraph [11], the visual representation and the semantic explanation are immediately abused.

Text-based image search and Content-based video search presents two main challenges:

- 1] Database image and videos includes temporal component; while query text does not. How can a retrieval system take into consideration to design effective method of comparing the query as a text to images and videos?
- 2] How can we integrate data based on features that can be compared directly to text, image, and video content?

II. PRESENT THEORY AND PRACTICES

A. Literature Survey

Manuscript received on June 06, 2021.

Revised Manuscript received on June 15, 2021.

Manuscript published on June 30, 2021.

* Correspondence Author

Sheetal Deepak Patil*, ME, Government College of Engineering, Aurangabad (Maharashtra), India. E-mail: psheetal065@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

TABLE -1: Finding from Existing Work

Author	Year	Proposed Technique	Input Data	Parameters Employed for Evaluation	Key Findings
Huang-Chia Shih	2007	Proposal of video attention ranking using the Visual and Contextual Attention model for the extraction of content-based sports videos.	Commercial baseball game sequences.	1)Visual Attention Model (VAM) 2)Contextual Attention Model (CAM) 3) Attention Rank (AR)	The proposed method produces promising results.
Ja-Hwung Su	2009	Proposed an innovative method to achieve video recovery based on effective content.	Collection of 15 video categories in real life.	1)DFPI-tree 2)SFPI-tree 3)BSW 4)ASW	The proposed method produces excellent performance for content-based video retrieval in terms of efficiency and effectiveness.
S.Padmakala	2011	Developed an effective video retrieval method based on content.	50 videos obtained from open direct video project	1) Video Segment Representation (VSR) 2)Optimal Frame Representation (OFR)	In terms of video retrieval, the proposed system delivered a promising and successful result.
B. V. Patel	2012	Low-level visual features derived from video frames are used in video analysis.	E-learning, sports, cartoon, movies.	1) Gray Level Co-occurrence Matrix I (GLCM) 2)Simple Color Histogram 3)Simple Region Growing 4)Auto Color Correlogram 5) Naive Vector	The results of experiments suggest that combining derived information improves video indexing and retrieval.
Thanh DUC NGO	2014	A method for retrieving videos based on their content has been proposed.	8 Video sequences from different video broadcasting stations.	1)Support Vector Machine 2)MA 3)GSC	The approaches provided are both reliable and efficient.
Madhav Gitte	2011	The Content Based Video Retrieval (CBVR) System is shown.	Video sequences	1)Euclidian Distance Algorithm 2)RGB, Local Color Histogram (LCH) and Global Color Histogram (GCH) 3) SVM	A method for improving the system's performance has been proposed.
N. Sudha Bhuvaneswari	2014	Extracting and storing video clips in preparation for a video query based on content.	Traffic related videos.	1)Sift Invariant Feature Transform (SIFT) 2)RGB	The proposed strategy boosts accuracy and efficiency.
D.Saravanan	2015	Using histogram clustering, a new method for retrieving video data quickly has been developed.	Video files	1)Histogram clustering 2)CHEMELEON graph	The method proposed is effective.
Ms.Deepti Bhatia	2015	Discrete Cosine Transform for Classification is used for video classification.	Sports and Cartoon Videos.	1)Discrete Cosine Transform 2) Naive Bayesian 3) K-Nearest Neighbor	The results of the experiments reveal that video classification is accurate.

Mr.Siddhant Kulkarni	2015	Distributed Multithreaded Video Classification using a Novel Model.	Video files	1)Simple Standalone Video Classification Model (SSVCM) 2)Multithreaded Video Classification Model (MVCM) 3)Distributed Multithreaded Video Classification Model (DMVCM)	The method proposed is effective.
----------------------	------	---	-------------	---	-----------------------------------

B. Background.

Methods	Cons	Pros
Classification Based Method	<ol style="list-style-type: none"> Active in the recovery of images. Use the binary classification method to identify whether the image is relevant or not. The classification model is learned with Pseudo Relevance Feedback 	In many real-world circumstances, the PRF's training examples are noisy and may not be sufficient for generating effective classifiers.
Clustering Based Method	<ol style="list-style-type: none"> It operates by sharing appropriate photographs with the crucial observation of visual qualities. Searching of result is based on the original text which are grouped by visual proximity 	There is no performance guarantee when the visual patterns are not clear which cause effect on results obtained.
Graph Based Method	<ol style="list-style-type: none"> It has increased attention and prove to be efficient Use random graph for similarity and rearrange the images accordingly Sorted list is obtained based on graph node weights 	It is true to say that it is based on low-level traits rather than the semantic connections between ranked lists.

III. COLOR FEATURES

Because color allows the human eye to distinguish between different sights depending on color, it is one of the most important low-level visual properties. Within the human visual range, color changes can be used to distinguish representations of real-world objects [21–24]. A modest quantity of representing colors using dominant color descriptor (DCD) [21] can replace the image's entire color knowledge. DCD is an MPEG-7 color descriptor that describes the suggestive color distribution and function in a simple, lightweight, and intuitive manner. The complexity of similarity computing is lowered as a result of choosing eight dominating colors from each image, calculating features using the histogram intersection approach, and reducing the complexity of similarity computing.

According to Duanmu [22], traditional techniques can retrieve images based on their labels and annotations, but this does not meet the needs of customers; as a result, the researchers concentrated on a different method of retrieving images: retrieving images based on their material. A two-stage clustering technique is used in the proposed method to modify a small image descriptor according to the meaning of the image.

Wang et al.[23] proposed using color to retrieve photos depending on their quality, which is derived by integrating color and texture features. This enables a more precise and flexible examination of how early humans process visual information [23 The combination of color and texture data provides a comprehensive feature set for color picture

retrieval techniques. According to the findings of the experiments, the proposed method captured photos more precisely than other standard ways. However, unlike other methods, the function dimensions are not as large, and the computational cost is considerable. The similarity calculation is calculated using a pairwise comparison for both low-level features, which may be a bottleneck [23].

An analysis of the completeness property of invariant descriptors was conducted by several research groups [24]. Zernike and pseudo-Zernike polynomials, which are orthogonal basis moment functions that maintain orthogonality and rotation invariance [24], can be used to represent the picture as a collection of mutually independent descriptors. Picture noise resistance was found to be higher in PZMs than in Zernike moments. First, a relationship is established between the pseudo-Zernike moments of the original image and images of the same shape but different orientation and scale. An absolute collection of scale and rotation invariants emerges from this relationship. In addition, this proposed strategy outperformed existing algorithms in pattern recognition [24].

Guo et al.[25] proposed a new indexing approach based on error diffusion block truncation coding features (EDBTC). To build image feature descriptors, EDTBC uses vector quantization (VQ) to process two color quantizers and a bitmap picture.



Color Histogram Function (CHF) and Bit Pattern Histogram Feature have been introduced to determine the similarity between the query image and the picture in the database (BHF). A VQ-indexed color quantizer and a VQ-indexed bitmap picture are used to calculate the CHF and BHF, respectively. The distance between CHF and BHF can be used to compare the two photos' similarity.

According to the findings of the studies, the suggested technique outperforms earlier BTC-based picture indexing and other contemporary picture retrieval systems. Image compression and CBIR image indexing are both possible with the EDBTC [25].

Jiexian et al. [26] suggested a multiscale distance coherence vector (MDCV) for CBIR. Because multiple shapes may have the same description, the distance coherence vector approach may not be able to completely remove noise. The proposed methodology starts by using the Gaussian function to construct the picture contour curve. Different operations such as translation, rotation, and scaling transformation are invariant to the proposed technique.

IV. TEXTURE FEATURES

Papakostas et al.[27] conducted studies on four datasets to illustrate the discrimination capacity of wavelet moments: COIL, ORL, JAFFE, and TRIESCH I. The suggested model (WMs) is evaluated using two different wavelet configurations, WMS-1 and WMs-2, with the former employing cubic B-spline and the latter employing Mexican hat mother wavelets. By using a feature selection strategy that only keeps successful traits, the classification skills of wavelet moments are considerably improved. Using 25, 50, 75, and 100% of the full datasets, the proposed model is compared to Zernike, pseudo-Zernike, Fourier–Mellin, and Legendre, as well as two others, and each moment family behaves differently in each dataset. The proposed model achieves better results in the classification of moment descriptors (wavelet Moments and moment invariants).

Because there are no known datasets for content-based image retrieval, Liu et al. [28] use Corel datasets to evaluate the proposed model (MSD) for image retrieval (CBIR). With a total of 15000 files, Corel-5000 and Corel-10000 are used to test retrieval output in HSV, RGB, and Lab color spaces. On the Corel-5000 and Corel-10000 datasets, the average retrieval and recall rates of the proposed model employing different color quantization levels and texture orientation quantization levels are tested, and our suggested model performs well in HSV and Lab color spaces but poorly in RGB color space. MSD uses 72 color and orientation quantization levels, as well as 6 for picture retrieval, to produce good results in terms of storage space, retrieval accuracy, and speed. The average retrieval and recall ratios of MSD are compared to other methods like as Gabor MTH on Corel datasets for evaluation, as these algorithms were built for image retrieval. The results show that our proposed model (MSD) outperforms other models.

V. SHAPE FEATURES

Shape is also an important low-level function because it helps with the recognition of real-world items and shapes. The application of shape characteristics in the domains of image

retrieval and image representation was thoroughly examined by Zhang and Lu [29].

Region-based and contour-based form features are the two basic classifications of shape features [30]. Trademark-based image retrieval [31] is one of the most fundamental fields where form features are applied for image representation.

VI. CONCLUSION

We conducted a thorough assessment of the literature on various CBIR and image representation techniques. The primary purpose of this study is to present an overview of several strategies that have been applied in different research models over the previous 12–15 years. Following this analysis, it can be concluded that image features are depicted using low-level visual features such as color, texture, spatial structure, and form. Due to the diversity and non-homogeneity of image attributes, a single feature representation cannot be utilized to represent image datasets. Image features are presented utilizing low-level visual elements such as color, texture, spatial structure, and form, according to this analysis. A single feature representation cannot be used to represent image collections due to the diversity and non-homogeneity of picture properties. In past CBIR and image representation investigations, traditional machine learning algorithms produced excellent results in a variety of domains. The optimization of feature representation in terms of feature dimensions can provide a good foundation for developing classification-based models with minimal overfitting. Recent CBIR research has moved to the use of deep neural networks, which have produced promising results on a variety of datasets and outperformed handcrafted features when fine-tuning the network. The key criteria for any deep network are large-scale image datasets and powerful computing machines. It's difficult and time-consuming to manage a large-scale image dataset for supervised deep network training. As a result, one of the potential future research topics in this subject is to assess the effectiveness of a deep network in unsupervised learning mode on a large-scale unlabeled dataset.

REFERENCES

1. S. Tsai, D. Chen, V. Chandrasekhar, G. Takacs, N.-M. Cheung, R. Vedantham, R. Grzeszczuk, and B. Girod, "Mobile Product Recognition," in Proc. ACM MM, 2010. Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (eds.). Machine Learning: An Artificial Intelligence Approach, Vol. I. Tioga, Palo Alto, CA, 1983
2. J. He, J. Feng, X. Liu, T. Cheng, T.-H. Lin, H. Chung, and S.-F. Chang, "Mobile Product Search with Bag of Hash Bits and Boundary Reranking," in Proc. CVPR, 2012. Michalski, R. S., Carbonell, J. G., and Mitchell, T. M. (eds.). Machine Learning: An Artificial Intelligence Approach, Vol. I. Tioga, Palo Alto, CA, 1983
3. D. Chen, G. Baatz, K. Koser, S. Tsai, R. Vedantham, T. Pylyk, K. Roimela, X. Chen, J. Bach, M. Pollefeys, B. Girod, and R. Grzeszczuk, "City-Scale Landmark Identification on Mobile Devices," in Proc. CVPR, 2011
4. Schroth, R. Huitl, D. Chen, M. Abu-Alqumsan, A. Al-Nuaimi, and E. Steinbach, "Mobile Visual Location Recognition," IEEE Signal Processing Magazine, vol. 28, no. 4, pp. 77–89, 2011
5. Amazon Flow, <http://flow.a9.com>. Accessed Apr. 2016

6. N. Kumar, A. C. Berg, P. N. Belhumeur and S. K. Nayar. Attribute and simile classifiers for face verification. In Proceedings of the IEEE International Conference on Computer Vision, 2009
7. M. Wang, G. Li, Z. Lu, Y. Gao, T.-S. Chua. When amazon meets google: product visualization by exploring multiple web sources. In ACM Transaction on Internet Technology, vol. 12, no. 4, 2014
8. M. Wang, K. Yang, X.-S. Hua, H.-J. Zhang. Towards a relevant and diverse search of social images. In IEEE Transaction on Multimedia, vol. 12, no. 8, pp. 829 - 842, 2010
9. Y. Liu and T. Mei. Optimizing visual search reranking via pairwise learning. IEEE Transactions on Multimedia, vol. 13, no. 2, pp. 280-291, 2011
10. F. Jing, C. Wang, Y. Yao, K. Deng, L. Zhang, and W.-Y. Ma. Group: Web image search results clustering. In Proceedings of the ACM Conference on Multimedia, 2006
11. A. Kovashka, D. Parikh, and K. Grauman. WhittleSearch: Image search with relative attribute feedback. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012
12. Huang-Chia Shih, Chung-Lin Huang, Jenq-Neng Hwang, "Video Attention Reranking Using Visual and Contextual Attention Model for Content-Based Sports Videos Mining," 1-4244-1274-9/07 ©2007 IEEE
13. Ja-Hwung Su, Yu-Ting Huang, Vincent S. Tseng, "Efficient Content-Based Video Retrieval by Mining Temporal Patterns," MDM/KDD'08, August 24, 2009, Las Vegas, NV, USA
14. S. Padmakala, Dr. G. S. Anandha Mala, M. Shalini, "An Effective Content Based Video Retrieval Utilizing Texture, Color and Optimal Key Frame Features," 978-1-61284-861-7/11 ©2011 IEEE
15. B. V. Patel, B. B. Meshram, "Content Based Video Retrieval," The International Journal of Multimedia & Its Applications (IJMA) Vol.4, No.5, October 2012
16. Thanh DUC NGO, DUY DINH, Shin'ichi SATOH "Scalable Approaches for Content Based Video Retrieval," DOI:10.2201/Niipi.2014.11.5
17. Madhav Gite, Harshal Bawaskar, Sourabh Sethi, Ajinkya Shinde "Content Based Video Retrieval System," International Journal of Research in Engineering and Technology, Volume: 03 Issue: 06 | Jun-2014
18. N. Sudha Bhuvaneshwari Ph.D, M. Madhanika, "Content Based Video Querying Technique For Video Retrieval and Video Making from Large Video Complication," International Journal of Computer Applications (0975 - 8887) Volume 106 - No.6, November 2014
19. D. Saravanan, Vaithyasubramanian, K. N. Jothi Vengatesh, Video Content Retrieval Using Histogram Clustering Technique," Procedia Computer Science 50 (2015) 560 - 565
20. Ms. Deepti Bhatia, Ms. Sonali Sabale, Mr. Siddhant Kulkarni, Ms. Sandhya Shinde "A Thorough Evaluation of Discrete Cosine Transform for Content Based Video Classification," 978-1-4799-6892-3/15 © 2015 IEEE
21. H. Shao, Y. Wu, W. Cui, and J. Zhang, "Image retrieval based on MPEG-7 dominant color descriptor," in Proceedings of the 9th International Conference for Young Computer Scientists ICYCS 2008, pp. 753-757, IEEE, Hunan, China, November 2008
22. X. Duanmu, "Image retrieval using color moment invariant," in Proceedings of the 2010 Seventh International Conference on Information Technology: New Generations (ITNG), pp. 200-203, IEEE, Las Vegas, NV, USA, April 2010
23. X.-Y. Wang, B.-B. Zhang, and H.-Y. Yang, "Content-based image retrieval by integrating color and texture features," Multimedia Tools and Applications, vol. 68, no. 3, pp. 545-569, 2014
24. H. Zhang, Z. Dong, and H. Shu, "Object recognition by a complete set of pseudo-Zernike moment invariants," in Proceedings of the 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 930-933, IEEE, Dallas, TX, USA, March 2010
25. M. Guo, H. Prasetyo, and J. H. Chen, "Content-based image retrieval using error diffusion block truncation coding features," IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 3, pp. 466-481, 2015
26. Jiexian, L. Xiupeng, and F. Yu, "Multiscale distance coherence vector algorithm for content-based image retrieval," The Scientific World Journal, vol. 2014, Article ID 615973, 13 pages, 2014
27. G. Papakostas, D. Koulouriotis, and V. Tourassis, "Feature extraction based on wavelet moments and moment invariants in machine vision systems," in Human-Centric Machine Vision, InTech, London, UK, 2012
28. G.-H. Liu, Z.-Y. Li, L. Zhang, and Y. Xu, "Image retrieval based on micro-structure descriptor," Pattern Recognition, vol. 44, no. 9, pp. 2123-2133, 2011
29. D. Zhang and G. Lu, "Review of shape representation and description techniques," Pattern Recognition, vol. 37, no. 1, pp. 1-19, 2004
30. D. Ping Tian, "A review on image feature extraction and representation techniques," International Journal of Multimedia and Ubiquitous Engineering, vol. 8, no. 4, pp. 385-396, 2013
31. Z. Hong and Q. Jiang, "Hybrid content-based trademark retrieval using region and contour features," in Proceedings of the 22nd International Conference on Advanced Information Networking and Applications-Workshops AINAW 2008, pp. 1163-1168, IEEE, Okinawa, Japan, March 2008



AUTHORS PROFILE

Sheetal Deepak Patil, I have completed ME from Govt Engineering College Aurangabad, Maharashtra, India. My Research Area in Image Processing and Vision Computing. I am a research fellow and currently working in IT industry.

