International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)
Vol. 11-No.3 (2025) pp. 4500-4506
http://www.ijcesen.com
ISSN: 2149-9144

Research Article

Efficient Image Retrieval using Dense-SIFT for Enhanced Object Segmentation and Key Point Localization

D. Jeyakumari[1*] , A. Nesarani[2], B Manojkumar[3], D. Suresh[4], R. Kiruthikaa[5]

[1]Professor, Department of Electronics and Communication Engineering , Hindusthan Institute of Technology , Coimbatore, Tamilnadu, India
* **Corresponding Author Email:** dgjeyakumari@gmail.com - **ORCID:** 0000-0003-3212-000X

[2]Associate professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vadeeswaram, Guntur, Andra Pradesh, India
Email: nesaraniabraham84@kluniversity.in - **ORCID:** 0000-0002-7604-9957

[3]Assistant Professor, Department of Electronics and Communication Engineering, Karpagam Institute of Technology, Coimbatore, Tamilnadu, India
Email: manoj.dreamtowin@gmail.com - **ORCID:** 0009-0000-4114-405X

[4]Assistant Professor, Department of Electronics and Communication Engineering,St.Joseph's Institute of Technology Chennai, Tamilnadu, India
Email: suresh.d.28@gmail.com - **ORCID:** 0000-0002-9251-4548

[5]Assistant Professor Department of Electronics and Communication Engineering,KGiSL Institute of Technology Coimbatore,Tamilnadu,India.
Email: kiruthikaar27@gmail.com - **ORCID:** 0009-0006-8748-7254

Article Info:

DOI: 10.22399/ijcesen.2507
Received : 22 February 2025
Accepted : 17 May 2025

Keywords

CBIR,
SIFT,
Feature,
Vision programming,
Descriptors,
D-SIFT.

Abstract:

In this paper a new technique for image retrieval is used. Traditional methods used for image retrieval are not supported for large set of databases. By using the features of the image such as color, shape and texture image can be retrieved efficiently. The Content Based image retrieval (CBIR) technique is the traditional technique used for image retrieval. Several kinds of detectors and descriptors such as SIFT, SURF, FAST, BRIEF, ORB, BRISK, FREAK are used for image retrieval. Among these techniques SIFT is quite powerful. The main drawback of the existing system is that it computes only at the interest points. The proposed system addresses about the D- SIFT algorithm in which the SIFT is computed at every pixel, or every kth pixel. The Density- Scale Invariant Feature Transform (D-SIFT) is a stand out amongst the most locally feature detector and descriptors which is utilized as a part of the majority of the vision programming. The main advantage of using Dense SIFT over SIFT is speed. The main goal is to segment the similar object from the two images. As the result of the proposed system the poorly localized points are removed by key point localization.

1. Introduction

Image processing involves changing the nature of an image either improve its pictorial information for human interpretation or render it more suitable for autonomous machine perception.
The digital image processing, which involves using a computer to change the nature of a digital image. digital image define as a two-dimensional function, f(x, y), where x and y are spatial (plane) coordinates, and the amplitude of f at any pair of coordinates (x, y) is called the intensity or gray level of the image at that point. When x, y, and the amplitude values of f are all finite, discrete quantities.
The field of digital image processing refers to processing digital images by means of a digital computer.
Note that a digital image is composed of a finite number of elements, each of which has a particular location and value and the elements are referred to as picture elements, image elements, and pixels. Pixel is the term most widely used to denote the elements of a digital image.

1.2. Task of Image Processing

The image processing has been performed in five tasks. Such as

(I) Enhancing the Representation of an Image

Enhancing the edges of an image to make it appear sharper. It Makes image as a more pleasant image. Sharpening edges is an essential component of printing, in order for an image to appear "at its best" on the printed page.

(II) Identify Matching from an Image

Matching being random errors in the image, it is a very common problem in data transmission all sorts of electronic components may affect data passing through image, and the results may be undesirable. Matching may take many different forms each type of Matching requiring a different method of removal.

(III) Identify Motion Blur from an Image

Motion blur may occur when the shutter speed of the camera is too long for the speed of the object. The photographs of fastmoving objects: athletes, vehicles for example, the problem of blur may be considerable.

(IV) Obtaining the Edges of an Image

Obtaining the edges of an image necessary for the measurement of objects in an image. The edges are used for measure their spread, and the area contained within the image. The edge result is necessary to enhance the original image slightly, to make the edges clearer.

(V) Identify Detail from an Image

The measurement or counting purposes, all the details of an image is not necessary. Example, a machine inspected items on an assembly line, the only matters of interest may be shape, size or color is used to simplify the image. Measure the size and shape of the animal without being distracted by unnecessary detail.

2. Related Works

This is the most common form of text search on the Web. Most search engines do their text query and retrieval using keywords.

The keywords based searches they usually provide results from blogs or other discussion boards. The user cannot have a satisfaction with these results due to lack of trusts on blogs etc. low precision and high recall rate.

In early search engine that offered disambiguation to search terms. User intention identification plays an important role in the intelligent semantic search engine.

The similarity assessment is fundamentally important to many multimedia information processing systems and applications such as compression, restoration, enhancement and copy detection etc. The image similarity assessment is to design algorithms for repeated and objective evaluation of similarity in a consistent manner with individual human judgment. The Peak signal-to-Matching ratio(PSNR), Human visual system(HVS)and Natural Scene Statistics(NSS) are efficient to measure the quality of an image evaluated with its original version, particularly for some image restoration applications. The existing methods mainly focus on evaluating the similarities between a reference image and its non-geometrically variation versions, such as decompressed and brightness/contrast-enhanced versions. Image Quality of a test image is strongly related to the virtual information present in the image and that the information can be quantified to measure the similarity between the test image and its reference image.

Drawbacks:

- Loss of Global Weighting. Predefined fixed weights are adopted to fuse the distances of different low-level visual features.
- Loss of Adaptive Weighting. adaptive weights for query images to fuse the distances of different low-level visual features. It is adopted by Bing Image Search.
- For our new approaches, two different ways of computing semantic signatures are compared.
- Not Visual: Query-specific visual semantic space using Reciprocal Hash Maps. For an image, a single semantic signature is computed from one SVM classifier trained by combining all types of visual features.

3. Proposed System

The proposed system Content-Based Image Retrieval (CBIR) uses D-SIFT algorithm the visual contents of an image such as color, shape, texture, and spatial layout to represent and index the image. Active research in CBIR is geared towards the development of methodologies for analyzing, interpreting cataloging and indexing image databases. In addition to their development, efforts are also being made to evaluate the performance of image retrieval systems.

The quality of response is heavily dependent on the choice of the method used to generate feature vectors and similarity measure for comparison of features. In this work we proposed an algorithm which incorporates the advantages of various other algorithms to improve the accuracy and performance of retrieval.

Halftone is the reprographic system that re-enacts nonstop tone symbolism through the utilization of specks shifting either in size or in separating subsequently creating a slope like effect. Halftone can likewise be utilized to allude particularly to the picture that is delivered by this process.

Where constant tone symbolism contains a vast scope of hues or greys the halftone process decreases visual multiplications to a picture that is printed with one and only shade of ink in spots of contrasting size (sufficiency balance) or dividing (recurrence balance). This proliferation depends on an essential optical figment the small halftone dabs are mixed into smooth tones by the human eye. At an infinitesimal level created high contrast photographic film additionally comprises of just two hues and not an unbounded scope of ceaseless tones.

The accuracy of color histogram based matching can be increased by using Color Coherence Vector (CCV) for successive refinement. The speed of shape based retrieval can be enhanced by considering approximate shape rather than the exact shape. In addition to this a combination of color and shape based retrieval is also included to improve the accuracy of the result.

Advantages:

- It Represent all of the descriptors of an image via sparse representation and assess the similarity between two images via sparse coding technique.
- The main advantage is, a feature descriptor is sparsely represented in terms of a Dictionary Score or transferred as a linear combination of Dictionary Score atoms, so as to achieve efficient feature representation and robust image similarity assessment.
- Best Results.
- High accuracy.
- High performance in search of related image re-ranking.

4. Module Description

1) Image Preprocessing and Feature Extraction

In the input module, the feature vector from the input image is extracted and that input image is stored in the image dataset.

The feature vector of each image in the dataset is also stored in the dataset whereas in the second module i.e. query module, a query image is inputted. After that the extraction of its feature vector is done.

During the third module i.e. in the process of retrieval, comparison is performed. The feature vector of the query image is compared with the each vector stored in the dataset.

The features which are widely used involve: texture, color, local shape and spatial information.

There is very high demand for searching image datasets of ever-growing size, This is reason why CBIR is becoming very popular.

2) D-Sift Feature Extraction for Reference and Test Images:

D-SIFT transforms image data into scale-invariant coordinates virtual to local features and generates large numbers of features that compactly cover the image over the full range of scales and locations. Shape is an important visual feature and it is one of the basic features used to describe image content. However, shape representation and description is a difficult task. This is because when a 3-D real world object is projected onto a 2-D image plane, one dimension of object information is lost. As a result, the shape extracted from the image only partially represents the projected object. To make the problem even more complex, shape is often corrupted with noise, defects, arbitrary distortion and occlusion. Further it is not known what is important in shape. Current approaches have both positive and negative attributes; computer graphics or mathematics use effective shape representation which is unusable in shape recognition and vice versa. In spite of this, it is possible to find features common to most shape description approaches. Basically, shape-based image retrieval consists of measuring the similarity between shapes represented by their features. Some simple geometric features can be used to describe shapes. Usually, the simple geometric features can only discriminate shapes with large differences; therefore, they are usually used as filters to eliminate false hits or combined with other shape descriptors to discriminate shapes

Each feature vectors are invariant to its geometrical variational versions and partially invariant to enlightenment changes and robust to geometric deformation**.**

3) Image Analysis:

In this module that have two functions as below

i)Scale-Space Extrema Detection

Searches over all scales and image locations. A difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation.

ii)Keypoint Localization

A key point has been found by comparing a pixel to its neighbors and is to perform a detailed fit to the nearby data for location, scale, and ratio of key curvatures. The low contrast points or poorly localized along an edges are removed by key point localization.

iii)Image Retrieval

The key points are transformed into a representation that allows for significant levels of local shape distortion and change in illumination.

The descriptor representation approach assessing the similarity between D-SIFT feature descriptors can be measured by matching their corresponding image by color, shape, size, texture and it will be displayed.

Algorithm for Color Retrieval
Step1: Read the image
Step2: Convert from RGB to HSV
Step3: Find HSV histogram and create vectors v1.
Step4: Read the vectors from database and compare one by one by one with vector v1.
Step5: Shortlist all the images which fall within the threshold.
Step6: find coherency of the query image for each color and create coherency vector c1.
Step7: Compare coherency vectors of all the short listed images from step5 with c1.
Step8: Store all matching images in results folder and also display them.

iv)Shape Retrieval

The proposed shape retrieval system based on the automatic segmentations process to get approximate information about the shape of an object. It begins by segmenting the image into 5 classes depending on their brightness. Then three attributes: Mass, Centroid and Dispersion for each class are calculated and stored as the shape vector. For retrieval the vectors of the query image and database images are compared and the most matching images are short listed as results.

Color Retrieval Color Retrieval System Works in Two Stages.

1) In the first stage, Histogram based comparison is done and matching images are short listed.
2) In the second stage, the Color Coherence Vectors of the short listed images (stage 1) are used to refine the results.
Numbers of coherent and non-coherent pixels for all color intensities are calculated in the image. Then size of coherency array, coherency array and no. of coherency pixels are stored as a vector.

D. Algorithm for shape Retrieval
Step1: read the image
Step2: convert it from RGB to grayscale
Step3: determine the range and number of classes.
Step4: calculate the number of pixels i.e. mass belonging to each class.
Step5: calculate the centroid and dispersion for each class.
Step6: compare centroid of each class of query image with the centroid of each class from database image and extract out
that class.
Step7: compare that class‟s mass and dispersion with respective class.
Step8: increase the count if it satisfies certain threshold.
Step9: consider second class and repeat steps 6-8 till all classes get over.
Step10: take another image from the database and repeat the comparison.
Step11: display the images with maximum count.

v)Size and Texture Similarity Measure

In this algorithm we propose that matching is done on color by color basis. By analyzing histograms, first calculate the number of colors in both query image and database image. Then both the images are matched by seeing if the proportions of a particular color in both the images are comparable. The image which satisfies most of the conditions is the best match. Retrieval result is not a single image but a list of images ranked by their similarities with the query image since CBIR is not based on exact matching.

If I is the database image and I‟ is the query image, then the similarity measure is computed as follows,
1. Calculate histogram vector vI = [vI1, vI2, ….vIn] and ccv vector cI = [cI1, cI2, ….cIn] of the database images.
2. Calculate the vectors vI‟ and cI‟ for the query image also.
3. The Euclidean distance between two feature vectors can then be used as the similarity measurement:
4. If $d \leq \tau$ (threshold) then the images match.
5. From all the matching images we display top 24 images as a result.
Segmenting the query image into 5 classes based on its brightness and calculates the Euclidean distance between the respective classes of query image and database image attributes. Mass, centroid and dispersion parameters are calculated for each class. These features are compared with database images stored features. The features values which are less than defined threshold are sorted based on increasing difference between query and database images then stored separately.

5. Experimental Setup

In this chapter, the reference image features were extracted by using Scale Invariant feature Transform. D-SIFT features are first extracted from a set of reference images and stored in a database. A new image is matched by individually comparing each feature from the new image to this previous database and finding candidate matching features based on Euclidean distance of their feature vectors. D-SIFT feature extraction first search over all scales and image locations.

The low contrast points are removed by key point localization and the stability is improved .Based on the local image gradient directions , one or more orientations are performed on image data that has been transformed relative to the assigned orientation ,scale, and location for each feature.
The technique for content-based image retrieval (CBIR) by exploiting the advantage of low-complexity ordered-dither block truncation coding using D-SIFT for the generation of image content descriptor. In the encoding step, our proposed work compresses an image block into corresponding quantizers and bitmap image. Two image features are proposed to index an image, namely, color co-occurrence feature (CCF) and bit pattern features (BPF), which are generated directly from the ODBTC encoded data streams without performing the decoding process. The CCF and BPF of an image are simply derived from the two Scale invariant quantizers and bitmap, respectively, by involving the visual codebook. Experimental results show that the proposed method is superior to the block truncation coding image retrieval systems and the other earlier methods, and thus prove that the proposed scheme is not only suited for image compression, because of its simplicity, but also offers a simple and effective descriptor to index images in CBIR system.

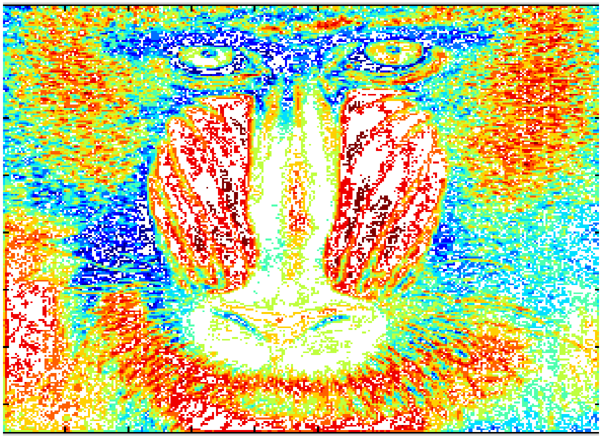

Figure.5.1. *D-SIFT Extrema detection*

In the keypoint descriptor generation, the key points are transformed into a representation that allows for significant levels of local shape distortion and change in illumination. A keypoint descriptor is created by first computing the gradient magnitude and direction at each image sample point in a region around the keypoint location. The key points are weighted by a Gaussian window, indicated by the overlaid circle and the samples are accumulated into orientation histograms summarizing the contents over 4x4 sub-regions with the length of each corresponding sum of the gradient magnitudes near that direction within the region.

Figure.5.2 *Descriptor generation*

To make the SIFT feature more compact, the bag-of-words(BoW) representation approach quantizes SIFT descriptors by vector quantization technique into a collection of visual words based on a pre-defined visual vocabulary or vocabulary tree. The vocabulary tree defines a hierarchical quantization that is built by hierarchical k-means clustering. A large set of representative descriptor vectors are used in the unsupervised training of the tree. The indexing descriptors extracted from local image region and is robust to background clutter. The local region descriptors are hierarchically quantized into a vocabulary tree. The vocabulary tree allows a larger and more discriminatory vocabulary to be used efficiently, which leads to a dramatic improvement in retrieval quality.

Figure.5.3. *BoW representation*

BoW representation approach quantizes SIFT feature descriptor by vector quantization into a collection of visual words based on the visual vocabulary tree. The BoW representation approach assessing the similarity between SIFT feature descriptors can be measured by matching their corresponding visual words by histogram matching.

6. Conclusion & Future Work

In the D-SIFT feature extraction, D-SIFT transforms image data into scale-invariant coordinates virtual to local features and generates large numbers of features that compactly cover the image over the full range of scales and locations. The low contrast points or poorly localized along an edges are removed by key point localization. A keypoint has been found by comparing a pixel to its neighbors and is to perform a detailed fit to the nearby data for location, scale, and ratio of key curvatures. To make the D-SIFT feature more compact, the bag-of-words (BoW) representation approach quantizes D-SIFT descriptors by vector quantization technique into a collection of visual words based on a pre-defined visual vocabulary or vocabulary tree. The future work focus to lead Color histogram and texture features based on a co-occurrence matrix are extracted to form feature vectors. Then the characteristics of the global color histogram, local color histogram and texture features are compared and analyzed for CBIR. Based on these works, a CBIR system is designed using color and texture fused features by constructing weights of feature vectors. This will helps for better feature extraction and fuzzy logic while gives better accuracy while matching image features. The relevant retrieval experiments show that the fused features retrieval brings better visual feeling than the single feature retrieval, which means better retrieval results.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Li-Wei Kang, Member, IEEE, Chao-Yung Hsu, Hung-Wei Chen, Chun- Shien Lu, Member, IEEE, Chih -Yang Lin, Member, IEEE, and Soo -Chang Pei, (2011). Feature-Based Sparse Representation for Image Similarity Assessment, IEEE Transactions on Multimedia, 13(5).

[2] Sivic J and Zisserman A, (2003). Video Google: A text retrieval approach to object matching in videos, Proc. IEEE Int. Conf. Computer Vision, Nice, France, 2; 1470–1477.

[3] C. Kim, Content-based image copy detection, *Signal Process.: Image Commun*., vol. 18, pp. 169–184, 2003

[4] Lowe D. G, (2004) Distinctive image features from scale-invariant keypoints, *Int. J. Compute*. Vision, vol. 60(2), 91–110.

[5] Maheshwari, R.U., B.Paulchamy, Pandey, B.K. et al(2024). Enhancing Sensing and Imaging Capabilities Through Surface Plasmon Resonance for Deepfake Image Detection. *Plasmonics*. https://doi.org/10.1007/s11468-024-02492-1

[6] Maheshwari, Uma, and Kalpanaka Silingam(2020). Multimodal Image Fusion in Biometric Authentication. *Fusion: Practice and Applications* 1(2); 7991.

[7] S. S, S. S and U. M. R(2022), Soft Computing based Brain Tumor Categorization with Machine Learning Techniques, *International Conference on Advanced Computing Technologies and Applications (ICACTA)*, Coimbatore, India, pp. 1-9, doi: 10.1109/ICACTA54488.2022.9752880.

[8] R. Uma Maheshwari, B. Paulchamy, Arun M, Vairaprakash Selvaraj, Dr. N. Naga Saranya and Dr . Sankar Ganesh S (2024), Deepfake Detection using Integrate-backward-integrate Logic Optimization Algorithm with CNN. *IJEER* 12(2), 696-710. DOI: 10.37391/IJEER.120248.

[9] Rajendran, U. M., & Paulchamy, J. (2021). Analysis and classification of gait characteristics. *Iconic Research and Engineering Journals*, 4(12).

[10] Paulchamy, B., Chidambaram, S., Jaya, J., & Maheshwari, R. U. (2021). Diagnosis of Retinal Disease Using Retinal Blood Vessel Extraction. *In International Conference on Mobile Computing and Sustainable Informatics: ICMCSI 2020*. pp. 343-359. Springer International Publishing.

[11] Boiman O., Shechtman E., and Irani M, (2008) In defense of nearest neighbor based image classification, in Proc. *IEEE Conf. Computer Vision and Pattern Recognition.*

[12] Yang J., Yu K., Gong Y., and Huang T., (2009) Linear spatial pyramid matching using sparse coding for image classification, in Proc. *IEEE Conf. Computer Vision and Pattern Recognition.*

[13] Hsu C. Y., Lu C.S and Pei S. C, (2009) “Secure and robust SIFT,” *in Proc. ACM Int. Conf. Multimedia*, pp. 637–640.

[14] Wright S. J., Nowak R.D and Figueiredo M. A. T, (2009) Sparse reconstruction by separable approximation, *IEEE Trans. Signal Process*., 57(7), 2479–2493,.

[15] Zhao W. L and Ngo C. W, (2009) Scale-rotation invariant pattern entropy for keypoint-based near-duplicate detection, *IEEE Trans. Image Process*., vol. 18(2), pp. 412–423.

[16] Chandrasekhar V., Makar M., Takacs G., Chen D. M., Tsai S. S., Grzeszczuk R and Girod B, (2010)

Survey of SIFT compression schemes, *in Proc. Int. Workshop Mobile Multimedia Processing, Turkey*.

[17] Yang A., Ganesh A., Ma Y and Sastry S, (2010) Fast l -minimization algorithms and an application in robust face recognition: A review, *in Proc. IEEE Int. Conf. Image Process., Hong Kong*.

[18] Kang L. W., Hsu C. Y., Chen H. W and Lu C. S, (2010) Secure SIFT based sparse representation for image copy detection and recognition, in *Proc. IEEE Int. Conf. Multimedia and Expo, Singapore*.

[19] Mairal J., Bach F., Ponce J and Sapiro G., (2010) Online learning for matrix factorization and sparse coding, *J. Mach. Learn. Res.*, vol. 11; 19–60.

[20] Hsu C. Y., Lu C. S and S. C. Pei S., C, (2011) Homomorphic encryption-based secure SIFT for privacy-preserving feature extraction, *in Proc. IS&T/ SPIE Media Watermarking, Forensics, and Security*.