Review about SIFT and Local Feature Extraction in Content Based Image Retrieval

Soran Ahmed Hasan[1], Asst. Prof. Dr. Gullanar M Hadi[1]

[1] Department of Software Engineering and Informatics, College of Engineering, Salahaddin University- Erbil, Kurdistan Region, Iraq

***Abstract*— As internet technology expands and the widespread use of digital devices, Content Based Image Retrieval CBIR has seen rapid development and application across a range of areas in computer vision and artificial intelligence Today, it's possible to retrieve related images efficiently and effectively from large scale databases using just an input image, In the last decade, there has been a significant push towards developing new CBIR theories and models, resulting in the establishment of many effective CBIR algorithms. CBIR is a crucial tool for locating images within a large dataset that share similarities with a certain query image., with the process generally involving the comparison of key features of the query image against those in the dataset to help rank and retrieve the most relevant images, the general aim of the paper is to explore the key concepts and methods central to Content Based Image Retrieval emphasizing the importance of sophisticated feature extraction techniques like SIFT,SURF,ORB, and how machine learning and deep learning are transforming the efficiency and accuracy of image retrieval, It sheds light on the significant shifts in how images are processed and retrieved, highlighting both the new opportunities and challenges emerging in the field.**

***Index Terms*— CBIR, Local Descriptors, SIFT, Feature Matching, Machine Learning**

I. Introduction

The growing number of different digital devices has generated several possibilities for capturing, sharing and retrieving digital photographs from extensive libraries, Previously accessing the Managing photos was a straightforward and effortless task because of the limited number of photographs in repositories Gradually, the process became more challenging as a result of the growing diversity and intricacy of images, coupled with a substantial volume of images, Novel methodologies and strategies have been devised to extract images based on user preferences, and image retrieval continues to be an important area in the fields of multimedia and computer vision(Li et al., 2021).

Content Based Image Retrieval is a system specifically developed to return images that are similar to the query image submitted by the user, The images that are obtained may have similarities in terms of texture, shape, or color composition, The architecture of Content Based Image Retrieval usually includes three separate stages: feature extraction and selection, similarity search, and indexing and retrieval of the most similar images Several innovative research have proposed distinct approaches for these three stages However, feature extraction and selection are the primary and critical steps in studying a picture, whether it is examined as a whole or in parts. Several writers have suggested feature descriptors that precisely capture the color, texture, and shape of an object in the image however the similarity measure is regarded as the second stage in minimizing the semantic disparity between the images. The algorithm involves measuring the difference between visual descriptors such as Euclidean or Manhattan distances, The process of ranking images involves three stages with indexing and retrieval being the third stage (Unar et al., 2019a)Feature descriptors are frequently employed sophisticated methods for extracting the visual attributes of an image, These descriptors can be categorized into two distinct classes Global descriptors and local descriptors refer to two types of feature representations used in computer vision and image processing, Global descriptors capture the overall characteristics of an image or a region, whereas local descriptors focus on capturing specific details or patterns within an image or a region. Global descriptors comprehensively analyze the entire image, taking into account its color, texture, and shape, Local descriptors possess the capability to analyze and extract particular objects inside a picture. In recent decades, widely-used local descriptors such as scale invariant feature transform (SIFT), histogram of oriented gradient (HoG) and speed-up robust feature (SURF), have been extensively employed for image retrieval. The image feature representations generated by the Scale Invariant Feature Transform (SIFT) or Speeded Up Robust Features (SURF) continue to hold great importance in the field of computer vision. There has been notable advancement in the characterization of visual features in recent years, encompassing both manual procedures and the application of machine learning methods, as well as the development of distinct image descriptors. The dominant phenomenon is based on Convolutional Neural Networks (CNN), also known as Deep Learning. Computer vision is

DOI:http://doi.org/10.24086/cocos2024/paper.1533

applied to extract things from a given scene. The object retrieval task generally consists of two sequential stages: visual object search and subsequent retrieval.(Amitha and Narayanan, 2020).The structure of this paper is as follows Section 2 discusses the evolution and historical background of CBIR. Section 3 introduces the core principles and fundamental operation of CBIR. Sections 4 and 6 the Local Feature Extraction approaches discuss the properties of SURF, SIFT, HOG, ORB, machine learning and deep learning in relation to CBIR. Following that, Section 7 delves into the performance comparison of operators in local Feature Extraction Techniques. The applications, obstacles, and open areas for research are presented in sections 8 and 9. The paper is ultimately concluded in Section 10.

II. BACKGROUND

A. Historical perspective of image retrieval

The historical aspect of image retrieval revolves around the development of techniques for finding images based on their content initially there was a focus on text based image retrieval (TBIR) which relied on annotations or metadata associated with the images However TBIR had limitations when it came to datasets and couldn't effectively capture the content of images, Content-based image retrieval (CBIR) appeared in the 1980s as a solution to address these difficulties. CBIR approaches are designed to extract elements such as color and texture from images in order to quantify their similarity. Color features such as color histograms and color moments offer information about the distribution of colors in an image, On the other hand texture features focus on the arrangement of gray intensities and can be quantified using techniques like local binary patterns (LBP) and gray level cooccurrence matrix (GLCM), These extracted features are then used to create feature vectors that represent the images and enable comparison for similarity in retrieval moreover a significant milestone in the history of image retrieval was marked by the introduction of algorithms, like Scale Invariant Feature Transform and similar algorithms revolutionized feature extraction by enabling robust analysis leading to improved accuracy and reliability of CBIR systems and over the years the combination of machine learning and deep learning methods has greatly improved CBIR systems (Kaur & Singh 2020; Li et al., 2021) These advancements have led to a grasp of content allowing for more advanced and nuanced image retrieval abilities. *Final Stage*

When you submit your final version, after your paper has been accepted, print it in two-column format, including figures and tables. You must also send your final paper via e-mail. Also, send a PDF page with complete contact information of all authors. Include full mailing addresses, telephone numbers, and e-mail addresses.

B. Transition from text-based to content-based methods

The shift from relying on text descriptions to utilizing content based image retrieval methods represented an advancement in the categorization and search of images in Initially finding images depended on information,This approach often fell short and failed to capture the true essence of the picture However with the introduction of Content Based Image Retrieval (Hadid et al., 2023) CBIR revolutionized image searching by examining the characteristics of an image itself such as its colors, textures and shapes this meant that we could bypass the limitations of text labels and gain an understanding of what we were seeking, This transition was made possible by advancements in algorithms like SIFT and technologies like machine learning resulting in an accurate and objective method for retrieving images that significantly enhanced the capabilities within this field.

C. Core Principles of CBIR

The basic structure of Content-Based Image Retrieval consists of fundamental phases and other optional stages as depicted in Figure1. The first phase entails the user submitting the query image which will subsequently experience the same applicable processes as all the images in the database, following a sequential order, Usually, these procedures are performed on the query image at the time of user submission and are called online processes, The same procedures can also be applied to dataset photos prior to submitting the query and are referred to as offline operations The structure of the framework may include an optional preprocessing stage that can involve operations like as resizing, segmentation, denoising, and rescaling Following the optional phase, the feature extraction stage ensues, which is widely regarded as the most critical stage, This stage entails transforming a visual concept into a numerical representation. The retrieved features can include fundamental characteristics such as shape ,color, texture, and spatial information, as well as local descriptors (Chavda & Goyani, 2019). Normalization or categorization is an optional preprocessing step that can be carried out following feature extraction. In the final stage, the retrieved attributes of the query image are compared to those of all other photos in the collection to determine the most appropriate matches, Ultimately Relevance feedback is an additional phase that enhances the outcomes by enabling user engagement in determining the relevance of the retrieved photos, Various techniques have been proposed to incorporate pertinent feedback for enhancing the operational efficiency of (CBIR) systems (Malik et al., 2019).

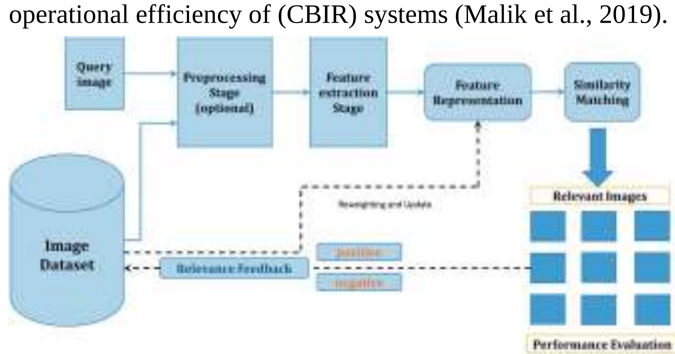

Figure 1.CBIR system Block diagram

Metadata significantly enhances the functionality of (CBIR) systems, It provides additional context through tags, descriptions, and annotations, improving retrieval accuracy and efficiency, CBIR systems leverage metadata to enable searches

DOI:http://doi.org/10.24086/cocos2024/paper.1533

based on specific image criteria like shape ,color, and texture, without manual annotation, This saves time and streamlines image management in large datasets, The integration of metadata not only boosts CBIR performance and usability but also negates the need for time-intensive manual image tagging, leading to a more efficient retrieval process (Unar et al., 2019b)

III. MAIN BODY

A. Local Feature Extraction techniques in CBIR

In computer vision, techniques such as object detection, classification, and retrieval entail the comparison of images by utilizing extracted features. Initially low level features such as color, shape, texture, and spatial layout (global features) were used (Celik & Bilge, 2017) However these features were limited as they weren't fully invariant to changes like translation, rotation, and scaling. To address these limitations, local descriptors were introduced, focusing on local image structures. These are robust to image processing and are better suited for computer vision tasks, being invariant to scaling, rotation, and partly to illumination and 3D camera views. The development of color-based invariants has further enhanced performance. Local descriptors are classified into distribution based, spatial frequency, and other types, with examples like SIFT, SURF, HoG,ORB, LBP, and LTP (Malik et al., 2019)

B. SIFT

Lowe proposed SIFT features (Lowe, 2004), which extract features that are somewhat invariant to light and 3D camera view, and invariant to scaling and rotation. These are also well-oriented in the frequency and spatial domains. Deterioration chance is decreased as a result. The extraction of the SIFT feature involves four distinct steps: detecting scale space extrema, localizing key points, assigning orientations, and describing the key points, The process of key point extraction in images involves four main steps. First, potential key points and their scales are identified using the Difference of Gaussian (DoG) applied to various scales of the image. Second, less significant points, like those with low contrast or on edges, are removed. Third, each key point is assigned one or more orientations based on local gradient directions Finally, a description of each key point is created using gradient magnitude and orientation, ensuring invariance to scaling, rotation, and illumination changes.Typically, when an object is represented in a 500x500 pixel image, around 2000 features are generated. Despite the presence of significant occlusion, this technique is still able to recognize the object in the image.(Ethan Rublee, 2011).The SIFT algorithm follows the following steps:

- o Identify the approximate position and size of the prominent feature point.
- o Precisely adjust their location and size
- o Establish the orientation(s) for each keypoint.
- o Establish the descriptors for each keypoint..

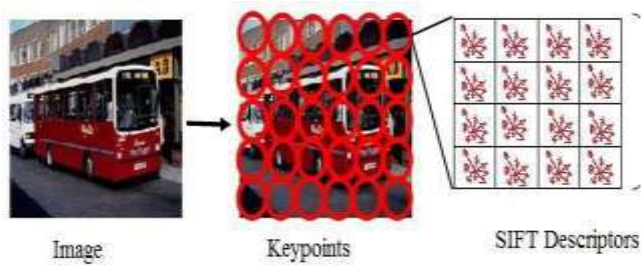

Figure 2. sift feature extraction

C. SURF

The development of local descriptors in image processing has led to the development of SURF (Speeded-Up Robust Features), an alternative to SIFT (Scale-Invariant Feature Transform). SURF is known for its reproducibility of key points across different viewpoints, contributing to its robustness. The SURF feature extraction process involves two stages: detection using the Hessian matrix and description using Haar Wavelet. The descriptor is designed with scale and rotational invariance, allowing for the identification of scaled features in another image. The detector aids in identifying significant aspects based on the description of image characteristics. The SURF feature detector uses a Gaussian second derivative mask at various scales, with the Laplacian operator for edge detection and the Gaussian operator for smoothing. The Laplacian of Gaussian (LoG) operator is used in the SURF algorithm, which can only achieve a 45 degree rotation for each axis, enhancing its resistance to rotation(Bay et al., 2008).

D. HOG

The (Histogram of Oriented Gradients) features, which were first introduced by Dalal for human detection (Dalal et al., 2005), are regarded as a simplified variant of SIFT due to their reliance on local gradient orientation computation. The procedure of collecting HoG features entails multiple sequential steps: adjusting the color and gamma value of the image, partitioning the image into cells and computing gradients for each cell, normalizing histograms for each block of cells, and ultimately obtaining HoG features by combining these blocks.

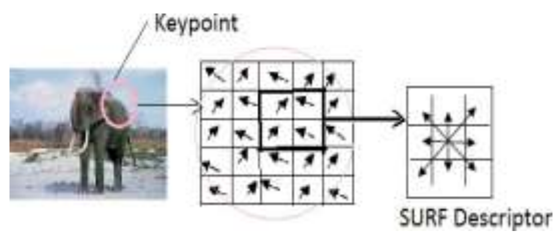

Figure 3. SURF feature Extraction

E. ORB

The (Oriented FAST and Rotated BRIEF) , developed in OpenCV Labs, combines the FAST keypoint detector and the binary BRIEF descriptor(Chhabra et al., 2020). It's designed to efficiently extract a limited number of high-quality features from images. Compared to SIFT and SURF, ORB requires less

DOI:http://doi.org/10.24086/cocos2024/paper.1533

computational power and operates faster, particularly surpassing SURF in speed, making it an efficient choice for feature extraction in various computer vision applications (Rublee et al., 2011).

F. MACHINE LEARNING IN CBIR

Artificial intelligence algorithms that allow computers to learn and improve from experience ,In recent times, CBIR systems have transitioned towards utilizing machine learning techniques to acquire a model capable of effectively handling new input data and providing accurate predictions. This advancement aims to enhance the performance of image search. This discussion will focus on the prevalent machine learning algorithms utilized in the CBIR area, along with the latest research contributions. This section will be divided into three distinct subsections unsupervised learning (specifically clustering), supervised learning (specifically classification), and deep learning. (Kaur & Singh, 2020).

1. Unsupervised learning

Following the completion of the feature extraction procedure and the development of the feature vector, clustering is carried out. Clustering involves grouping picture descriptors into distinct groups that differ semantically from one another. Clustering is classified as an unsupervised learning algorithm due to its lack of prior information. Determine the appropriate cluster for the data of the photos using K-means . K-means and hierarchical clustering are the predominant clustering algorithms utilized in Content-Based Image Retrieval (CBIR), particularly when systems rely on local feature extraction techniques. Typically, these methods are accompanied with a clustering procedure to determine the semantic category to which the image is assigned.

2. Supervised learning

Supervised learning methods, such as Support Vector Machines (SVM), are employed in the fields of recognition of patterns and image categorization. They possess previous understanding of picture clusters and tags, rendering them into categorization assignments. .There are two categories: linear and non-linear. Linear Support Vector Machine (SVM) is capable of separating characteristics into two distinct groups, whereas non-linear SVM employs kernel functions to facilitate the separation(Sharif et al., 2019).

3. Deep learning

Traditional CBIR techniques use color, shape, and texture features, along with local descriptors like SURF and SIFT. With the rise of deep learning and image databases, Convolution neural network (CNN) has become a popular choice in computer vision for efficient image retrieval (Naeem et al., 2023). Deep learning surpasses human performance in object recognition by bridging the semantic gap in image retrieval. However, transitioning from image classification to retrieval presents challenges. CBIR with CNN focuses on creating effective image representations by training and testing models to improve image retrieval accuracy. Challenges arise from dataset variability and the complexity of building effective training sets. A solution involving CNN recognizes similarities in images from different locations and incorporates geo-location for precise separation and categorization.

G. CHARACTERISTICS OF LOCAL DESCRIPTORS

SIFT is a robust feature extraction technique that is invariant to scale, rotation, and affine transformations. It performs well in images with a simple background and represents them without noise. However, it is not fast enough for real-time applications. Various variants of SIFT have been proposed to address this issue, such as root-SIFT, color SIFT affine SIFT, edge SIFT, CSIFT, NSIFT, and PCA-SIFT.(Dewan & Thepade, 2020), SURF is a suitable alternative for SIFT and is widely regarded as an enhanced iteration of SIFT. It exhibits higher speed and greater durability, necessitating less time for feature calculation and matching. The indexing system employed is predicated on the Laplacian sign. Nevertheless, SURF exhibits subpar performance when it comes to rotation. It exhibits greater resilience in the face of variations in lighting and size. Integrating SURF with other descriptors, such as FREAK, can improve the accuracy of categorization (Celik & Bilge, 2017).

ORB is a binary feature descriptor that is quicker to compute compared to SIFT and SURF ,It is rotation and multiscale invariant and suitable for low-complexity and low-power hardware. Each keypoint is represented as a bit, making it efficient and reliable. However, ORB has constraints in its descriptive capacity and its ability to maintain scale invariance under specific circumstances (Unar et al., 2019b),HOG captures the distribution of gradient orientations in an image, providing information about pixel orientations. limited in terms of spatial information about neighboring pixels.Co-occurrence histogram of oriented gradient (CoHOG) is an extension of HOG that incorporates spatial information.CoHOG, when combined with other descriptors like SURF, can enhance the performance of content based image retrieval systems. HOG is not directly applicable to multispectral images.(Hameed et al., 2021).

H. COMPARATIVE ANALYSIS OF FEATURE EXTRACTION TECHNIQUES

1. Sift with CNN

This excerpt describes using standard settings to extract CNN features from Google Net's pool5 layer and SIFT features aggregated using VLAD. Both features are dimensionally aligned using PCA. From figure 2 Results show CNN and SIFT have comparable performance. An experiment combining these features demonstrated significant performance enhancement, indicating their complementarity. However, The example additionally suggests that this combination may not consistently include all comparable images, highlighting the necessity for advanced integration approaches. We assess the complementary nature of CNN and SIFT characteristics for image retrieval tasks. In order to take use of the complementary nature, we suggest using the CCS model to depict scene-level information(Yan et al., 2016)

DOI:http://doi.org/10.24086/cocos2024/paper.1533

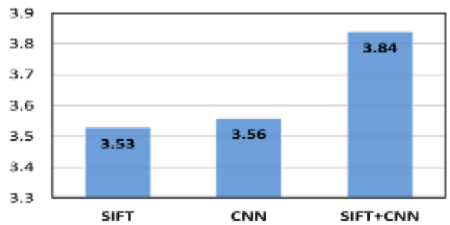

Figure 4. compares the performances of SIFT, CNN, and a simple combination of both.

2. *SIFT WITH SURF*

A comparative analysis was conducted between the SURF and SIFT detectors, focusing on their performance ,The parameters for comparison were response time and accuracy. A Penn State College of Information Technology Sciences and Technology dataset was used for the experiments. These tests show that SURF generates fewer but more significant local features, demonstrating its greater efficiency, stability, and speed for image retrieval. Our findings indicate that SURF outperforms SIFT by at least 5% in terms of precision and recall. Regarding response time, it has been demonstrated that SURF can get 40 photos in 4 seconds, whereas SIFT can only return 33 images, indicating a 21% improvement. Thecomparisons showed in figure 3(a,b,c,d)(Rublee et al., 2011)

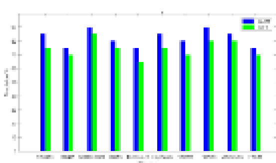

Figure 5. -a. based on Precision

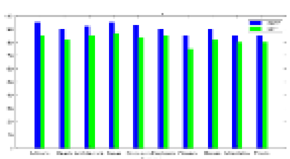

Figure 5. -b. based on Recall

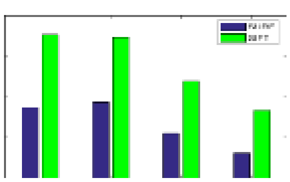

Figure 5. -c. Number of features detected by SIFT ,SURF

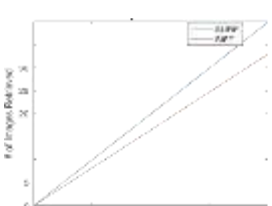

Figure 5. -d. Time Response of SIFT ,SURF

3. *SIFT with ORB*

ORB and SIFT features are used for efficient retrieval of images from large datasets. However, due to their size and complexity, K-means clustering and Locality-Preserving Projections (LPP) are applied to reduce space and complexity. K-means clusters descriptors into 32 groups, while LPP reduces them to 4 or 8 components. The system's performance, measured in precision, Root-mean-square error(RMSE), and time efficiency, showed high precision rates for both Wang and Corel datasets, outperforming existing CBIR systems. For photos in the Wang and Corel databases, a combination of SIFT and ORB feature vectors is examined, achieving the maximum precision rates of 99.53% and 86.20% for the Wang and Coral databases, respectively (Chhabra et al., 2020; Rublee et al., 2011)the Tables (1,2) show the RMSE result.

Figure 6. a. Extracted Features by ORB key points

Figure 6. b. Extracted Features by SIFT key points

Features 10-class dataset	No of Feature vector dimensions	Decision tree (%)	Random forest (%)	MLP (%)
ORB	4	24.74	21.83	23.92
	8	26.95	21.63	23.18
SIFT	4	25.47	22.30	25.44
	8	27.29	21.63	22.37
ORB + SIFT	4 + 4	16.51	14.18	17.24
	8 + 8	17.28	15.67	30.76

Figure 7. .a RMSE for Wang dataset(Rublee et al., 2011)

Features 40-class dataset	No of Feature vector dimensions	Decision tree (%)	Random forest (%)	MLP (%)
ORB	4	10.32	8.82	10.16
	8	10.89	8.78	10.59
SIFT	4	5.52	3.28	10.04
	8	5.55	2.74	9.69
ORB + SIFT	4 + 4	4.92	2.77	5.37
	8 + 8	5.00	3.34	5.46

Figure 3. b RMSE for corel dataset(Rublee et al., 2011)

I. APPLICATIONS OF CBIR

In literature, Content-Based Image Retrieval (CBIR) applications are divided into three main categories, each determined by the user's search objective (Srivastava et al., 2023):

Associative Search: This category encompasses a search process in which the user engages in exploratory behavior without a specific objective, but rather aims to uncover interesting items. The approach is iterative, continuously improving the search by incorporating user comments regarding relevance. An early example involved using a rough

DOI:http://doi.org/10.24086/cocos2024/paper.1533

sketch as a query image in a visual interface for database access, indicating a general, not specific, search interest.

Search by a Specific Image: Here, the user is intent on finding a particular image. This could be a specific image in mind, a similar image to one they already have, or a target image. This category includes diverse applications such as searching art catalogs or finding specific items like stamps, industrial components, and artwork.

Search by Category : focuses on retrieving images from a specific category. The user starts with an image and seeks other images within the same category, Examples include image retrieval based on line patterns, pattern based retrieval (like stripe patterns in clothing), classifying trademarks and categorizing entities such as human faces, sunflowers, bicycles, airplanes, cameras, and so on. This category encompasses a broader range of classifications and pattern recognitions. Content Based Image Retrieval is extensively employed in several industries such as fashion, medical imaging and satellite photography, It is also utilized in popular search engines like Google, Bing, and Yahoo. The system operates by examining and classifying photos according to their visual attributes, such as color, texture, and shape The efficacy of CBIR systems relies on their ability to accurately interpret certain facets of images. CBIR driven by developments in internet technology and digital devices has a wide range of applications in numerous areas, The applications of this technology range from retrieving fashion images and identifying individuals to searching for products in ecommerce, retrieving remote sensing images, and searching for trademark images. Each of these applications utilizes the capabilities of Content Based Image Retrieval to effectively handle and analyze distinct categories of images that are pertinent to sectors such as medical, military, commerce, and art Domains used to categorize and arrange photographs based on distinct visual attributes (Li et al., 2021).

IV. DISCUSSION

The exploration of SIFT and local feature extraction in the context of Content Based Image Retrieval (CBIR) has illuminated the pivotal role these techniques play in the advancement of image retrieval systems. SIFT, along with other local descriptors such as SURF, HOG, and ORB, has significantly contributed to the evolution of CBIR by enhancing the precision of image searches. The effectiveness of these techniques in identifying, matching, and retrieving images based on content similarity is paramount, especially in vast databases where traditional text-based methods fall short. Furthermore, the integration of machine learning and deep learning has further propelled the capabilities of CBIR systems, enabling more nuanced and efficient searches that closely mimic human visual understanding.

The comparative analysis of feature extraction techniques like SIFT with CNN, SIFT with SURF, and SIFT with ORB underscores the complementary nature of combining traditional and modern approaches. This synergy not only improves the performance of CBIR systems but also opens new avenues for addressing the semantic gap challenge inherent in these systems. Moreover, the role of local descriptors in object detection and classification exemplifies the critical importance of feature extraction in the broader domain of computer vision.

V. CONCLUSION

The desire to develop an efficient image retrieval system based on image content arises from the existence of extensive image databases and the lack of an effective text-based image retrieval approach. This article performed a comprehensive study of various works and techniques in the domain of Content Based Image Retrieval additionally explored the fundamental stages of the CBIR architecture and the latest strategies employed to minimize the semantic gap, Traditional CBIR techniques use shape, texture and color features, along with local descriptors like SURF and SIFT remains critical for the performance of CBIR systems therefore recent trends show a move towards hybrid methods combining traditional local feature descriptors with advanced machine learning and deep learning techniques. With the rise of deep learning and image databases, Convolution neural network (CNN) has become a popular choice in computer vision for efficient image retrieval. Deep learning surpasses human performance in object recognition by bridging the semantic gap in image retrieval. However, transitioning from image classification to retrieval poses challenges. CBIR with CNN focuses on creating effective image representations by training and testing models to improve image retrieval accuracy, The advancements in theory, technology, and application of CBIR are clear. However, other obstacles persist, particularly with the advent of huge amounts of data and the use of sophisticated deep learning methods. These factors will significantly impact the role of artificial intelligence in our everyday lives in the future, Ultimately this article highlighted key factors that impact the performance of CBIR and focused on several approaches that could contribute to the development of an innovative CBIR system, Developing an algorithm that attains a high level of retrieval accuracy while also lowering computing expenses is a challenging problem.

VI. RECOMMENDATION

As CBIR systems continue to evolve, leveraging machine learning techniques such as K-Nearest Neighbors, SVM, Decision Trees, and Random Forest has proven instrumental in overcoming numerous challenges, particularly in object recognition and natural language processing tasks within CBIR. These techniques have demonstrated significant success in classification tasks, underscoring the potential of machine learning in enhancing CBIR systems' functionality (Kaur & Singh, 2020). Nonetheless, the performance of these systems largely hinges on the quality of images within databases. Issues such as noise and poor visibility can notably impair retrieval accuracy, pointing to the critical need for high-quality image data (Srivastava et al., 2023). Amidst the rapid advancements in technology and the growing importance of big data and deep learning, CBIR faces ongoing challenges

DOI:http://doi.org/10.24086/cocos2024/paper.1533

that necessitate innovative solutions for its integration into daily life through artificial intelligence (Hadid et al., 2023).

To address these challenges and propel CBIR forward, the following strategic actions recommend:

1. Continue exploring and refining the use of advanced machine learning algorithms in CBIR to further improve object recognition and natural language processing capabilities.
2. Implement advanced preprocessing techniques and quality control measures to mitigate issues associated with noise and poor visibility in image databases.
3. Expand CBIR systems to include diverse data types such as text and audio alongside visual content, enriching the retrieval experience and broadening the systems' applicability.
4. Invest in the enhancement of deep learning technologies for better feature extraction and analysis, ensuring CBIR systems can effectively handle complex image data.
5. Prioritize the development of ethical guidelines and privacy-preserving technologies to safeguard users and maintain trust in CBIR applications.
6. Work towards creating larger and more varied datasets for deep learning, enabling the training of more robust and accurate CBIR models
7. Explore learning strategies that mimic human learning processes for more personalized and efficient small-scale retrieval scenarios.
8. Enhance algorithms and technologies for database searching to improve speed and accuracy, particularly for large datasets.

References

Amitha, I. C., & Narayanan, N. K. (2020). Object Retrieval in Images using SIFT and R-CNN. *2020 International Conference on Innovative Trends in Information Technology, ICITIIT 2020*. https://doi.org/10.1109/ICITIIT49094.2020.9071557

Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, *110*(3), 346–359. https://doi.org/10.1016/j.cviu.2007.09.014

Celik, C., & Bilge, H. S. (2017). Content based image retrieval with sparse representations and local feature descriptors : A comparative study. *Pattern Recognition*, *68*, 1–13. https://doi.org/10.1016/j.patcog.2017.03.006

Chavda, S. M., & Goyani, M. M. (2019). Recent evaluation on Content Based Image Retrieval. *International Journal of Computer Sciences and Engineering*, *7*(4), 325–329. https://doi.org/10.26438/ijcse/v7i4.325329

Chhabra, P., Garg, N. K., & Kumar, M. (2020). Content-based image retrieval system using ORB and SIFT features. *Neural Computing and Applications*, *32*(7), 2725–2733. https://doi.org/10.1007/s00521-018-3677-9

Dalal, N., Histograms, B. T., & Triggs, B. (2005). *Histograms of Oriented Gradients for Human Detection*. 886–893. https://doi.org/10.1109/CVPR.2005.177ï

Dewan, J. H., & Thepade, S. D. (2020). Image Retrieval Using Low Level and Local Features Contents: A Comprehensive Review. In *Applied Computational Intelligence and Soft Computing* (Vol. 2020). Hindawi Limited. https://doi.org/10.1155/2020/8851931

Hadid, M. H., Hussein, Q. M., Al-Qaysi, Z. T., Ahmed, M. A., & Salih, M. M. (2023). An Overview of Content-Based Image Retrieval Methods and Techniques. *Iraqi Journal for Computer Science and Mathematics*, 66–78. https://doi.org/10.52866/ijcsm.2023.02.03.006

Hameed, I. M., Abdulhussain, S. H., & Mahmmod, B. M. (2021). Content-based image retrieval: A review of recent trends. In *Cogent Engineering* (Vol. 8, Issue 1). Cogent OA. https://doi.org/10.1080/23311916.2021.1927469

Jothy, R., & Radha, R. (2021). *A Comparative Analysis of SIFT, SURF and ORB on Sketch and Paint based images*. https://www.researchgate.net/publication/368189389

Kaur, P., & Singh, R. K. (2020). An efficient analysis of machine learning algorithms in CBIR. *Proceedings of International Conference on Computation, Automation and Knowledge Management, ICCAKM 2020*. https://doi.org/10.1109/ICCAKM46823.2020.9051528

Li, X., Yang, J., & Ma, J. (2021). Recent developments of content-based image retrieval (CBIR). *Neurocomputing*, *452*, 675–689. https://doi.org/10.1016/j.neucom.2020.07.139 Lowe, D. G. (2004). Accepted for publication in the. In *International Journal of Computer Vision*.

Malik, N. U. R., Airij, A. G., Memon, S. A., Panhwar, Y. N., Abu-Bakar, S. A. R., & El-Khoreby, M. A. (2019). Performance Comparison between SURF and SIFT for Content-Based Image Retrieval. *Proceedings of the 2019 IEEE International Conference on Signal and Image Processing Applications, ICSIPA 2019*. https://doi.org/10.1109/ICSIPA45851.2019.8977732

Naeem, A., Anees, T., Ahmed, K. T., Naqvi, R. A., Ahmad, S., & Whangbo, T. (2023). Deep learned vectors' formation using auto-correlation, scaling, and derivations with CNN for complex and huge image retrieval. *Complex and Intelligent Systems*, *9*(2), 1729–1751. https://doi.org/10.1007/s40747-022-00866-8

Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. *2011 International Conference on Computer Vision*, 2564–2571. https://doi.org/10.1109/ICCV.2011.6126544

Sharif, U., Mehmood, Z., Mahmood, T., Javid, M. A., Rehman, A., & Saba, T. (2019). Scene analysis and search using local features and support vector machine for effective content-based image retrieval. *Artificial Intelligence Review*, *52*(2), 901–925. https://doi.org/10.1007/s10462-018-9636-0

Srivastava, D., Singh, S. S., Rajitha, B., Verma, M., Kaur, M., & Lee, H. N. (2023). Content-Based Image Retrieval: A Survey on Local and Global Features Selection, Extraction, Representation, and Evaluation Parameters. *IEEE Access*, *11*, 95410–95431. https://doi.org/10.1109/ACCESS.2023.3308911

Unar, S., Wang, X., Wang, C., & Wang, M. (2019a). New strategy for CBIR by combining low-level visual features with a colour descriptor. *IET Image Processing*, *13*(7), 1191–1200. https://doi.org/10.1049/iet-ipr.2019.0098

Unar, S., Wang, X., Wang, C., & Wang, Y. (2019b). *A decisive content based image retrieval approach for feature fusion in visual and textual images* ☆. *179*, 8–20. https://doi.org/10.1016/j.knosys

Yan, K., Wang, Y., Liang, D., Huang, T., & Tian, Y. (2016). CNN vs. SIFT for image retrieval: Alternative or complementary? *MM 2016 - Proceedings of the 2016 ACM Multimedia Conference*, 407–411. https://doi.org/10.1145/2964284.2967252

DOI:http://doi.org/10.24086/cocos2024/paper.1533