UNIVERSITY "ST. KLIMENT OHRIDSKI" - BITOLA
FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGIES - BITOLA
REPUBLIC OF NORTH MACEDONIA

Proceedings of the 13th International Conference on Applied Internet and Information Technologies

AIIT 2023

13 October, 2023, Bitola, Republic of North Macedonia

AIIT
International Conference

UNIVERSITY "ST. KLIMENT OHRIDSKI" - BITOLA
FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGIES - BITOLA
REPUBLIC OF NORTH MACEDONIA

13TH INTERNATIONAL CONFERENCE ON APPLIED INTERNET AND INFORMATION TECHNOLOGIES

AIIT 2023
PROCEEDINGS

Bitola, 2023

Proceedings publisher and organizer of the conference:
University “St. Kliment Ohridski” - Bitola, Faculty of Information and Communication Technologies - Bitola, Republic of North Macedonia

For publisher:
Blagoj Ristevski, PhD
Faculty of Information and Communication Technologies - Bitola
University “St. Kliment Ohridski” - Bitola, North Macedonia

Proceedings editors:
Kostandina Veljanovska, PhD
Eleonora Brtka, PhD

Conference Chairmans:
Kostandina Veljanovska, University “St. Kliment Ohridski”, Faculty of Information and Communication Technologies - Bitola, Republic of North Macedonia - chair
Eleonora Brtka, University of Novi Sad, Technical Faculty “Mihajlo Pupin”, Zrenjanin, Serbia – co-chair

Technical preparation of the proceedings:

Andrijana Bocevska, PhD

Kostandina Veljanovska, PhD

Cover design:

Andrijana Bocevska, PhD

e-Proceedings
ISBN 978-608-5003-03-7

CIP - Каталогизација во публикација
Национална и универзитетска библиотека "Св. Климент Охридски", Скопје

004-049.8(062)

INTERNATIONAL conference on applied internet and information technologies AIIT 2023 (13 ; 2023 ; Bitola, Republic of North Macedonia)
Proceedings / 13th International conference on applied internet and information technologies AIIT 2023, 13 October, 2023, Bitola, Republic of North Macedonia ; [editors Kostandina Veljanovska, Eleonora Brtka]. - Bitola : University "St. Kliment Ohridski", Bitola Faculty of information and communication technologies, 2023. - 411 стр. : илустр. ; 30 см
Библиографија кон трудовите
ISBN 978-608-5003-03-7

а) Информатичка технологија -- Примена -- Собири
COBISS.MK-ID 62559493

C O N T E N T

Content-Based Image Retrieval: Contemporary Trends and Challenges

Buen Bajrami[1], Zoran Kotevski[2] and Kostandina Veljanovska[3]

[1,2,3] *"St. Kliment Ohridski" University - Bitola, Faculty of Information and Communication Technologies - Bitola, Studentska bb., 7000 Bitola, North Macedonia*

[1]*bajrami.buen@uklo.edu.mk;* [2]*zoran.kotevski@uklo.edu.mk;* [3]*kostandina.veljanovska@uklo.edu.mk*

Abstract:
Content-Based Image Retrieval (CBIR) is a process that enables finding similar images in large sets of databases based on an image query. The purpose of CBIR engines is to mimic humans in the image classification process, thus a high computational cost is required to reach the proper selection of features, which must be as unique as possible. In the last two decades, the research related to CBIR has increased, but even today it is considered as a difficult process. For this reason, the researchers in this field have developed various models and techniques that help in rendering the image, trying to make the results as accurate as possible. Some of these techniques include Local Binary Pattern (LBP) histogram, Local Difference Binary (LDB), Local Tetra Pattern (LTrP), etc. In this paper, we review the latest research in the field of CBIR, we compare their performance based on several factors, such as calculation time, image acquisition time and accuracy of results and we conclude the paper with a discussion about which models have shown the best performance and what are their advantages and disadvantages for which we make some recommendations for the future research.

Keywords:
Image Retrieval, Content-Based Image Retrieval

1. Introduction

When we discuss image retrieval we have two technologies in mind: Text-Based Image Retrieval (TBIR), which enables image retrieval based on the metadata that the images contain, and the newer and more advanced Content-Based Image Retrieval (CBIR), which enables image retrieval through various processes based only on the visual appearance of the image, extracting their details such as color, fractures, shapes, etc. CBIR is a process that enables finding similar images in large sets of databases based on an image query. The most basic need in CBIR is to search and sort images from a given archive through human-machine communication. So the purpose of CBIR engines is to mimic humans in the image classification process. To increase the accuracy rate the requirements of the end user must be as concrete as possible. Also in this aspect, a high computational cost is required to reach the proper selection of features which must be as unique as possible, otherwise, the wrong choice of the model and then of the image features negatively affects the result by providing images that are not so similar to the query image. In CBIR, all high-level visuals are represented by feature vectors consisting of numerical values. Image feature vectors are used as input to Machine Learning (ML) algorithms through training and testing models. In the last two decades, the research related to CBIR has increased rapidly, but even today it is treated as a complex paradigm. In this manner, various models and techniques are developed that help in rendering the image, trying to make the results as accurate as possible. Some of the techniques used for CBIR include Local Binary Pattern (LBP) histogram, Local Difference Binary (LDB), Local Tetra Pattern (LTrP), Local Neighbor Pattern (LNP), Color Volume Histogram (CVH), Deep Learning (DL) and Color Moments (CM). We have also researched several models and frameworks such as Hybrid+SVM+RF, Hybrid+CFBPNN+RF, GMAF, LSH, Color Histogram, GMM, etc. We have analyzed the way they work, and we have compared the results obtained from different experiments based on several factors, such as calculation time, image acquisition time, and accuracy of results. Most of these techniques and models have improved CBIR technology by increasing the level of reliability and accuracy, providing satisfactory results. The

remainder of this paper is organized as follows. Section 2 presents state of the art in CBIR technologies and models. Section 3 concludes the paper with a discussion about the performance of the presented CBIR paradigms.

2. Latest research in CBIR

Leung, Ma and Zhang [1] propose adaptive multimedia indexing using Naïve Bayes classification retrieval of multimedia objects. The core of this search framework is based on capturing human judgment based on user queries to develop semantic indexes related to search terms. Thus, we can say that the lesson is concluded as a success or failure depending on whether the number of positive labels or negative labels, whichever of them reaches the majority. The authors estimate that stochastic methods should be used in multimedia information searches, because if they are missing we will always get the same results. A multi-agent framework is proposed where positive and negative labels are performed by agents and the result will be based on a stochastic method of agents competing with each other. All these tags occur independently. As a future challenge, it remains to do other research related to similar stochastic methods and to add Markov dependence as part of the analysis to increase the accuracy of the process.

Alrahhal and Supreethi [2] presented a CBIR method to find an image from a set of databases. This method covers several areas such as image segmentation, extracting features from the image, and converting these features into semantic features. In this paper, the authors have focused on extracting low-level (color, texture and shape) and high-level features from images. They perform CBIR in two main phases. The first phase enables offline search where the system extracts the feature from all images and stores them in the DB. The second phase is the online phase, where the user enters the query image and the system extracts features from this image and measures the similarity by calculating the distance between the query image (feature vector) and all images in the DB. The authors have analyzed several models that enable the classification of results, four of them have been analyzed in this paper: LBP, LDB, LTrP and LNP. For image classification, the authors use four types of algorithms, including Linear Discriminant, Support Vector Machine, K-Nearest Neighbor classification and Ensemble Classifiers. Also, an experiment was done with three databases, including Color database (Corel 1k), Texture database (Vistex database) and Faces database. The results obtained in the performed experiments show that the most accurate model is LNP compared to other models, in terms of average recall. The development of the LNP model in video retrieval remains an unsolved future work.

Hua et al. [3] developed a feature descriptor that enables conversion from the RGB color space to the HSV color space. The HSV color space closely mimics human color perception and can be interpreted as a cylinder. The authors find it easier to calculate the volume of the cylinder, so the most color areas are highlighted using the volume of the cylinder. Commonly used color spaces include RGB, Lab, LUV, YUV, HSV, YIQ, and YCbCr. Of these, RGB is the most widely used, but it is not a uniform color space. Therefore, the authors have proposed a new visual descriptive method called color volume histogram (CVH) which is based on visual perception. According to the experimental results, the authors claim that this histogram provides greater accuracy than other local histograms or text models. LBP and Multitext histogram (MTH) were used for comparison to validate the performance of this method. The authors also claim that CVH provides a better representation of edges, as well as a better spatial representation of colors compared to LPB, but in the representation of local structural information LPB performs better.

Hou and Wang [4] proposed an improved Gaussian Mixture Model (GMM) and presented a framework of standard image retrieval, where after the system accepts a query image, it enables image retrieval and starts preprocessing. After this stage, the image is divided into blocks. Then the classification according to complexity begins where different methods are used for image extraction and feature vectors are coded by Fisher vectors. In the end, the image acquisition is realized through the similarity index of the feature vectors. The authors have made improvements in several aspects of the Gaussian Mixture algorithm including the feature encoding algorithm, Gaussian mixture model initialization method and new Gaussian distribution generation. After the experiments carried out in PASCAL VOC2012 where a comparison was made between the standard Gaussian algorithm and the

improved one, it is proven that there is a significant improvement in image restoration through the improved algorithm. To compare these two frameworks, the Friedman test was used, which is non-parametric, but takes the performance rank of the algorithms as value. Therefore, we conclude that the framework proposed by the authors in this paper provides greater accuracy, takes less computation time, and increases efficiency.

Saha et al. [5] proposed a new descriptive model of CBIR, which is enabled using DL techniques. Briefly, this process works by collecting the input image and processing it to remove noise. Then, the processed image is extracted and classified using a feed-forward convolutional neural network. According to the experiments carried out in the paper, the technique proposed by the authors, where the processed image is extracted and classified, shows an accuracy performance of 95% and a precision of 79%.

Dhingraa and Bansal [6] proposed two CBIR models. First, Color moment is used, which provides spatial features of an image, and then LPB contains information about scale, noise resistance and brightness variability. To increase the accuracy of this system they use techniques that speed up the process of classifying the similarity between images. Two databases are used in this paper. 1) Corel 1-K: which contains 1000 images divided into 10 categories such as buses, buildings, beaches, food, mountains, etc. and 2) Oxford Flower, which contains 1360 images in JPEG format, divided into 17 categories, each of them containing 80 images. These data are examined through two proposed models: 1) Hybrid + Support Vector Machine (SVM) + Relevance Feedback (RF); and 2) Hybrid + Cascade Forward Back Propagation Neural Network (CFBPNN) + Relevance (RF). These are two innovative hybrid models, in which intelligent techniques have been incorporated. The SVM model is based on ML, while the CFBPNN model is based on deep learning. According to the results obtained from the experiments carried out with the two databases, the authors concluded that the CFBPNN model gives more accurate results but is slower in calculation than the SVM model.

Wagenpfeil et al. [7] presented a general framework that unites existing algorithms for image and video retrieval in a unified model of indexing, annotation and semantic retrieval. The authors developed a design concept based on "User Centered System Design", which is presented using Unified Modelling Language (UML) and use case design for the main activities. In this use case, the Multimedia Feature Vector Graph (MMFVG) is added, which helps in adding other metadata to digital content. General Multimedia Analysis Framework (GMAF) is a framework that combines and uses existing Multimedia processing systems for image, video, and textual information. New algorithms can be integrated into GMAF very easily. The actual results of GMAF processing show that the level of detail is greatly increased due to the recursive application of the algorithms. Furthermore, MMFVG is designed to be presented in Resource Description Framework (RDF) and queries can be written in the official SPARQL language.

Magliani et al. [8] employed a kNN graph, which makes a connection of all the nodes with each other, but takes a lot of time as a process. In this case, the Locality Sensitive Hashing (LSH) method was used, which saves time in building the graph compared to brute-force or divide-and-conquer methods. In this kNN graph, which is built based on database images, diffusion is applied, which helps distinguish nodes, assigning them different values depending on the similarity of the images. The diffusion applied in this paper is similar to the Google PageRank algorithm where a graph is selected with the help of diffusion which is used iteratively. To make comparisons between the query image and the data set image, the Euclidean distance is used. Each individual is assigned a probability, including genes. The best individuals are assigned from a buffer, and by comparing generations of individuals, in the end, only the individuals with the highest probability remain. This genetic algorithm is implemented using DEAP1. The authors used three image datasets for the experiment. Oxford5k contains 5063 images belonging to 11 classes. Paris6k contains 6412 images belonging to 12 classes. Flickr1M [19] contains 1 million Flickr images used for large. By comparing the results, it was proven that the genetic algorithm gives the same or better results compared to other algorithms such as random search, grid search, and PSO. The method proposed in this paper, utilizing kNN graphs, a diffusion process, and a genetic algorithm, has applications in CBIR technology, enhancing both time and quality optimization.

In [9], Tzelepi and Tefas present a new model of RF which uses deep Convolutional Neural Networks (CNNs). So, the goal is to use the user feedback from the CNN, which modifies its structure and provides better representations in the image return. The reason for using deep CNN is that recently a great advancement has been seen in image classification, digit recognition and pedestrian detection.

CNNs belong to deep learning algorithms which are based on deep neural network architecture. Also to improve the retrieval quality, the authors proposed retraining the convolutional architecture on a data set with relevant image statistics and the tested database classes. The presented method for the improvement of deep CNN affects the permanent improvement of the CBIR system. This is based on two stages. In the first stage, the system collects information from the feedback of various users and stores it. This information consists of questions and images relevant and irrelevant to these questions. Then, in the second phase, the system builds targets for each image based on user queries. In this paper, two image retrieval datasets are used for the experiment: the 102 category flower dataset, consisting of 8189 images divided into 102 categories, and the Inria Holidays 3 dataset, consisting of 1491 images, divided into 500 classes. After obtaining results from the experiments, the authors claim that there is a great improvement in the proposed model.

Shikha et al. [10] proposed a unique hybrid system that is based on content and extracting different attributes such as texture, color, and shape with the help of a Gray-level co-occurrence matrix (GLCM). For the development of this hybrid model, the authors have analyzed several different techniques such as Color Moment, Gray Level Co-occurrence Matrix, Region-props Process, Extreme Learning Machine, and Relevance Feedback. As for Color Moment, it has also been used as a technique to extract color regardless of angle, rotation, or scale. While extracting the shape features, the Mass, Centroid, Mean, Variance, and Dispersion parameters were calculated. Four large databases are used in this paper, and most of them are used in most of the experimental works. They are Corel-1K, Corel-5K, Corel-10K and GHIM-10. In forming a query or input image, all images of all datasets are used. If the acquired images are similar or consistent with the input image, then we say that this system is effective and the image retrieval system has been successful. The similarity between the input image and the output images is found by comparing all the images of the four databases. Similarity is then calculated using three distance metric techniques. Usually, the Euclidean distance is used as a basis because it gives a much more accurate and faster result compared to the Manhattan distance and the Minkowski distance, which very often give false results. Finally, the authors conclude that this Hybrid model is very effective and provides a concrete CBIR solution.

In [11], Magliani et al. use four large datasets to perform experiments, including Holidays with 1491 high-quality images, Flickr1M with 1 million images, SIFT1M consists of 1 million 128D Sift descriptors, GIST1M consists of 1 million 960D Gist descriptors. First, an analysis of the results obtained after the experiment was carried out for the first two datasets, i.e. Holidays + Flickr1M, where they refer to the acquisition time and acquisition accuracy. As the authors report, LSH and Multi-probe LSH obtained the best results but required a huge average query time. PP-index reduced the retrieval time but with a loss in terms of accuracy. The FLANN technique is also used, which is an open-source library for ANN, which is very successful in nearest-neighbor matching. It has achieved an accuracy rate of 83.97%. In the second scenario, the PP-index has achieved a withdrawal of 94.32%, but it takes a lot of time, i.e. 17 sec. LOPQ has a low accuracy of 19.93%, with a computation time of only 3msec. FLANN has achieved better results than the LOPQ technique but is still weaker than the PP-index.

Arun et al. [12] deal with the categorization of image extraction techniques that are part of the Bag of Visual Words (BoVW) model. In this model, the visual dictionary is built through K-Means clustering which is obtained from local image descriptors. Then these images are categorized using visual word histograms. The BoVW-based representation is derived from the given query and each image dataset is then classified to determine the degree of similarity between the images. The distance function is one of the most frequently used metrics for BoVW-based similarity determination. Six datasets are used throughout this paper. INRIA holiday dataset, Scene-15 dataset, Oxford dataset, GHIM-10K dataset, IAPR TC-12 dataset, and SUN-397 dataset. Depending on the capacity of the datasets, samples were also taken. To more easily extract the features of the images of interest, the authors have proposed the Hessian-Affine detector and the Scale Invariant Feature Transform (SIFT) descriptor which provide reasonable solutions. For obtaining the similarity between two histograms, it has been observed that the best metric is Chi-square. After reviewing the literature, the authors have proven that three approaches that influence the increase in the effectiveness of image restoration: (i) approaches that minimize the quantization error, (ii) approaches that minimize the semantic loss and (iii) approaches that incorporate spatial information of visual words. One of the challenges is that during the visual learning of the dictionary and the local encoding of the descriptor, some information is lost during the process. Each of the existing approaches tries to address only one specific limitation of the

BoVW model. A unique BoVW-based system is missing which would address all the issues that are still unresolved.

In [13], Devi and Parmar present various techniques that are part of the CBIR technology. Here they explain that there are four stages of image retrieval from a database based on geometric properties in the input image. They are the image database generation phase, input images from scanner-thinning editing, outline-based image retrieval phase, and identifying global shape similarity. The authors also give brief descriptions of CBIR features which are Texture Features, Color Features, Spatial Location Features, Shape Features, and Local Image Features.

Kumar and Esther [14] address three image feature extraction techniques, Gabor, Wavelet, and Histogram. Color Histogram is one of the most used techniques in CBIR, which is based on RGB space. In Matlab, the maximum number of bins used by Color Histogram is 256. The similarity between the Query image and the database image is calculated through distance metrics. The Gabor technique, as we have mentioned above in other papers, extracts information from an image. This technique is a multi-scale, multi-resolution filter. This two-dimensional filter is presented as a sinusoidal signal. By applying the Wavelet method, they divide the image into four sub-images with three bands: diagonal, vertical, and horizontal. These sub-images contain information about the texture.

Atlam et al. [15] performed a comparison between three techniques for extracting image features including Color Histogram, HSV Color Histogram, and Color Histogram Equalization. The authors have experimented using the WANG database, which is a subset of the Corel database, using 1000 images, divided into 10 categories of 100 images each. Euclidean distance and the correlation coefficient were used as similarity metrics, while the requirement was that 20 similar images be returned for each query image. Time, accuracy, and error rate were also calculated. After the experiment they concluded that HSV color histogram and color histogram give the best results but with different calculation times. As a metric for measuring similarity, the correlation coefficient turned out to be more accurate than the Euclidean distance.

3. Conclusions

After reviewing the literature, some conclusions can be drawn about CBIR. In recent years, significant progress has been made in this field, leading to an increase in the number of large databases designed for CBIR. Many authors have developed specialized methods aimed at enhancing the ease of use and accuracy of image retrieval, while also reducing computation time. Below is a list of some of the main methods for CBIR that proved to be quite practical and recommended to use.

1. LNP model - after experiments and comparisons with other similar models such as LBP, LDB, and LTrP, the most accurate model is LNP.
2. Regarding the HSV color space we have analyzed two models from which CVH can describe color, texture, shape, and other spaces. We can also say that in the representation of colors, and edges, as well as the spatial representation of colors, the CVH method is better than LPB, but in the representation of local structural information LPB is better than CVH.
3. Gaussian Mixture is also a fairly developed method in image restoration.
4. In the CBIR technology based on the conducted experiments, it turns out that the genetic algorithm gives the same or better results compared to other algorithms such as random search, grid search, and PSO.
5. Deep CNNs have also influenced the development of CBIR. So a new model of Relevance Feedback is proposed which uses CNN, where the goal is to use feedback from CNN, which modifies its structure and provides better representations in the image return.
6. A hybrid system developed in [10] analyzes different techniques such as Color Moment, Gray Level Co-occurrence Matrix, Region-props Process, Extreme Learning Machine, and Relevance Feedback, and turns out to be quite accurate in image retrieval.
7. One of the other widely used techniques in CBIR is Color Histogram, which is based on RGB space.

These are seven of the most important developments that have had an immediate effect on the improvement and expansion of image retrieval in general. In the end, we have managed to show a very

real state of development of CBIR, also presenting the orientation of this technology which is multidimensional. We have listed the most commonly used techniques in CBIR, through which many advances have been made that increased the quality and use of CBIR. However, in each of them, we have listed the shortcomings as well as research challenges that should be addressed in future research.

References:

[1] Clement H. C. Leung, R. F. Ma & Jiayan Zhang (2020). Adaptive Multimedia Indexing Using Naïve Bayes Classification. ADVCOMP 2020: The Fourteenth International Conference on Advanced Engineering Computing and Applications in Sciences, pp. 12-17.

[2] Maher Alrahhal, K.P. Supreethi (2019). Content-Based Image Retrieval using Local Patterns and Supervised Machine Learning Techniques. CSE Dept., JNTUH Collage of Engineering, IEEE, pp. 118-124.

[3] Ji-Zhao Hua, Guang-Hai Liu, Shu-Xiang Song (2019). Content-Based Image Retrieval Using Color Volume Histograms. International Journal of Pattern Recognition and Artificial Intelligence, Vol. 33, No. 11.

[4] Yong Hou, Qingjun Wang (2018). Research and Improvement of Content-Based Image Retrieval Framework. International Journal of Pattern Recognition and Artificial Intelligence. Vol. 32, No. 12, 1850043.

[5] Nilanjana Saha, P. Thiruvannamalai Sivasankar, Rahama Salman, Badria Sulaiman Alfurhood, Aishwary Awasthi, & Sukhwant Singh Bindra. (2022). Novel Technique in Content-Based Image Retrieval using Classification by Deep Learning in Artificial Intelligence. International Journal of Intelligent Systems and Applications in Engineering, 10(2s), pp. 256-259.

[6] Shefali Dhingraa, Poonam Bansal (2022). Designing of a rigorous image retrieval system with amalgamation of artificial intelligent techniques and relevance feedback. Journal of Intelligent & Fuzzy Systems. vol. 42, no. 2, pp. 1115-1126.

[7] Stefan Wagenpfeil, Felix Engel, Paul Mc Kevitt, Matthias Hemmje (2021). AI-Based Semantic Multimedia Indexing and Retrieval for Social Media on Smartphones. Information. Vol. 12, Issue 1: 43.

[8] Federico Magliani, Laura Sani, Stefano Cagnoni, Andrea Prati (2019). Genetic algorithms for the optimization of diffusion parameters in content-based image retrieval. Proceedings of the 13th International Conference on Distributed Smart Cameras, pp. 1-6.

[9] Maria Tzelepi, Anastasios Tefas (2018). Deep convolutional learning for Content-Based image retrieval. Neurocomputing. Vol. 275 pp: 2467-2478.

[10] B Shikha, P. Gitanjali, D. P. Kumar (2020). An extreme learning machine-relevance feedback framework for enhancing the accuracy of a hybrid image retrieval system. International Journal of Interactive Multimedia and Artificial Intelligence. Vol. 6, No. 2, pp: 15-28.

[11] Federico Magliani, Tomaso Fontanini, Andrea Prati (2021). Bag of indexes: a multi-index scheme for efficient approximate nearest neighbor search. Multimedia Tools and Applications. Vol. 80, pp. 23135-23156.

[12] K. S. Arun, V. K. Govindan, S. D. M. Kumar (2020). Enhanced bag of visual words representations for Content-Based image retrieval: a comparative study. Artificial Intelligence Review. Vol. 53, pp. 1615–1653.

[13] Pooja Devi, Mahesh Parmar (2017). A survey on cbir techniques and learning algorithm comparison. International Journal of Latest Trends in Engineering and Technology. Vol. 8, No. 1, pp. 197-205.

[14] Ashok D. Kumar, J. Esther (2011). Comparative study on CBIR based by color histogram, Gabor and wavelet transform. International Journal of Computer Applications. Vol. 17, No. 3, pp. 37-44.

[15] Hany Fathy Atlam, Gamal Attiya, Nawal El-fishawy (2013). Comparative Study on CBIR based on Color Feature. International Journal of Computer Applications. Vol. 78, No. 16, pp. 9-15.