Intelligent Image Indexing for Fast, Scalable, and Accurate Visual Information Retrieval

Abdelkrim Saouabe [1,2,*] and Said Tkatek [1]

[1] Computer Science Research Laboratory, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco
[2] Akkodis Research, Paris, France
Email: Abdelkrim.saouabe@uit.ac.ma (A.S.); said.tkatek@uit.ac.ma (S.T.)
*Corresponding author

***Abstract*—The efficient management and analysis of large-scale digital image collections have become critical challenges across various domains, including photographic archives, e-commerce platforms, and social networks. The effectiveness of image retrieval systems heavily depends on the choice of indexing methods, which influence both the speed and accuracy of visual information retrieval. Among the most prominent approaches, Content-Based Image Retrieval (CBIR) has emerged as a key technique, leveraging digital signatures derived from visual attributes such as color, shape, and texture. This study introduces an innovative CBIR-based automatic image indexing framework designed to enhance not only indexing automation but also retrieval accuracy and computational efficiency. The proposed approach has been rigorously evaluated on multiple image databases and benchmarked against state-of-the-art methods. Experimental results demonstrate significant improvements in retrieval speed, precision, and scalability, making this method a promising solution for real-world applications. By optimizing image management systems in terms of reliability and efficiency, this approach has the potential to transform diverse fields, including digital archiving, online advertising, and social media analytics. The adoption of this advanced indexing paradigm enables businesses and institutions to handle ever-growing volumes of digital images more effectively while ensuring seamless access to relevant visual information for end-users.**

***Keywords*—visual information retrieval, Content-Based Image Retrieval (CBIR), image processing, feature selection, vectorial similarity**

I. Introduction

With the explosion of visual data generated daily, image retrieval has become a major challenge in many fields. Traditional approaches, based on keywords associated with images, quickly show their limitations when it comes to accurately exploiting visual content. This shortcoming has led to the development of Content-Based Image Retrieval (CBIR) systems, which use visual features such as colors, shapes and textures to improve the relevance of results. These systems have a wide range of applications, including in medicine for image-assisted diagnosis, in e-commerce for visual product searches, and on social networks for content organization and recommendation. These advances have aroused growing interest among researchers, who are striving to explore the capabilities, uses and limits of these systems, to maximize their impact in diverse contexts. These systems, which enable images to be retrieved based on their visual features such as color, shape or texture, are at the heart of much research. Fig. 1 shows the classic architecture of the CBIR system.

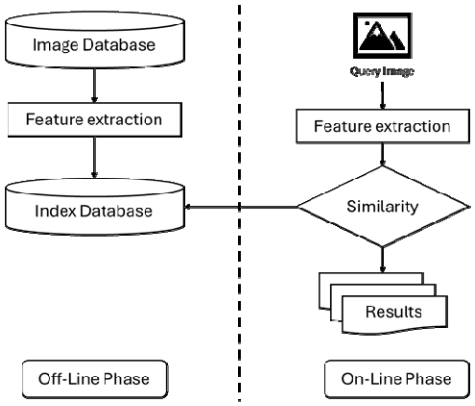

Fig. 1. CBIR system classic architecture.

The originality of our approach lies not only in the optimal synergy between these two phases but also in the distinctive hybrid strategy adopted. By combining local and global descriptors across the three fundamental visual dimensions color, texture, and shape our method captures both fine-grained pixel-level details and higher-level structural information, thereby yielding a richer and more discriminative representation of visual content. Moreover, unlike deep learning-based approaches, our solution is computationally lightweight, requiring neither high-end hardware nor large-scale annotated training datasets. This makes it particularly attractive for systems with limited resources or for large-scale applications where efficiency is critical. Finally, by transforming large tensors into compact vectors, our approach achieves significant dimensionality reduction, optimizing memory usage and enabling scalability to massive image databases.

In the literature, various aspects of CBIR systems have been explored and developed. For example, Ramamurthy

Manuscript received July 24, 2025; revised September 29, 2025; accepted October 11, 2025; published March 26, 2026.

and Chandran [1] focused on a CBIR system using shape and texture content for medical images, with the aim of assisting pathologists in the diagnostic process. In addition, Kumar and Singh [2] developed a CBIR system for eye image analysis, using the Local Binary Pattern (LBP) algorithm to extract texture and intensity features. Furthermore, Öztürk [3] introduced a content-based medical image retrieval method using a Siamese network, combined with deep feature selection to generate hash codes for images.

In the moving image domain, Paiz-Reyes *et al*. [4] proposed a cloud-hosted GIF image retrieval system. Another example of a cloud-based CBIR implementation is presented, where Mahmoudi *et al*. [5] introduce a cloud-based platform integrating multiple feature extraction algorithms tailored for CBIR systems. They also propose an efficient combination of Scale-Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) descriptors, which enhances the extraction and matching of image features, thereby improving the overall image retrieval process. To gain a complete understanding of the state-of-the-art in CBIR systems, an in-depth study was carried out and presented [6, 7]. This analysis enabled us to gain a better understanding of the work carried out by other researchers, while clearly identifying the advantages and disadvantages of each proposed solution. By drawing on this knowledge, we can contribute to the improvement of existing CBIR systems and propose new approaches adapted to current user needs.

In this article, we propose an innovative architecture grounded in the paradigm of CBIR, which constitutes the foundation of our research and experimental work. The architecture is structured around two complementary and interdependent phases. The first, referred to as the automatic indexing phase (offline), involves a sequence of processing steps designed to analyze images, extract the most relevant visual features, and generate representative digital signatures. This phase also integrates a discriminative feature selection mechanism, ensuring both robustness and reliability for subsequent retrieval tasks. The second, known as the online search phase, leverages these signatures to process user queries in real time. Through efficient comparison mechanisms and matching algorithms, it guarantees fast and accurate retrieval while offering a seamless interaction via a user-friendly interface tailored to the needs of end-users.

Together, these contributions position our architecture as a high-performance, resource-efficient, and scalable solution, offering a pragmatic yet innovative alternative to existing state-of-the-art CBIR methods.

The remainder of this article is organized as follows: Section II describes the automatic indexing (offline) phase, detailing the processes and algorithms used to analyze images, extract features, and generate digital signatures for later retrieval. It also introduces the search (online) phase, where these signatures support real-time image queries through comparison mechanisms, matching algorithms, and user interfaces. Section III evaluates system performance, analyzing mean Average Precision (mAP), search speed, and resource efficiency, and comparing results with state-of-the-art approaches. Finally, Section IV concludes the article by summarizing our findings, discussing their implications, and outlining directions for future research.

II. CBIR-Hybrid Architecture (Our Approache)

CBIR-Hybrid (CBIRH) architecture, based on the traditional CBIR model, features a major innovation: the integration of local and global descriptors for automatic image indexing. This hybrid approach captures both the fine details and global characteristics of images, offering a more complete and accurate representation of visual content. Local descriptors focus on specific details within images, such as patterns or textures present in particular regions [8]. For example, methods such as SIFT are used to identify points of interest or fine textures [5]. In contrast, global descriptors examine the image, capturing more general information such as color distribution or dominant color [9]. Combining these two types of descriptors create digital vectors that reflect visual information in each image in a richer, more balanced way. Fig. 2 provides an overview of CBIRH architecture, which is divided into two phases that will be described in greater detail later. It illustrates the various system components and their interactions, providing a visual representation that facilitates understanding of the overall process and the proposed innovations.

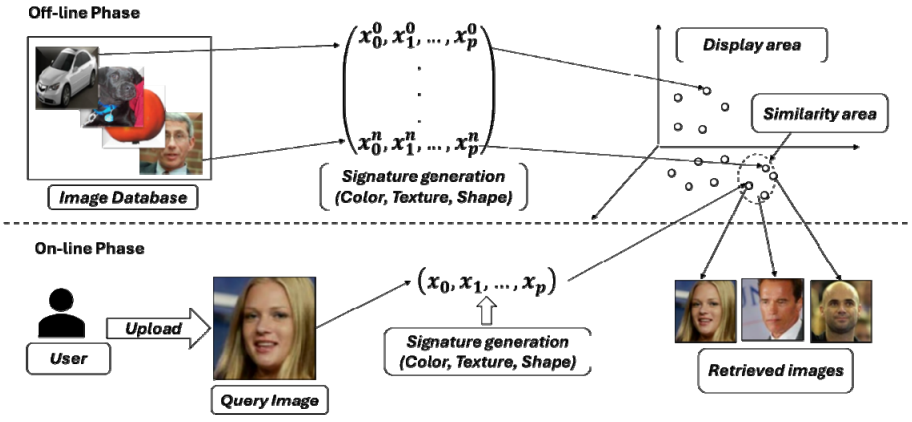

Fig. 2. CBIRH architecture.

A. The Offline Phase or Automatic Indexing Phase

The offline phase or the automatic indexing phase is an essential stage in the operation of CBIR systems. During this phase, the system operates autonomously, without direct interaction with the user. The main objective is to prepare images for future searches by creating efficient numerical representations, known as feature vectors, that enable images to be identified and compared according to their visual attributes. This phase is crucial, as it lays the foundations for the CBIR system's performance in the online search phase. Automatic indexing takes place in the background, which means it can be carried out without real-time constraints, enabling more intensive and detailed image processing. The process takes place in two distinct stages:

1) Image dataset

For our study, we have chosen to use the Pattern Analysis, Statistical modelling and Computational Learning (PASCAL) Visual Object Classes (VOC) dataset, which is recognized as an essential reference in the field of computer vision. This dataset is widely used for various tasks such as object detection [10], semantic segmentation [11] and image classification [12]. It was developed as part of the PASCAL VOC challenge, a series of competitions orchestrated by the PASCAL project to evaluate the performance of computer vision algorithms [13]. The PASCAL VOC 2012 version is particularly widespread and features a wealth of annotations, including labeled images belonging to twenty common object classes. These classes cover a wide range of object types, from vehicles and animals to everyday objects. The annotations provided in this dataset include bounding boxes, pixel-level segmentation masks and class annotations that facilitate detailed analysis of the objects present in each image. Its diversity and the precision of its annotations make it a valuable tool for researchers and engineers specializing in visual recognition.

To simplify the stages of our study and facilitate analysis, we have selected five specific classes from the PASCAL VOC database: car, fruit, person, cat and dog. For each class, we have chosen 100 annotated images, which will enable us to carry out a targeted analysis while ensuring adequate representation of the various objects under study. This approach will help us draw accurate conclusions about the performance of the detection and classification models we develop.

2) Feature extraction

The automatic indexing phase is crucial to the operation of CBIR systems. It enables the visual information in images to be represented in digital form by descriptors, which encapsulate and structure the data [14]. This step is not limited to extracting visual attributes but also aims to organize this information into digital signatures to facilitate subsequent searches. The quality of indexing depends on the effectiveness of the system in identifying the essential characteristics of images [15]. These characteristics include color, texture and shape, each offering a unique perspective on the visual content. The study will detail the descriptors used to improve image indexing. The Table I summarizing the descriptors selected for this study, organized according to their type of feature. This choice is based on an in-depth analysis of several state-of-the-art descriptors [6]. The descriptors selected were carefully chosen according to the specific requirements of the analysis, with the aim of optimizing both the accuracy and efficiency of the image retrieval system.

TABLE I. TABLE OF DESCRIPTORS SELECTED FOR THIS STUDY

Color descriptors	Texture descriptors	Shape descriptors
Color histogram	Texture histogram	SIFT
Dominant color	Co-occurrence matrix	Hu moments
	Gabor filters	

a) Color visual feature descriptors

Color is often the first criterion used for image retrieval, playing an essential role in the identification and classification of visual content. Although the Red-Green-Blue (RGB) color space is widely used, other spaces such as Hue-Saturation-Value (HSV) or the CIELab and CIELuv spaces offer better correspondence with human perception [16]. The latter offer a more intuitive representation of colors, making them increasingly popular in image processing systems. Choosing the right color space is crucial, as it depends on the specific data and algorithms used [17]. Indeed, color representation can have a considerable influence on the performance of an image retrieval system [18].

Various descriptors are available to characterize image colors. Among these, color histograms and dominant colors are particularly effective methods for estimating the color density of an image. Color histograms represent the distribution of colors in an image by counting the number of pixels for each color in a given color space, usually RGB or HSV [19]. This method is easy to understand and implement, while offering a reliable estimate of color density. What's more, it allows rapid comparison between images, making it very practical for image retrieval. On the other hand, the dominant color descriptor can be used to identify the most representative hues in an image, frequently using clustering algorithms to group similar colors [9]. In our study, we opted for the *K*-means machine learning model with a K of 3, after applying the kink method to determine the optimal number of clusters. This approach simplifies the analysis by focusing on the most significant colors, which is particularly relevant for facilitating the search of images according to specific tonalities or moods, this descriptor improves the user experience in image retrieval systems. Fig. 3 shows some examples of the results obtained by our dominant color algorithm on images from the database.

After extracting three dominant colors from each image in our database, we proceed to generate vectors for each image using the Eq. (1):

$$F = \{(C_i, P_i)\} \; avec, \quad i = 1,2,3 \tag{1}$$

where C_i is a vector representing the dominant colors, assigned to the percentages P_i of pixels in the image.

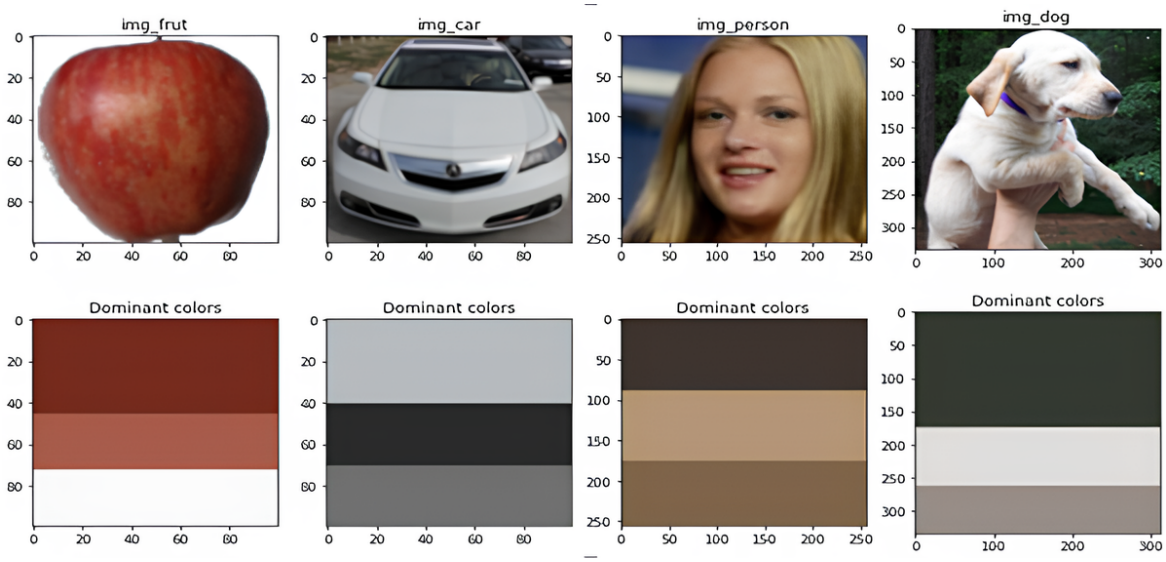

Fig. 3. Identifying dominant colors in different image classes.

b) Texture visual feature descriptors

Texture is often modeled as a spatial structure resulting from the organization of primitives, each with a random appearance. It can be either periodic or random in appearance [18]. Due to the imprecise definition of texture, the measures associated with it display an even greater diversity than those for colors. These measures aim to capture descriptors of the image or its parts, considering variations along a specific direction and scale. They are particularly useful for analyzing homogeneously textured regions. What's more, texture measurements are invariant to image rotation and scale changes [16]. Texture can be characterized by various attributes, such as pattern regularity, contrast and periodicity. In the context of content-based search, texture can be used to distinguish areas of similar color, but different meanings [14]. In our study, we rely on three texture descriptors.

Texture histograms are essential in image analysis, as they quantify and describe the textural patterns present. By illustrating the distribution of gray levels or intensities, they highlight important characteristics such as the regularity, contrast and complexity of textures [20]. This digital representation not only enables different images to be compared, but also textural similarities and differences to be identified, making it easier to integrate into content-based image retrieval systems. Fig. 4 shows an example of how texture histograms can be applied to some of the images in our database.

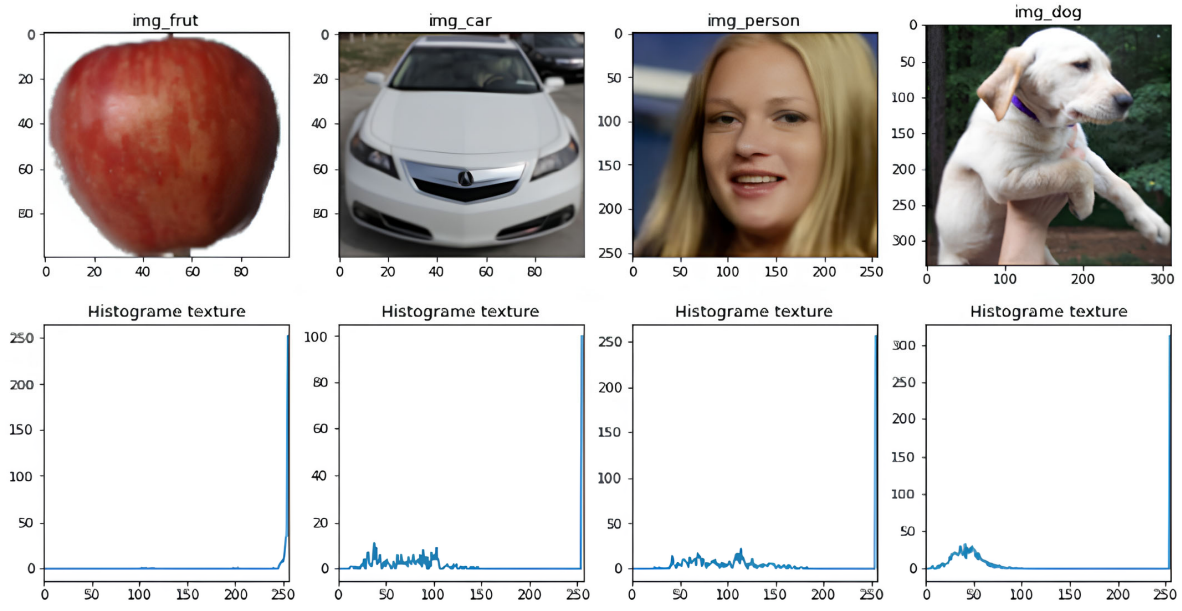

Fig. 4. Applying texture histograms to database images.

The Grey Level Co-occurrence Matrix (GLCM) is another powerful tool for texture analysis [21]. It captures the relationship between pixels based on their intensity levels and relative positions, by counting the frequency of pixel pairs in given directions and distances. By deriving characteristics such as homogeneity and contrastivity, it enables in-depth texture analysis, essential for applications such as segmentation and pattern recognition [22].

The indices of this matrix correspond to the gray levels of the analyzed texture. The matrix is defined such that $P_{d,\theta}(i,j)$ represents the frequency with which a pair of points, separated by a distance d in a directionθ, have the gray levels I_i, I_j respectively. To ensure accurate relative frequencies, the elements of the matrix must be normalized by dividing them by the total number of point pairs separated by the distance d in the directionθ across the entire image [18]. After applying the co-occurrence matrix to the images in our database, we calculated the following values to generate the feature vector associated with each image.

$$Energy = \sum_{i=1}^{M}\sum_{j=1}^{N}\left(P_{d,\theta}(i,j)^2\right) \tag{2}$$

$$Homogeneity = \sum_{i=1}^{M}\sum_{j=1}^{N}\left(\frac{P_{d,\theta}(i,j)}{1+(i-j)^2}\right) \tag{3}$$

$$Entropy = \sum_{i=1}^{M}\sum_{j=1}^{N}\left(\log P_{d,\theta}(i,j)P_{d,\theta}(i,j)\right) \tag{4}$$

$$Contrast = \sum_{i=1}^{M}\sum_{j=1}^{N}\left((i-j)^2 P_{d,\theta}(i,j)\right) \tag{5}$$

Gabor filters, meanwhile, are sophisticated instruments that extract textural features by mimicking the response of neurons in the human visual cortex. These filters can detect patterns in different orientations and frequencies, offering a multiscale representation that facilitates the identification of varied textures [23]. Their robustness to scale and orientation transformations, combined with their efficiency in isolating relevant features, makes them particularly useful in image classification [22].

In Fig. 5, we present an illustration of the convolutional product between a Gabor filter in Fig. 6 and an image from our database. After applying the filter, we proceed to calculate key statistical measures, including the energy, mean, and standard deviation, which are then used to generate the feature vector assigned to the image. These computed values capture the textural characteristics and contribute to the overall texture-based analysis. In summary, these textural analysis tools—texture histograms, co-occurrence matrices and Gabor filters—provide rich and varied descriptors, enabling in-depth understanding of the patterns present in images. Their integration into image processing systems enhances the accuracy and relevance of results, making their use indispensable in a variety of application fields.

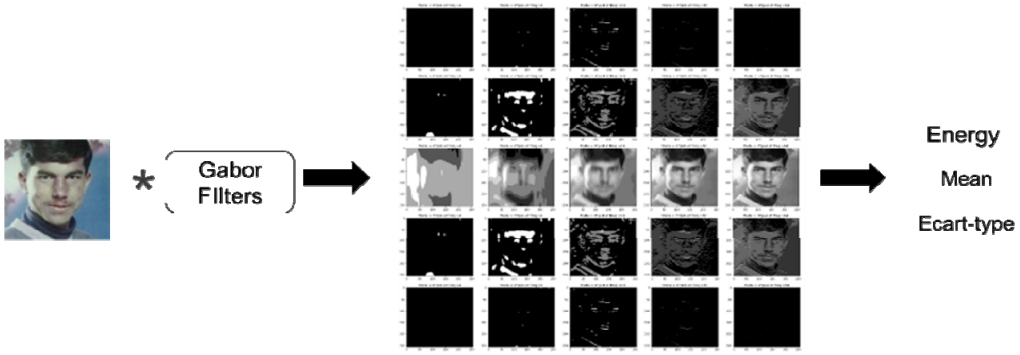

Fig. 5. Implementation of Gabor filters on a sample image.

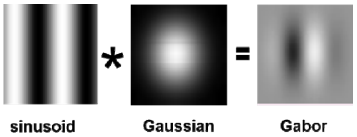

Fig. 6. Gabor filters: Product of a sine wave and a Gaussian function.

c) Shape visual feature descriptors

An effective shape descriptor must be invariant to rotation, translation and scale, to ensure that the features extracted from an image remain relevant, whatever geometric transformations it undergoes [6]. These invariances are essential for applications such as object recognition or image retrieval, where shapes may appear at different angles, sizes or positions in the image. In our study, we have chosen to use two descriptors that share this capacity for invariance to geometric transformations. Firstly, we use SIFT, a particularly powerful method for extracting points of interest, as it retains invariance to scale and rotation [5]. This property is crucial for identifying images captured from a variety of camera angles, where objects may appear larger or smaller, or rotated relative to their original position [18]. It can detect minutiae and extracts robust descriptors, enabling reliable matching between images. SIFT, proposed by Lowe [24], is an algorithm that detects points of interest in an image and extracts distinctive features for object

recognition. This descriptor is represented by a 128-dimensional vector, computed from the orientations and moduli of the gradients about a feature point. The neighborhood is divided into 16 4×4-pixel regions, for which histograms of gradient orientations are created. Each 4×4-pixel block produces an 8-class histogram, representing the main orientations between 0 and 360 degrees. Together, these histograms form the SIFT descriptor, a vector of 128 coordinates (4×4×8). This process takes place in four stages, from the detection of points of interest to feature extraction [18].

Detection of scale-space extrema: The image is convolved with a Gaussian kernel. The scale-space of an image is therefore defined by the function:

$$L(x,y,\sigma) = G(x,y,\sigma) \times I(x,y) \tag{6}$$

where $I(x,y)$ is the original image.

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{7}$$

With one normalization, this is equivalent to solving:

$$\frac{\partial y}{\partial x} = \Delta I \tag{8}$$

where ΔI represents the Laplacian of I.

The pre-selection of points of interest and their scale is made by detecting the local extrema of Gaussian differences:

$$D(x,y,k\sigma) = \big(G(x,y,k\sigma) - G(x,y,\sigma)\big) \times I(x,y) \tag{9}$$

$$D(x,y,k\sigma) = L(x,y,k\sigma) - L(x,y,\sigma) \tag{10}$$

where $D(x, y, \sigma) \approx (k-1)\Delta I$, when $k \to 1$.

The extrema are sought in small neighborhoods in position and scale (typically 3×3×3). An interpolation step aims to improve the location of points of interest in space and scale. Then, an analysis of the eigenvalue ratios of the "2×2". Hessian matrix is used to eliminate points of interest located in areas of insufficient contrast or on edges with too little curvature.

Choice of descriptor orientation: This step involves assigning an orientation to each point. This orientation corresponds to the majority orientation of the spatial intensity gradients calculated in the vicinity of the point of interest at the scale previously determined. A point of interest can be assigned several orientations. This results in a redundancy of descriptors.

Calculation of descriptors: Finally, for a given position, scale and orientation, each point of interest is assigned a descriptor. For each image, the norm of the spatial gradient $m(x,y)$ and the orientation of the spatial gradient (x,y) at that scale are computed, as shown in Eqs. (11) and (12).

$$m(x,y) = \sqrt{\big(L(x+1,y) - L(x-1,y)\big)^2 + \big(L(x,y+1) - L(x,y-1)\big)^2} \tag{11}$$

$$\theta(x,y) = \tan^{-1}\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right) \tag{12}$$

Applying the SIFT method to the images generates a set of points of interest, each represented by a 128-dimensional vector. These vectors capture distinctive features, facilitating in-depth image analysis. However, each image may contain a variable number of points of interest depending on its visual complexity, resulting in a significant increase in data volume when the method is applied to many images. As shown in Fig. 7.

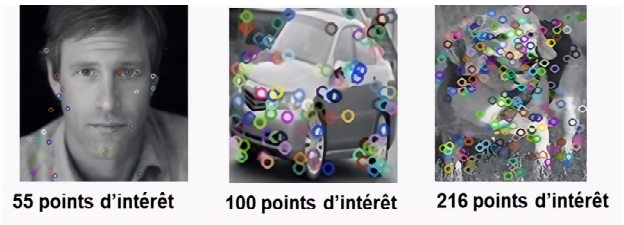

Fig. 7. Application of SIFT to three images with a variable number of points of interest.

This explosion of feature vectors complicates subsequent analysis steps, such as data storage, processing and classification. Although SIFT is a powerful tool for feature extraction, it is essential to implement effective strategies to manage this mass of data and guarantee reliable results. To this end, we have chosen to apply a clustering algorithm based on the K-means machine learning method, with $K = 3$, after using the kink method to determine the optimal number of clusters. This extracts the three most significant points of interest per image. By concatenating these three vectors, we obtain a 384-dimensional vector (128×3) per image, which simplifies analysis while preserving essential information.

In addition to SIFT, we also use Hu moments, another powerful descriptor which, like SIFT, is invariant to rotation, translation and scale. Hu moments are mathematical functions derived from the geometric moments of an image, and they capture global information about the shape of an object [9]. This approach makes it possible to recognize shapes, even when the object is moved, enlarged or oriented differently. Here are the seven mathematical equations that describe Hu's moments as shown in Eqs. (13)–(19).

$$M_1 = \mu_{20} + \mu_{02} \tag{13}$$

$$M_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \tag{14}$$

$$M_3 = (\mu_{30} - 3\mu_{21})^2 + (3\mu_{21} - 3\mu_{03})^2 \quad (15)$$

$$M_4 = (\mu_{30} + \mu_{21})^2 + (\mu_{21} + \mu_{03})^2 \quad (16)$$

$$M_5 = (\mu_{30} - 3\mu_{21})(\mu_{30} + \mu_{12})\left((\mu_{30} + \mu_{21})^2 - 3(\mu_{21} + \mu_{03})^2\right) + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})\left(3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2\right) \quad (17)$$

$$M_6 = (\mu_{20} - \mu_{02})\left((\mu_{30} + \mu_{21})^2 - (\mu_{21} + \mu_{03})^2\right) + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{03} + \mu_{21}) \quad (18)$$

$$M_7 = (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})\left((\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})\right) - (\mu_{30} - 3\mu_{21})(\mu_{12} + \mu_{03})\left(3(\mu_{30} + \mu_{12})^2 - (\mu_{12} + \mu_{03})^2\right) \quad (19)$$

Table II illustrates the application of the Hu moments descriptor to a fruit image from the database, before and after a 90° rotation. The results confirm that the moment values remain unchanged despite the applied transformation, demonstrating the rotation invariance of this descriptor. This property is essential for pattern recognition tasks, as it enables an object to be correctly identified regardless of its orientation in the image.

TABLE II. APPLYING TO HU'S MOMENTS TO A 90° ROTATED IMAGE

Image	M_1	M_2	M_3	M_4	M_5	M_6	M_7
	3.099344	8.660422	12.073033	11.569857	26.873674	17.250029	−23.416231
	3.099363	8.660905	12.075507	11.569796	26.838309	17.086204	−23.422268

By combining these two descriptors, we enhance the robustness of our feature extraction model, ensuring that extracted shapes remain consistent and reliable across a wide range of observation conditions.

3) Feature selection

In this section, we focus on feature selection after concatenating all the features extracted from the previously used descriptors. We have combined the descriptor vectors into a single vector for each image, resulting in a vector size of 511 values. This significant increase in feature vector size complicates subsequent analysis steps, such as data storage, processing and classification. To manage this mass of data and guarantee reliable results, it is essential to implement an efficient features selection method. To this end, we compared two selection methods: Principal Component Analysis (PCA) and a machine learning model, which identifies important features.

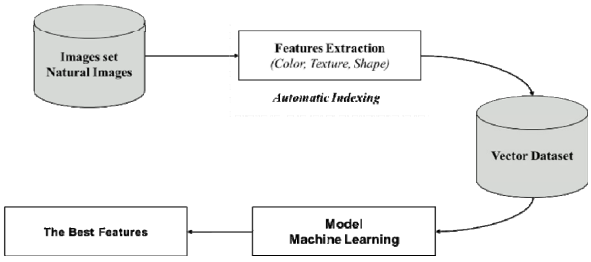

Fig. 8. Steps in the optimal characteristics' selection process.

To ensure a reliable and efficient comparison, we established a structured approach, which we illustrate in Fig. 8. This approach will enable us to determine the most appropriate method for feature selection, thus facilitating subsequent analysis steps.

a) Principal Component Analysis (PCA) method

PCA is a statistical method used to simplify and explore complex data sets [25]. In our study, it is applied to images as individuals and to visual descriptors such as color, texture and shape as variables. PCA will enable us to analyze the linear correlations between these descriptors, to better understand the relationships that exist between them. For example, PCA can reveal strong associations between certain color and texture combinations or show that certain shapes are recurrent in images with a specific color distribution. The main aim of PCA is to reduce the dimensionality of the data while retaining most of its variance, enabling the identification of principal components and revealing significant interactions between different visual features in images. Another key advantage of PCA is the ability to visualize results graphically. A particularly useful tool in this respect is the eigenvalue screen, which helps determine the optimal number of principal components to retain.

Fig. 9 illustrates the eigenvalue scree plot, where each principal component is represented by its corresponding eigenvalue, representing the proportion of variance or percentage of inertia explained by that component. The higher the eigenvalue, the greater the proportion of total data variance captured by the principal component. In our study, we have 16 features with eigenvalues ranging from 1% to 24%.

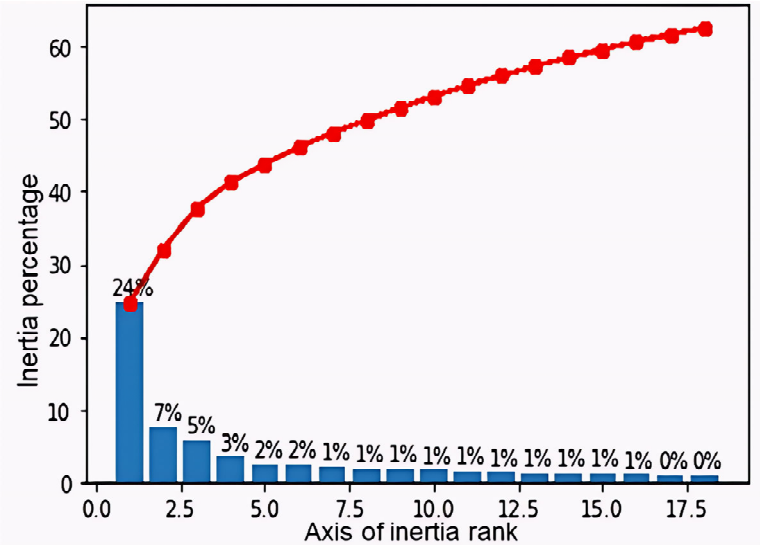

Fig. 9. Eigenvalue collapse.

By plotting these eigenvalues against their rank, the scree reveals the "fall" or decrease in variance explained by each component. This makes it possible to identify the point at which the decrease in eigenvalues becomes less pronounced, generally referred to as the "kink", as indicated by the red curve. This point is crucial for selecting the number of principal components to retain, ensuring that the analysis retains a significant amount of information while simplifying the complexity of the data.

Following the application of PCA on our feature vectors, we have reduced their dimension from 511 to 16 values. This reduction facilitates model simplification while preserving the essential information required for analysis. We then evaluated the performance of these simplified vectors using a machine learning model, specifically the random forest model. The model was trained on 75% of the data, while the remaining 25% was used to test its ability to generalize on new data. To assess the model's performance, we chose two main metrics: the confusion matrix and accuracy. The confusion matrix is a visual tool for evaluating the performance of a classification model, providing an overview of the errors made, as well as the types of confusion, such as false positives and false negatives. As for accuracy, this is calculated as the ratio of correct predictions to all positive predictions, thus measuring the proportion of true positive predictions among all positive predictions made by the model.

b) Important features method

Random forest is a highly effective machine learning algorithm, suitable for both classification and regression tasks. By constructing an ensemble of decision trees, it enhances prediction accuracy and effectively manages noisy or incomplete data. One of its most valuable features is its ability to assess the significance of various attributes in a dataset. In our approach, Random Forest's feature importance will serve as the second method for selecting relevant features.

Important Features (IF) in random forest is typically measured using a metric known as Gini importance. This metric evaluates the total reduction in Gini impurity each time a specific feature is used to split the data. A higher Gini importance indicates that the feature plays a more critical role in the model's performance.

After training the random forest model on our database containing feature vectors, we used the feature importance calculation functionality to assess the contribution of each attribute in the prediction process. This functionality, integrated into the random forest model, enables us to measure the influence of each feature in terms of accuracy or impact on the decisions made by the model's trees. The feature importance calculation assigns a percentage of accuracy to each attribute, highlighting those that play a major role in the model's overall performance. This analysis allows us not only to identify the most relevant features, but also to better understand the structure of the data. To illustrate these results, we have generated a bar chart that visualizes the relative importance of each feature. This chart provides a clear overview of the most influential attributes, facilitating interpretation of the results and enabling informed decisions to be made about the selection or elimination of certain features in later stages of the project.

The resulting plot in Fig. 10 will display the features ranked by their importance in descending order. As mentioned earlier, this method enabled us to rank features in descending order of importance. This means that each attribute was evaluated according to its contribution to the overall performance of the model, with the most influential features being placed at the top of the ranking. To determine an optimal threshold of importance at which we could select the most relevant features, we applied the elbow method [26]. This technique involves plotting the importance of features and identifying the point at which a further increase in importance no longer brings a significant gain in terms of performance. In our case, the optimal threshold was set at an important value equal to 0.00263.

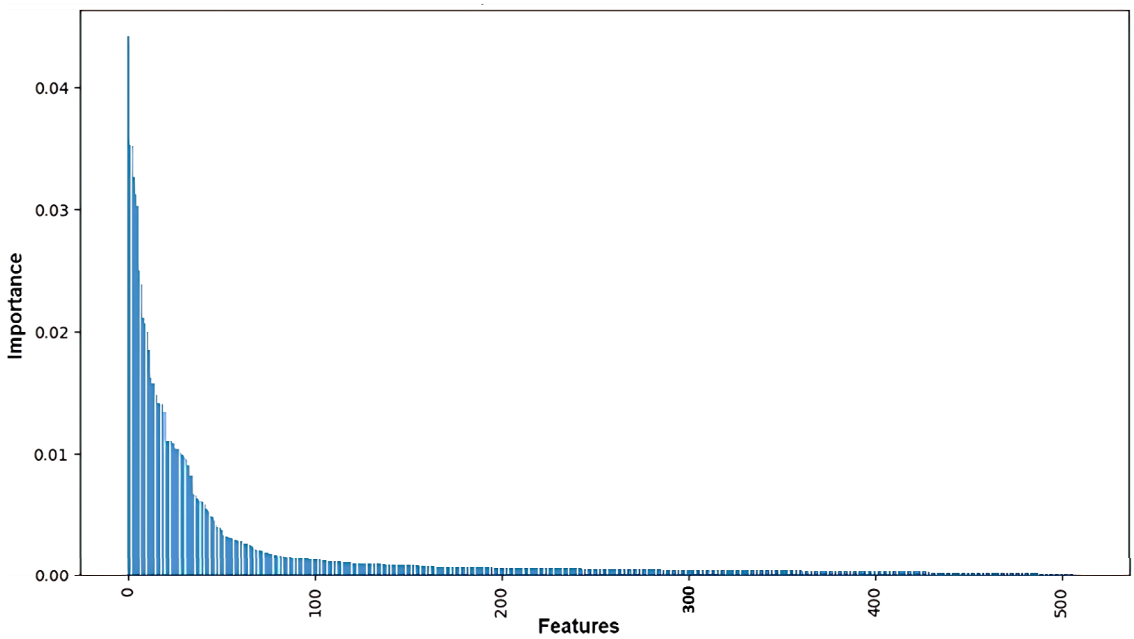

Fig. 10. Importance of all features organized in descending order.

Using this threshold, we were able to select 65 features considered most relevant to our model. These attributes contribute significantly to predictions, while the others, whose importance is below the defined threshold, are deemed less useful or redundant for the task in hand. This step made it possible to reduce the dimensionality of the data, from 511 to 65 attributes, while preserving the essential information, thus optimizing model performance and reducing computational complexity. Fig. 11 shows the features ranked in descending order of importance after selection.

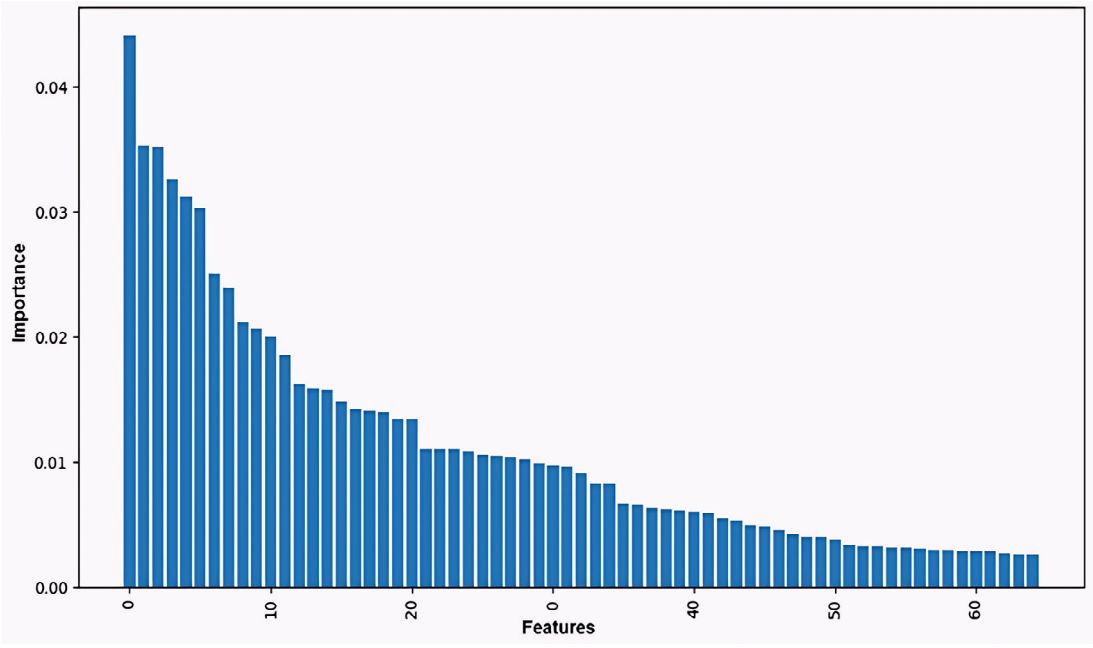

Fig. 11. Importance of features after threshold filtering, ranked in descending order.

c) Comparison of feature selection methods

In this section, we evaluate the performance of the random forest model on three separate datasets. The first set consists of the initial vector database, containing 511 features. The second set was reduced using the PCA method, which reduces the dimensionality of the data while retaining the essential information. Finally, the third data set is the result of the last feature selection method we applied, which identified the most relevant features according to their importance for the method (IF). To compare the model's performance on each of these datasets, we used four evaluation metrics:

- Accuracy (overall precision) measures the proportion of correct predictions out of all predictions made.

- Precision assesses the model's ability to limit false positives by measuring the ratio of correct predictions among those marked as positive.
- Recall indicates the model's ability to capture true positives among all positive samples, in other words, its sensitivity.
- F1-Score, a synthetic metric combining precision and recall in a single indicator, is particularly useful in cases of class imbalance.

Evaluating the model on these databases not only allows us to understand the impact of dimension reduction and feature selection on overall performance, but also to identify the feature selection method that maximizes model efficiency while maintaining good generalization capability. Table III summarizing the results obtained for each feature selection method and on each of the metrics mentioned.

TABLE III. EVALUATION RESULTS FOR ALL DATASETS BY METRICS

Databases	Number of features	Accuracy	Precision	Recall	F1-Score
Initial database	511	96.36%	79.98%	80.9%	80.01%
Features Selection by PCA	16	84.38%	59.86%	61.66%	60.26%
Features Selection by IF	65	92.96%	82.32%	82.52%	82.14%

The interpretation of the results reveals important insights into the impact of feature reduction on the performance of the Random Forest model. When the model is evaluated on the initial database with 511 features, it achieves a very high accuracy of 96.36%. This shows that it is capable of correctly predicting many cases. However, while this overall performance is impressive, precision 79.98% and recall 80.9% indicate that there is a certain balance to be maintained between correct predictions and errors, particularly with false positives and true positives. The F1-Score 80.01% reflects this balanced performance but suggests that the model could be refined to improve the accuracy of results.

On the other hand, when the database is reduced to 65 features by selecting the most important ones, accuracy decreases slightly to 92.96%. This reduction may be due to the loss of some relevant information contained in the deleted features. However, despite this moderate drop in accuracy, precision 82.32%, recall 82.52% and F1-Score 82.14% all increased compared with the initial base. This indicates that, although the overall model is slightly less accurate, it is now more efficient in detecting positive cases and makes fewer errors when identifying cases as positive. The increase in F1-Score underlines a better balance between the model's ability to limit false positives and capture true positive cases.

After comparing the results, we were able to confirm that the method (IF) feature selection method represents the best compromise in terms of performance and complexity. The 65 features were shown to be sufficiently informative to enable the model to maintain high performance while reducing the dimensionality of the data. This also improves the computational efficiency of the model, making it faster and easier to train. The 65 features we have extracted from our data analysis will play a crucial role in becoming the true "signatures" of our image database. These signatures, which bring together the essential information needed to describe each image, will form the basis of our scientific contribution to optimizing image retrieval systems by exploiting them as descriptors. We aim to enrich our Content-Based Image Retrieval-Hybrid (CBIRH) approach, by integrating these significant features. Our aim is to offer a more powerful solution for image retrieval, facilitating smarter, more tailored searches. This breakthrough has the potential to transform image retrieval applications, particularly in fields such as e-commerce, medicine and visual archives.

4) Similarty method

The similarity between the two images is established by measuring the distance between the vectors of the 65 extracted features. The notion of similarity is based on the principle that the smaller the distance between the characteristics of two images, the more similar the images are perceived to be. Conversely, a larger distance indicates a greater difference between the images. This means that the system favors images whose characteristics are closest to those of the query image, enabling the user to obtain relevant, contextualized results.

In the literature, the chi-squared method has proven to be particularly well-suited for computer vision applications due to its ability to accurately measure the dissimilarity between two feature distributions [27]. Unlike other distance metrics, the chi-squared method accounts for the relative importance of each feature, allowing for a more refined evaluation of the differences between images. This approach is especially relevant in systems where certain features play a more significant role than others in the comparison process. Furthermore, this measure not only improves the efficiency of image ranking but also better captures the visual nuances between images. As a result, users benefit from a more precise search experience, where the results are ranked in a more intuitive way, aligning with visual expectations. This enhances the relevance of the images returned in response to a query. The chi-squared distance is calculated using the following formula:

$$D_{X^2}(P,Q) = \sum \frac{(P_i - Q_i)^2}{P_i - Q_i} \tag{20}$$

where P and Q represent the two vectors of 65 values, and P_i and Q_i are the corresponding values of the two vectors.

B. The Online Phase or Search Phase

In this online phase of the search system, the user uploads an image to be analyzed by the system. The system examines the characteristics of the target image and compares them with those of all the images available in the database. The result is an ordered list of images, sorted according to their similarity to the query image. We have now assembled all the components needed to complete our CBIRH architecture. This architecture is based on two fundamental elements: the image signature generation tool, based on the 65 extracted features, and the method for calculating the distance between these signatures, enabling us to measure the similarity between images. These two components form the heart of the system, guaranteeing efficient and precise image searches based on visual content.

1) Test databases

To evaluate the performance of CBIRH architecture on widely recognized image datasets that are commonly used in both the literature and the field of image retrieval. Testing on these well-established datasets is crucial for assessing the effectiveness and robustness of the proposed approach. It allows us to compare our method with other state-of-the-art techniques under standardized conditions, ensuring the validity and relevance of the results. To evaluate the effectiveness of our system, we are using a variety of databases, including [7]:

- Coil 100 and Coil 10K, offering a wide range of objects and lighting conditions,
- Corel-1K, well-known for its complexity and visual diversity,
- ZUBUD, a set of building images,
- GHM, a rich database used for content-based image retrieval.

These databases will enable us to test our model in a variety of contexts, covering different types of objects, scenes and visual conditions. By evaluating the ability of our architecture to retrieve similar images from these datasets, we will be able to measure performance indicators such as accuracy and speed of execution. These results will form the basis for judging the robustness and efficiency of our CBIRH solution and will guide us towards possible optimizations.

2) Evaluation method and metrics

To evaluate the CBIRH architecture, we implemented a rigorous methodology to formalize the test process. As part of this approach, we have selected 20 separate images, representing different objects, of varying quality and from different contexts. The aim is to carry out 20 tests, with each image serving as a query during the online phase. Each query image will undergo specific processing (see Fig. 2) to extract a unique signature. This signature will then be compared with the other signatures stored in the vector database, using the Chi-2 distance calculation. Once the distances between the query vector and the target vectors have been calculated, the system sorts the vectors in ascending order of similarity. Finally, for each query image, we will limit the number of images retrieved to 20, ensuring that only the most relevant images closest to the initial query are presented to the user. This approach will enable us not only to evaluate the performance of the CBIRH architecture, but also to identify its strengths and weaknesses in handling complex queries.

To evaluate the performance of CBIRH architecture, we have chosen to use two key metrics: accuracy and execution time. These two criteria offer a comprehensive and balanced assessment of our system's capabilities in image retrieval. Accuracy is a key measure of efficiency of an image search system. It is calculated as the number of relevant results returned by the system in relation to the total number of results returned. In other words, it indicates the extent to which the images retrieved meet the user's expectations. Execution time is another important metric, measuring the speed with which the system processes a query and returns results. In image retrieval, users expect answers quickly, making this metric more critical. We timed the time required for each step in the process, including processing the query image, calculating distances, sorting results, and final image retrieval. Low execution time is essential for a smooth and responsive user experience.

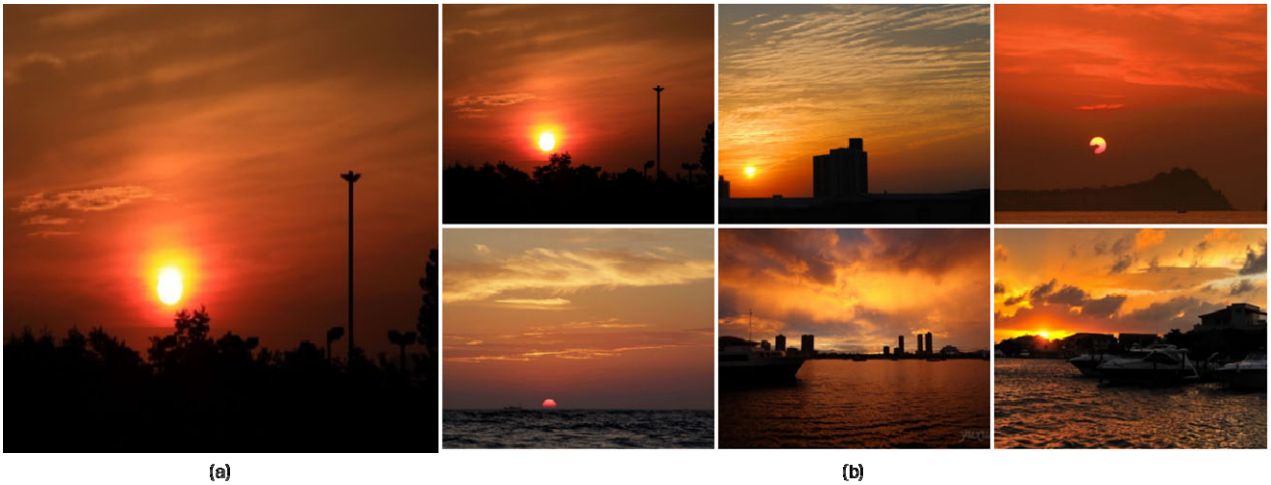

Fig. 12. CBIRH illustration: (a) query image, (b) similar images retrieved for a sunset.

Fig. 13. Example of CBIRH search: (a) query image, (b) images automatically extracted by the system for a mountain range.

1) Retriaval results and discussions

To illustrate the performance of our approach in concrete terms, we tested the CBIRH system on several representative visual queries. In the first case in Fig. 12, the query consists of an image of a sunset. The images returned by the system all display a striking visual correspondence, faithfully reproducing the warm color palette, ranging from bright yellow to intense orange-red. This result highlights the CBIRH's ability to capture subtle characteristics related to color and lighting variations, which are crucial elements in the semantic perception of images.

A second example in Fig. 13 concerns an image of a mountain range submitted to the system. The images retrieved all show similar mountainous landscapes, with a high degree of fidelity in the representation of relief and natural textures. This success illustrates the ability of CBIRH to effectively recognize and distinguish complex geographical structures, while maintaining remarkable visual consistency. These examples clearly show that the CBIRH system combines relevance and speed: the results not only match the initial query but also reflect the underlying visual and semantic intent. These observations highlight the maturity and robustness of our approach, as well as its potential for a variety of practical applications in the field of image search.

Using the selected test databases and the evaluation metrics mentioned above, we evaluated the CBIRH architecture. mAP results are shown in Fig. 14 showing the accuracy rate achieved by our system when tested on each selected database. A high accuracy rate indicates that most of the images retrieved by the system were relevant to user queries. For example, in tests with different query images, we found variations in accuracy depending on the nature of the images and their context. This enables us to assess the extent to which the CBIRH architecture can deliver relevant results.

In terms of execution time, the results are shown in Fig. 15 presents the time taken to process each query and return the corresponding results. We measured this time at every stage of the process, including image processing, distance calculation and result sorting. Short execution times are essential to guarantee a pleasant user experience, as users expect quick responses to their queries.

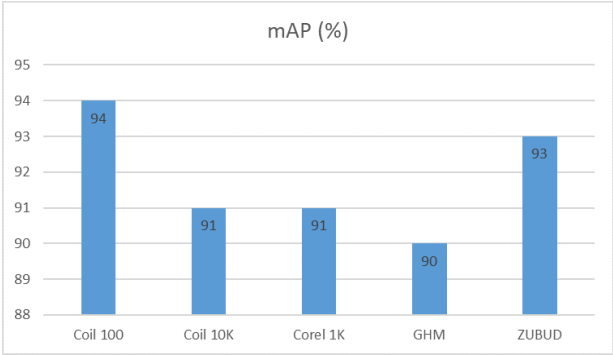

Fig. 14. Evaluation of image retrieval using (mAP).

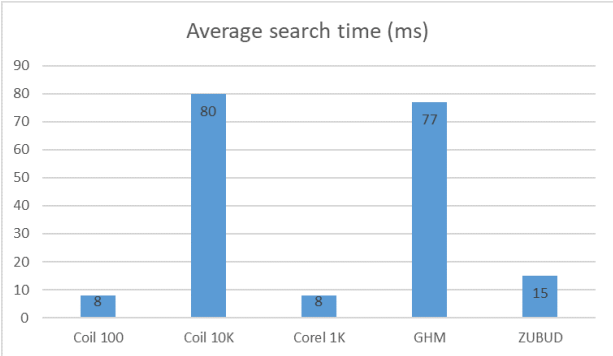

Fig. 15. Time search results of image retrieval.

Using these two metrics, we have obtained a global view of the performance of CBIRH architecture. The joint analysis of accuracy and execution time provides valuable information on system performance. For example, high accuracy combined with long execution time may indicate that the system is generating relevant results but needs to be optimized to improve its speed.

This combination enables us to strike a balance between quality of results and operational efficiency. In this way, we can assess not only the relevance of the

images retrieved, but also the speed at which they are presented to the user. A good compromise between these two dimensions is essential to deliver an image retrieval experience that is both efficient and satisfying.

III. Comparative Analysis of CBIRH Architecture and State-of-the-Art Approaches

The Table IV compares different state-of-the-art approaches to image retrieval, based on average response time (in milliseconds) and average accuracy (in percent). Analysis of these results shows significant differences in the performance of the methods evaluated, both in terms of speed and efficiency.

The Content-Based Image Retrieval Based on Multiple Extended Fuzzy-Rough Framework (CBIR-MEFRF) approach has the longest recovery time (2000 ms), with a relatively low accuracy of 70% [28]. It appears to be the least efficient of the techniques presented, which may indicate a less optimized or more complex process. It may be suitable in situations where speed is not a priority, although its low accuracy makes it less competitive against the other methods. Sketch Based Image Retrieval (SBIR) [29, 30], with a response time of (1800 ms) and an accuracy of 78%, offers a slight improvement on IRFSM in terms of accuracy. Although its efficiency is better, this approach is still far from the performance of modern techniques in terms of speed and accuracy.

TABLE IV. Comparisons of the Proposed Approach with State-of-Art Methods

References	Techniques	Retrieval Time (ms)	mAP
Islam *et al*. [28]	CBIR-MEFRF	2000	70
Zhang *et al*. [29]	SBIR	1800	78
Varish *et al*. [31]	IR-QBCI	1150	81
Ghrabat *et al*. [32]	GDBM-IR	1100	82
Faiyaz [33]	CBIR-SMANN	980	88
Ours	CBIRH	374	91.8

The Image Retrieval scheme using Quantized Bins of Color Image components (IR-QBCI) and Adaptive Tetrolet Transform [31] and Greedy learning of Deep Boltzmann Machine (GDBM)'s Variance and Search Algorithm for efficient Image Retrieval (GDBM-IR) [32] approaches are becoming more competitive, with response times of (1150 ms) and 1100 ms respectively, and accuracies of 81% and 82%. Their performance is significantly better than that of the first two methods, offering a good balance between speed and reliability. These methods are interesting intermediate options, combining efficiency and accuracy without the compromises noted for the previous techniques.

CBIR-Similarity Measure via Artificial Neural Network interpolation (CBIR-SMANN) [33] boasts significantly higher accuracy 88% and a recovery time of (980 ms), making it a reliable option combining speed and accuracy. It stands out from previous methods for its enhanced efficiency, ideal for applications requiring an optimum compromise between accuracy and speed.

Finally, the CBIRH approach sets itself apart by outperforming all other techniques. With an exceptionally short recovery time of (374 ms) and a high accuracy of 91.8% (calculated as the average of the results obtained from the five test databases), it offers the best combination of speed and reliability of all the methods analyzed. CBIRH is therefore the most powerful and optimized option for applications requiring maximum responsiveness and accuracy, making it an ideal choice for environments where speed and reliability are paramount.

IV. Conclusion and Futur Works

In this article, we have developed a strategy combining specific methods and skills to propose a new approach to image retrieval, based on CBIR system. This research includes a literature review, the design of a hybrid CBIR architecture, called CBIRH, and the evaluation of its performance. First, the literature review on image indexing systems identified the major challenges of existing CBIR methods, particularly in terms of accuracy and efficiency.

This diagnosis guided us in formulating a new approach capable of addressing these limitations. CBIRH architecture is based on two essential components: the computation of image descriptors that generate a unique feature vector for each image, and a selection step that selects the most relevant features to optimize the comparison between images in the database and those sought. This model improves search speed and accuracy. Finally, the evaluation of CBIRH, using average accuracy and average execution time metrics, validated its performance on five databases. These tests confirm the robustness and efficiency of our approach, which makes a significant contribution to the field of image retrieval, with prospects for improvement for applications requiring high precision and responsiveness.

In our future work, we aim to further improve the performance of approach. To increase mAP and reduce processing time, we plan to integrate Vision Transformer architectures and generative AI. These models will generate contextual annotations for images, offering a richer description contextually adapted to user queries. This evolution should enhance the accuracy of search results by aligning the answers provided with the specific context of each user query, paving the way for more precise and intelligent image retrieval.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Abdelkrim Saouabe: Conceptualization, software,

validation, methodology, formal analysis, writing—original draft. Said Tkatek: Supervision, reviewing. All authors had approved the final version.

REFERENCES

[1] B. Ramamurthy and K. R. Chandran, "CBMIR: Content based medical image retrieval using shape and texture content," *Advances in Modeling B*, vol. 56, no. 2, pp. 84–95, 2013.

[2] M. Kumar and M. Singh, "CBMIR: Content based medical image retrieval system using texture and intensity for eye images," *Int. J. Sci. Eng. Res.*, vol. 7, no. 9, pp. 2229–5518, 2016.

[3] Ş. Öztürk, "Hash code generation using deep feature selection guided siamese network for content-based medical image retrieval," *Gazi University Journal of Science*, vol. 34, no. 3, pp. 733–746, 2021.

[4] E. Paiz-Reyes, N. Nunes-de-Lima, and S. Yildirim-Yayilgan, "GIF image retrieval in cloud computing environment," in *Proc. International Conference Image Analysis and Recognition*, 2018, pp. 261–268.

[5] S. A. Mahmoudi, M. A. Belarbi, E. W. Dadi, S. Mahmoudi, and M. Benjelloun, "Cloud-based image retrieval using GPU platforms," *Computers*, vol. 8, no. 2, p. 48, 2019.

[6] A. Saouabe, S. Tkatek, H. Oualla, and I. Mourtaji, "Image indexing approaches for enhanced content-based image retrieval: An overview," in *Proc. 2024 International Conference on Ubiquitous Networking (UNet)*, 2024, pp. 1–7. https://ieeexplore.ieee.org/abstract/document/10794711/

[7] A. Saouabe, S. Tkatek, H. Oualla, and C. S. Henriquez, "Large-scale image indexing and retrieval methods: A PRISMA-based review," *International Journal of Advanced Computer Science & Applications*, vol. 15, no. 7, 2024. doi: 10.14569/IJACSA.2024.0150732

[8] C. Leng, H. Zhang, B. Li, G. Cai, Z. Pei, and L. He, "Local feature descriptor for image matching: A survey," *IEEE Access*, vol. 7, pp. 6424–6434, 2018.

[9] G. Xie, B. Guo, Z. Huang, Y. Zheng, and Y. Yan, "Combination of dominant color descriptor and Hu moments in consistent zone for content based image retrieval," *IEEE Access*, vol. 8, pp. 146284–146299, 2020.

[10] P. Garcia-Fernandez, D. Cores, and M. Mucientes, "Enhancing few-shot object detection through pseudo-label mining," *Image and Vision Computing*, vol. 154, 105379, 2025.

[11] Y. Li, Y. Wu, C. Liu, and X. Wu, "IAC-ReCAM: Two-dimensional attention modulation and category label guidance for weakly supervised semantic segmentation," *Image and Vision Computing*, vol. 136, 104738, 2023.

[12] K. Tong and Y. Wu, "Rethinking PASCAL-VOC and MS-COCO dataset for small object detection," *Journal of Visual Communication and Image Representation*, vol. 93, 103830, 2023.

[13] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal Visual Object Classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, 2010. doi: 10.1007/s11263-009-0275-4

[14] B. Soumia, "Indexing and searching for images by content in large image corpora," Ph.D. dissertation, Université Mohamed Boudiaf des Sciences et de la Technologi, Oran, Algeria, 2017. https://www.ccdz.cerist.dz/admin/notice.php?id=00000000000000860421000903 (in French)

[15] N. Bourkache, S. Sidhom, and M. Laghrouche, "Comparative study in indexing applied to medical images," in *Proc. 5th. International Symposium ISKO-Maghreb (2015) on Knowledge Organization in the Perspective of Digital Humanities: Researches and Applications*, 2015, p. 191. https://inria.hal.science/hal-01255992/ (in French)

[16] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications—clinical benefits and future directions," *International Journal of Medical Informatics*, vol. 73, no. 1, pp. 1–23, 2004.

[17] A. Shifa, M. B. Imtiaz, M. N. Asghar, and M. Fleury, "Skin detection and lightweight encryption for privacy protection in real-time surveillance applications," *Image and Vision Computing*, vol. 94, 103859, 2020.

[18] K. Houari and K. Mohamed-Khireddine, "Content-based image search," M.S. thesis, Univ. Mentouri Constantine, Constantine, Algeria, 2010. https://www.academia.edu/download/81149923/HOU5734.pdf (in French)

[19] R. Dowerah and S. Patel, "Comparative analysis of color histogram and LBP in CBIR systems," *Multimedia Tools and Applications*, vol. 83, no. 5, pp. 12467–12486, 2023. doi: 10.1007/s11042-023-15955-0

[20] J. Peeples, W. Xu, and A. Zare, "Histogram layers for texture analysis," *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 4, pp. 541–552, 2021.

[21] E. R. Vimina and K. P. Jacob, "Content based image retrieval using low level features of automatically extracted regions of interest," *Journal of Image and Graphics*, vol. 1, no. 1, pp. 7–11, 2013. doi: 10.12720/joig.1.1.7-11

[22] H. Fairouz, "Image database indexing," Ph.D. dissertation, Fac. Sci., Univ. Ferhat Abbas Sétif 1, Sétif, Algeria, 2021. http://dspace.univ-setif.dz:8888/jspui/handle/123456789/3738 (in French)

[23] D. Mahanta, D. Hazarika, and V. K. Nath, "Local bit-plane domain 3D oriented arbitrary and circular shaped scanning patterns for bio-medical image retrieval," *Journal of Image and Graphics*, vol. 11, no. 2, pp. 212–226, 2023.

[24] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004. doi: 10.1023/B:VISI.0000029664.99615.94

[25] A. Razzaque and A. Badholia, "PCA based feature extraction and MPSO based feature selection for gene expression microarray medical data classification," *Measurement: Sensors*, vol. 31, 100945, 2024.

[26] S. Mozgai, C. Kaurloto, J. Winn *et al*., "Machine learning for semi-automated scoping reviews," *Intelligent Systems with Applications*, vol. 19, 200249, 2023.

[27] A. N. Khusna, "Siamese neural networks with chi-square distance for trademark image similarity detection," *Scientific Journal of Informatics*, vol. 11, no. 2, pp. 429–438, 2024.

[28] S. M. Islam, M. Banerjee, S. Bhattacharyya, and S. Chakraborty, "Content-based image retrieval based on multiple extended fuzzy-rough framework," *Applied Soft Computing*, vol. 57, pp. 102–117, 2017. doi: 10.1016/j.asoc.2017.03.036

[29] Y. Zhang. X. Qian, and X. Tan, "Sketch-based image retrieval by salient contour reinforcement," *IEEE Transactions on Multimedia*, vol. 18, no. 8, pp. 1604–1615. https://ieeexplore.ieee.org/document/7469342

[30] Y. Liu, D. Dai, K. Zou *et al*., "Prior semantic-embedding representation learning for on-the-fly FG-SBIR," *Expert Systems with Applications*, vol. 255, 124532, 2024.

[31] N. Varish, A. K. Pal, R. Hassan *et al*., "Image retrieval scheme using quantized bins of color image components and adaptive tetrolet transform," *IEEE Access*, vol. 8, pp. 17639–117665, 2024. https://ieeexplore.ieee.org/abstract/document/9121956

[32] M. J. J. Ghrabat, G. Ma, Z. A. Abduljabbar, M. A. Al Sibahee, and S. J. Jassim, "Greedy learning of Deep Boltzmann Machine (GDBM)'s variance and search algorithm for efficient image retrieval," *IEEE Access*, vol. 7, pp. 169142–169159, 2019. doi: 10.1109/ACCESS.2019.2948266

[33] F. Ahmad, "Deep image retrieval using artificial neural network interpolation and indexing based on similarity measurement," *CAAI Transactions on Intelligence Technology*, vol. 7, no. 2, pp. 200–218, 2022. doi: 10.1049/cit2.12083