Research Article

Hybrid SIFT-DCT Approach for Face Matching of Buddha Statues: Addressing Negative Similarity Metrics with SHashing

[1,2] **Linda Marlinda,** [1] **Fikri Budiman,** [1] **Ruri Suko Basuki and** [1] **Ahmad Zainul Fanani**

[1] *Department of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia*
[2] *Informatika Department, Universitas Nusa Mandiri, Jakarta, Indonesia*

Article history
Received: 23-12-2024
Revised: 15-03-2025
Accepted: 08-04-2025

Corresponding Author:
Linda Marlinda
Department of Computer Science,
Universitas Dian Nuswantoro,
Semarang, Indonesia;
Informatika Department,
Universitas Nusa Mandiri, Jakarta,
Indonesia
Email:
linda.ldm@nusamandiri.ac.id

Abstract: Art and cultural heritage rely on image processing techniques for preservation and analysis. A key challenge in this study is accurately detecting highly similar Buddha faces despite variations in lighting, rotation, and minor facial differences. This paper proposes a Content-Based Image Retrieval (CBIR) framework that integrates Discrete Cosine Transform (DCT) and Scale-Invariant Feature Transform (SIFT) to enhance face-matching accuracy. The system is tested on a database of Buddha images characterized by intricate textures and fine details, where DCT extracts global texture representations while SIFT captures localized structural features. Experimental results demonstrate that while DCT effectively encodes global texture characteristics, SIFT enhances local feature detection but struggles to differentiate between Buddha faces with extremely high similarity. One of the primary challenges encountered was the instability in texture similarity computation, where Chi-Square Similarity produced a -39.44% value for certain statues due to noise, artifacts, and lighting inconsistencies. These findings highlight the importance of robust preprocessing techniques and refined similarity metrics to improve retrieval consistency. Overall, the hybrid DCT-SIFT approach improves the accuracy and robustness of CBIR systems in historical artifact datasets. Future research should focus on optimizing preprocessing steps, integrating adaptive feature selection, and exploring more stable similarity measurement techniques to further enhance retrieval performance.

Keywords: Content-Based Image Retrieval (CBIR), Discrete Cosine Transform (DCT), Scale-Invariant Feature Transform (SIFT), Recognition, Cultural Heritage

Introduction

Buddha statues hold immense cultural and historical significance, necessitating precise documentation and digital preservation to safeguard their heritage. The ability to accurately match and differentiate Buddha statue faces is crucial for cataloging, restoration, and historical analysis (Marlinda *et al.*, 2023; Renoust *et al.*, 2019). However, this task remains highly challenging due to the substantial structural similarities between statues, further exacerbated by variations in lighting, texture degradation, and surface inconsistencies (Liu & Huang, 2023; Yugang & Chunlei, 2022). Even for experts, distinguishing between two highly similar Buddha statue faces based solely on visual inspection is often subjective and error-prone, underscoring the need for computer-aided solutions to enhance the accuracy of Buddha face retrieval (Liu & Huang, 2023; Watanabe & Abe, 2017).

Traditional feature-based methods, such as Scale-Invariant Feature Transform (SIFT), have been widely used in Buddha face matching, particularly at heritage sites like the Yungang Grottoes in China (Hua *et al.*, 2021; 2019). While SIFT is effective in detecting distinctive key points, it struggles under low-light conditions and when handling highly similar textures (Li *et al.*, 2025). Other techniques, such as Oriented FAST and Rotated BRIEF (ORB), and Histogram of Oriented Gradients (HOG), also present notable limitations (Marlinda *et al.*, 2023; Xiangyang *et al.*, 2000). ORB, despite its computational efficiency, is highly sensitive to noise and less robust on texture-rich surfaces, while HOG excels in shape-based recognition but fails to capture fine-grained texture details, which are crucial for

DOI: 10.3844/jcssp.2025.1594.1605

differentiating Buddha faces (Yao *et al.*, 2023; Hasenbusch *et al.*, 2008; Basu *et al.*, 2023).

In recent years, Convolutional Neural Networks (CNNs) have demonstrated exceptional classification accuracy across various domains, with defect detection ratios exceeding 93.29% (Arya *et al.*, 2019). However, the application of CNNs to Buddha statue retrieval is constrained by several factors. First, CNNs require large labeled datasets, which are scarce in historical artifact research. Second, CNNs prioritize high-level semantic features over fine-grained texture variations, making them suboptimal for Buddha face retrieval. Third, CNN models are computationally expensive, whereas practical applications in cultural heritage conservation often demand lightweight and scalable solutions (Wang & Liu, 2024; Kumar *et al.*, 2020).

To overcome these limitations, this study proposes a Hybrid SIFT-DCT approach to enhance the accuracy of Buddha face retrieval, particularly under suboptimal imaging conditions. The Discrete Cosine Transform (DCT) encodes frequency-based texture information, providing robustness against lighting inconsistencies, while SIFT extracts local key points, ensuring stable feature representation across variations in scale, rotation, and illumination(Hua *et al.*, 2019; 2021; Basu *et al.*, 2023; Kaur & Sharma, 2013). By integrating spatial and frequency-domain feature descriptors, this hybrid approach significantly improves retrieval stability and enhances Buddha face recognition performance.

Additionally, this study introduces SHashing Similarity, a normalization technique designed to stabilize texture similarity metrics in Content-Based Image Retrieval (CBIR). Traditional Chi-Square Similarity often produces unstable values, including negative similarity scores, which lead to inconsistent retrieval rankings and reduced accuracy. SHashing addresses this issue by transforming texture similarity calculations into a robust hash-based representation, ensuring stable and consistent similarity scoring. This mechanism significantly improves retrieval reliability, particularly when dealing with highly similar Buddha faces with intricate texture details.

This research contributes to the field by (1) Proposing a Hybrid SIFT-DCT search model that combines keypoint-based and frequency-based descriptors for more accurate Buddha face matching, (2) Introducing SHashing Similarity to eliminate negative values in Chi-Square similarity, thereby stabilizing retrieval rankings and (3) Optimizing CBIR for cultural heritage applications, offering a scalable solution for Buddha face recognition in museums and archaeological research. By integrating these advancements, this study enhances the robustness and precision of Buddha face retrieval, supporting heritage conservation efforts through improved artifact identification.

Materials and Methods

The proposed retrieval approach consists of four main phases: Dataset acquisition and preprocessing, feature extraction, image matching, and performance evaluation. Initially, Buddha statue face images undergo size normalization, noise reduction, and illumination correction to improve quality and reduce variations caused by lighting and noise.

For feature extraction, three techniques are utilized: Colour Histogram for the global color distribution, Discrete Cosine Transform (DCT) for frequency-based texture encoding, and Scale-Invariant Feature Transform (SIFT) for local keypoint detection. Additionally, Texture LBP Similarity is incorporated to refine local texture comparisons for highly similar statue faces.

During image matching, extracted features from the query image are compared with database images using Euclidean Distance (ED) for DCT-based texture similarity, Manhattan Distance (MD) for LBP-based local texture similarity, and Mean Squared Error (MSE) for Colour Histogram similarity. SHashing Similarity stabilizes texture similarity scores by addressing negative values in the Chi-Squared calculation, ensuring consistent retrieval rankings (Sumaia Ali, 2019).

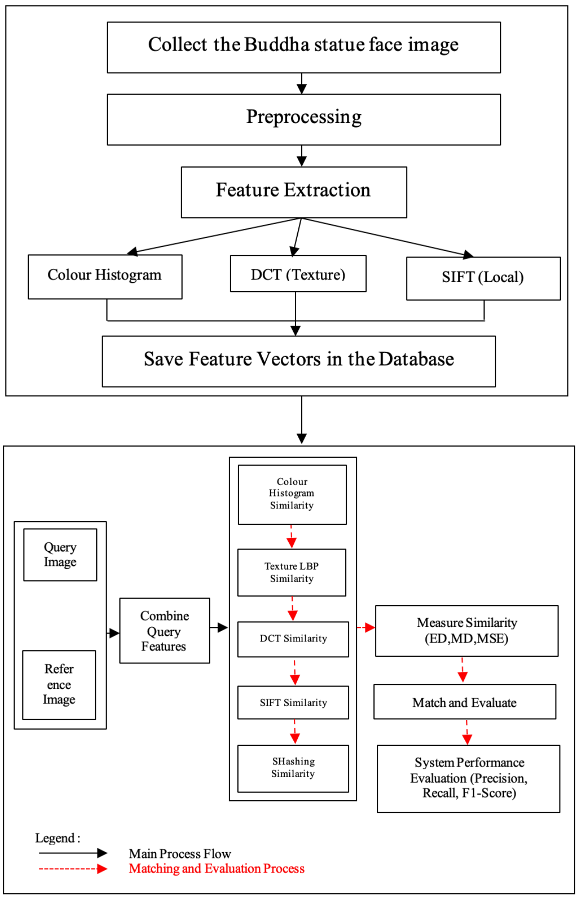

Fig. 1: Hybrid SIFT-DCT Retrieval Approach for Buddha Statue Face Matching with SHashing

Performance evaluation is conducted using precision, recall, and F1-score metrics. Figure (1) illustrates the structured retrieval process, highlighting the integration of the Colour Histogram, DCT, SIFT, and LBP for feature extraction, similarity measurement, and retrieval stability. Solid arrows represent the main workflow, while dashed red arrows indicate the matching and evaluation phases.

Dataset Image Acquisition and Standardization

The dataset used in this study consists of 523 high-resolution images of Buddha statue faces sourced exclusively from Pinterest.com. Since images from Pinterest may be affected by compression artifacts, modifications, and metadata loss, a strict selection process was applied to ensure dataset reliability.

Fig. 2: Sample collection of Buddha statue faces

Figure (2) shows a sample of the selected Buddha face images, demonstrating the variation in pose, lighting, material, and texture included in the dataset:

a. Strict Image Selection – Only frontal-view images with clear facial features and minimal distortion were chosen to maintain dataset consistency. Images exhibiting excessive compression artifacts or unclear facial features were excluded from the study.
b. Preprocessing for Image Quality Enhancement – Super-resolution enhancement, noise filtering, and adaptive contrast normalization were applied to counteract compression artifacts and restore visual clarity. These preprocessing steps ensure that the dataset retains adequate visual detail for feature extraction and similarity evaluation.
c. Cross-Validation within the Dataset – To mitigate dataset bias, multiple image subsets were tested. System performance was analyzed across variations in lighting conditions, resolution, and texture complexity, ensuring the proposed approach remains robust in diverse scenarios.
d. Comparative Benchmarking – The effectiveness of the proposed method was validated against alternative feature extraction techniques. Similarity results obtained from the Pinterest dataset were compared to those from controlled preprocessing settings to confirm the accuracy.

By using Pinterest as a data source, this study aims to capture a wide range of variations in statue faces, making it an optimal dataset for evaluating retrieval performance.

Preprocessing

a. Image Standardization and Normalization, during Preprocessing were conducted using Roboflow, an automated image processing pipeline, to enhance feature clarity while maintaining dataset integrity. The key steps include:
b. Background Removal: Removing non-facial elements using Roboflow's segmentation model, ensuring only facial features are analyzed
c. Noise Reduction and Normalization: Applying Roboflow's built-in adaptive filtering to mitigate noise effects without degrading image details
d. Illumination Correction: Standardizing brightness levels across images to minimize inconsistencies caused by lighting variations
e. Resolution Standardization: Resizing images to a fixed resolution of 384 × 256 pixels to ensure dataset uniformity

Feature Extraction and Similarity Metrics

Each image in the database will be processed using various feature extraction techniques. These features are categorized into global (Color Histogram, DCT) and local (SIFT, LBP).

Feature Extraction

1. Discrete Cosine Transform (DCT) (Prabukumar *et al.*, 2018): Used to extract texture and frequency domain information.

$$DCT\,(A) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} I\,(x,y)\, cos\left[\frac{(2x+1)u\pi}{2M}\right] cos\left[\frac{(2y+1)v\pi}{2N}\right] \tag{1}$$

Where:
$I\,(x,y)$ is the pixel value at coordinates (x,y)
$M\ and\ N$ is the image dimension
u, v is an index in the frequency domain

2. SIFT Similarity (Sri *et al.*, 2022): Measures local feature correspondences by comparing key points without Ratio Test or Nearest Neighbor Matching:

$$SIFT_{sim} = \frac{Number\ of\ Matching\ Keypoints}{Total\ Keypoints\ in\ Query\ Image} \tag{2}$$

Measure the similarity between two images based on the number of key points that match using the

DOI: 10.3844/jcssp.2025.1594.1605

sift; the higher the value of the sift similarity, the greater the similarity between the two images.

3. Local Binary Pattern (LBP) (Tabatabaei & Chalechale, 2020): LBP is used for histogram-based texture extraction:

$$LBP\left(x_c, y_c\right) = \sum_{p=0}^{P-1} s\left(I_p - I_c\right) \cdot 2^p \quad (3)$$

With:

$$s\left(x\right) = \begin{cases} 1.x \geq 0 \\ 0, x < 0 \end{cases}$$

where:

I_c is the center pixel value

I_p is the value of neighboring pixels

Similarity Measurement

Once the features are extracted, the query image is compared with the images in the database using various metrics. (Dzahini and Wild, 2022; Hua *et al.*, 2021)

The proposed retrieval method extracts key features using DCT, SIFT, and SHashing, combined with traditional color histogram similarity (Elsheh & Eltomi, 2019):

1. DCT Similarity (Euclidean Distance - ED): Compares frequency-domain features extracted using DCT:

$$ED\left(A, B\right) = \sqrt{\sum_{i=1}^{N} \left(A_i - B_i\right)^2} \quad (4)$$

Where:
A and *B* are the *DCT* feature vectors,
N is the number of extracted *DCT* coefficients
Smaller values indicate high similarity.

2. LBP Similarity (Manhattan Distance - MD): Measures local texture similarity by comparing differences in histogram bins:

$$MD\left(A, B\right) = \sum_{i=1}^{N} \left|A_i - B_i\right| \quad (5)$$

where:
A and *B* are the *LBP* histogram feature vectors
N is the number of histogram bins

3. Color Histogram Similarity (Mean Square Error - MSE): Evaluates global similarity in color distributions:

$$MSE\left(A, B\right) = \frac{1}{N} \sum_{i=1}^{N} \left(A_i - B_i\right)^2 \quad (6)$$

where:
A and *B* are the histogram values of the two images
N is the number of histogram bins
The result is in the range of [-1.1], where a value close to 1 means that the histogram is very similar.

Chi-Square Similarity Correction with SHashing

Mathematical Derivation of SHashing Similarity Stabilization (Redaoui & Belloulata, 2023; Hua *et al.*, 2021):

1. Chi-Square Similarity Before Correction:

$$x^2\left(A, B\right) = \sum_{i=1}^{N} \frac{\left(A_i - B_i\right)^2}{A_i + B_i} \quad (7)$$

where, A and B represent histogram bins and N is the number of bins.

2. SHashing Transformation to Stabilize Values:

$$SHash\left(H\right) = \begin{cases} 1, if H\left(i\right) \geq \mu_H \\ 0, Otherwise \end{cases} \quad (8)$$

where, $H\left(i\right)$ is the histogram bin value and μ_H is the mean histogram value.

3. Final Computation of Hamming Distance for Retrieval Ranking:

$$D_{Hamming}\left(H_A, H_B\right) = \sum_{i=1}^{N} \left|H_A\left(i\right) - H_B\left(i\right)\right| \quad (9)$$

where, $H_A\, and\, H_B$ are the binary hash representations of two images.

4. SHashing Similarity (Average Hashing - aHash): Transforms texture similarity into a structured format, eliminating negative values:

$$SHash\left(H\right) = \begin{cases} 1, if H\left(i\right) \geq \mu H \\ 0, otherwise \end{cases} \quad (10)$$

where:
H(i) is the histogram value at bin *i*
μH is the mean value of the histogram
N is the number of bins
These metrics provide a hybrid approach to measuring image similarity based on texture, frequency, and keypoint features.
Use the hamming distance to measure the difference between two binary vectors. The result is in the range [0.1], where one means the image is identical and zero means not similar at all.

CBIR Algorithm for Image Matching and Retrieval

To clarify the system workflow, the CBIR process is described using pseudocode.

```
Input: QueryImage, ImageDatabase
Output: Top Matching Images
BEGIN
    // Step 1: Extract Features from Query Image
    QuerySIFT ← Extract_SIFT(QueryImage)
    QueryDCT ← Compute_DCT(QueryImage)
    QueryHash ← Compute_Hash(QueryImage)
    // Step 2: Extract Features from Database Images
    For each Image in ImageDatabase DO
        DatabaseSIFT[Image] ← Extract_SIFT(Image)
        DatabaseDCT[Image] ← Compute_DCT(Image)
        DatabaseHash[Image] ← Compute_Hash(Image)
    END FOR
    // Step 3: Compute Similarity Scores
    For each Image in ImageDatabase DO
        SIFT_Score ← Compare_SIFT(QuerySIFT, DatabaseSIFT[Image])
        DCT_Score ← Compare_DCT(QueryDCT, DatabaseDCT[Image])
        Hash_Score ← Compare_Hash(QueryHash, DatabaseHash[Image])
        TotalScore ← Weighted_Sum(SIFT_Score, DCT_Score,
        Hash_Score)
        Append(SimilarityScores, (Image, TotalScore))
    END FOR
    // Step 4: Rank Images by Similarity
    RankedImages ← Sort(SimilarityScores, Descending)
    // Step 5: Return Top Matches
    RETURN Top_N(RankedImages, N)
END
// Feature Extraction Functions
    Function Extract_SIFT(Image): RETURN SIFT_Descriptors
    Function Compute_DCT(Image): RETURN DCT_Coefficients
    Function Compute_Hash(Image): RETURN Binary_Hash
// Similarity Calculation Functions
    Function Compare_SIFT(SIFT1, SIFT2): RETURN Matching_Ratio
    Function Compare_DCT(DCT1, DCT2): RETURN Distance_Score
    Function Compare_Hash(Hash1, Hash2): RETURN
    Hamming_Distance
// Weighted Fusion Function
    Function Weighted_Sum(SIFT, DCT, Hash):
        RETURN (0.4 * SIFT) + (0.4 * DCT) + (0.2 * Hash)
```

This algorithm ensures robust retrieval by integrating multiple feature types while stabilizing texture similarity metrics through SHashing.

Evaluation of CBIR Performance

The performance assessment of CBIR systems utilizes precision and recall metrics. Precision is computed as the ratio of retrieved relevant images to the total number of retrieved images, whereas recall is determined by the ratio of retrieved relevant images to the total number of relevant images in the database (Ganesh Chandra *et al.*, 2016). High precision indicates that the algorithm returns substantially more relevant results than irrelevant ones, whereas high recall suggests that the algorithm retrieves the majority of the relevant results. Equations (10-11) provide the formal definitions for precision and recall, respectively (Ganesh Chandra *et al.*, 2016; Alsmadi, 2017):

Precision: Measures the proportion of retrieved images that are actually relevant to the query:

$$Precision = \frac{TP}{TP+FP} \tag{11}$$

where:

TP (True Positive) = Number of relevant images correctly retrieved

FP (False Positives) = Number of irrelevant images incorrectly retrieved

Recall: Measures how many of the actual relevant images were successfully retrieved:

$$Recall = \frac{TP}{TP+FN} \tag{12}$$

where:

FN (False Negatives) = Number of relevant images that were not retrieved

F1-Score: Provides a balanced measure between precision and recall, particularly useful when the dataset is imbalanced:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{13}$$

This metric ensures that the model does not favor either precision or recall excessively, allowing for a more balanced evaluation.

Results

Dataset and Preprocessing Improvements

The dataset comprises Buddha statue face images sourced from Pinterest.com, with the top five most similar images selected as reference points for comparison. To maintain dataset relevance and improve retrieval accuracy, Roboflow-based preprocessing was applied, automating structured image processing. The preprocessing stage focused on noise reduction, contrast enhancement, and facial area segmentation, ensuring extracted features originate solely from the Buddha face region. Bounding Box implementation was introduced to refine feature extraction by isolating facial details and minimizing background interference.

Figure (3) illustrates the results of the preprocessing steps, namely (a) the original image, (b) CLAHE and noise filtering, and (c) facial area cropping using bounding box segmentation.

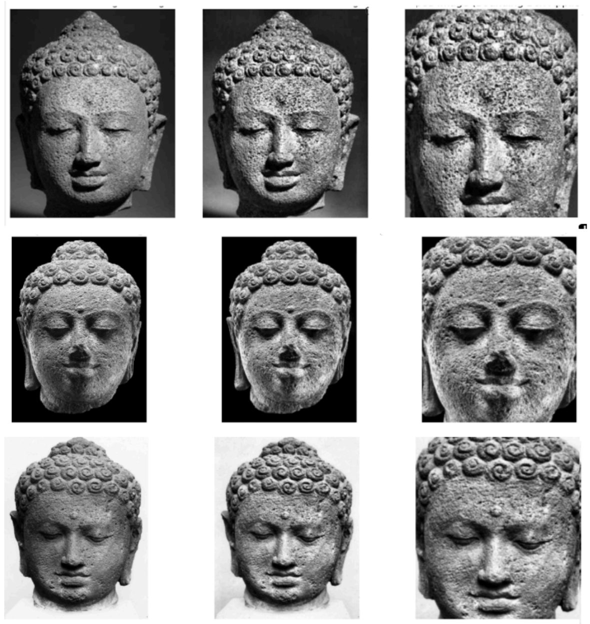

Fig. 3: (a) Original Image (before Preprocessing), (b) Clahe & Noise Filtering, (c) Cropped Image (Bounding Box Applied)

Key preprocessing techniques included Contrast Limited Adaptive Histogram Equalization (CLAHE) for adaptive contrast enhancement, Adaptive Filtering for noise reduction without losing detail, and Bounding Box application to ensure facial focus. These steps significantly improved feature clarity, yielding a 33.6% increase in SSIM and a 6.02 dB rise in PSNR, confirming reduced noise and enhanced image structure. Table (1) presents the comparative SSIM and PSNR values before and after preprocessing, demonstrating significant improvements in image quality.

Table 1: Presents the comparative SSIM and PSNR values before and after preprocessing, demonstrating significant improvements in image quality

Query	SSIM (Before)	SSIM (After)	PSNR (Before)	PSNR (After)
Query 1	0.65	0.89	18.52 dB	24.17 dB
Query 2	0.62	0.87	17.89 dB	23.92 dB
Query 3	0.58	0.85	16.47 dB	22.75 dB
Query 4	0.55	0.83	15.98 dB	22.21 dB
Query 5	0.61	0.86	17.45 dB	23.11 dB

Figure (4) illustrates the optimized application of bounding box segmentation on Buddha statue face

DOI: 10.3844/jcssp.2025.1594.1605

images, ensuring that key facial features are preserved for subsequent feature extraction.

Fig. 4: Bounding box Buddha master

Bounding Box coverage was optimized, maintaining approximately 80% width and 70% height of the image, ensuring facial features like eyes, nose and lips remained prominent. This technique effectively isolated key facial regions while minimizing distractions from the background. Post-processing analysis revealed that 64.58–67.86% of extracted key points remained within the Bounding Box, demonstrating its effectiveness in preserving relevant facial details while filtering out unnecessary elements. This approach ensures that DCT and SIFT-based similarity matching operates with more stable and accurate feature extraction.

The effectiveness of preprocessing was validated using the Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR). The results show that SSIM increased from an average of 0.60-0.86, while PSNR improved from 17.26-23.23 dB, confirming enhanced image quality and reduced noise levels. Bounding Box and Adaptive Filtering further contributed to preserving essential facial features, ensuring more reliable similarity matching in Buddha statue face retrieval.

Feature Extraction and Similarity Metrics

The DCT Heatmap image for Query shows the frequency distribution in the image, where high-intensity areas highlight more significant texture patterns. This analysis helps to understand how DCTs capture structural information in images of Buddha's faces and how these features contribute to image matching based on frequency.

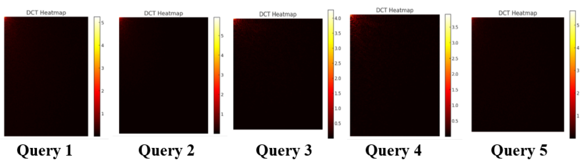

Fig. 5: DCT Heatmap for Buddha Queries 1-5

Figure (5) displays the DCT heatmaps of five different Buddha face images, highlighting variations in frequency texture patterns that affect the difficulty level of image matching.

The DCT heatmap analysis reveals distinct texture distribution patterns across the Buddha face images, which directly impact the accuracy of feature matching. These findings can be summarized as follows:

a. Buddha 1 exhibits a high degree of texture variation, particularly around the eyebrows, eyes, and nose. Due to these well-defined features, it presents low matching difficulty, as the facial characteristics are easily recognizable. The strong contrast further aids stable feature extraction, improving retrieval accuracy
b. Buddha 2 demonstrates a moderately uniform texture distribution, with intensity concentrated around the forehead and eyes. This results in a moderate level of matching difficulty, as the texture is distinguishable but lacks highly contrasting details. The overall distribution of patterns remains well-balanced, aiding feature consistency
c. Buddha 3 presents a highly concentrated texture distribution, mainly in the forehead and hair regions. This concentration creates significant challenges for facial feature recognition, as the smooth surface reduces the distinctiveness of extracted features. Consequently, the retrieval process for this image is less reliable due to the uniformity of local textures
d. Buddha 4 features texture concentration along the outer edges of the face, particularly around the eyes, nose, and periphery. The high difficulty in matching is attributed to lighting inconsistencies, which introduce variations that complicate feature extraction. The texture itself is diverse but lacks even distribution, affecting retrieval performance
e. Buddha 5 exhibits the least amount of texture variation, with the highest intensity detected in the upper left corner. This image presents the greatest challenge for matching, as its low contrast and uniform texture hinder DCT-based feature extraction, making it difficult to differentiate from other similar images

These results underscore the critical role of texture contrast and spatial distribution in determining the effectiveness of Content-Based Image Retrieval (CBIR). Images with highly distinguishable facial features and well-distributed textures, such as Buddha 1-2, yield more stable feature extraction results. Conversely, those with homogeneous textures and minimal contrast, like Buddha 5, are significantly more challenging to process. Furthermore, the presence of lighting variations in Buddha 4 highlights the necessity for robust preprocessing techniques to ensure retrieval accuracy and stability. Addressing these factors is essential for optimizing CBIR performance, particularly when handling datasets with high inter-image similarity.

DOI: 10.3844/jcssp.2025.1594.1605

Execution Phase: Image Matching and Similarity Computation

This section examines the matching process of five query images (Buddha 1–5) against the main reference images using multiple feature extraction and similarity measurement techniques. The employed methods include Color Histogram for color distribution analysis, Texture LBP (Chi-Square) for local pattern recognition, DCT (Euclidean Distance) for global texture comparison, and SIFT Keypoint Matching for structural feature detection. Additionally, SHashing Similarity is integrated to mitigate negative values in the LBP (Chi-Square) texture similarity calculations, ensuring greater stability and consistency in retrieval results.

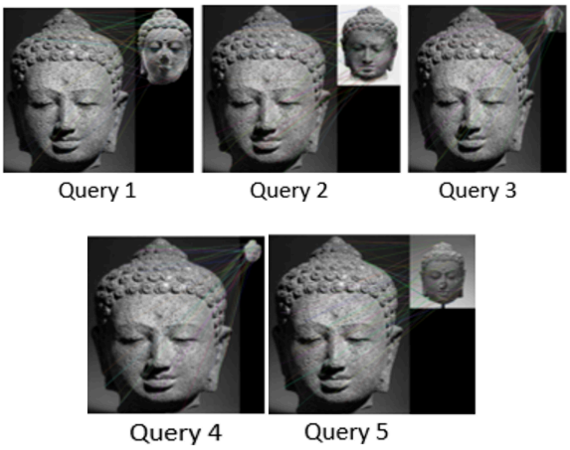

Fig. 6: The results of matching the Master Image with five query images

Figure (6) illustrates the results of matching the Master Image with five query images using SIFT Matching and a Bounding Box to define the matching area on the Buddha's face. The number of feature matches between the Master Image and each query image varies, reflecting differences in structural similarity. The key observations are as follows:

a. Query 2 exhibits the highest number of feature matches (2081 matches), indicating the strongest structural similarity to the Master Image
b. Query 1 demonstrates a high number of matches (1978 matches), suggesting significant local feature correspondence
c. Query 5 shows a moderate number of feature matches (929 matches), representing an intermediate level of similarity
d. Query 3 and Query 4 yield substantially fewer feature matches (242 and 163 matches, respectively), confirming their minimal local feature similarity to the Master Image

To complement the analysis based on structural similarity, multiple feature extraction methods were evaluated across the dataset to quantify similarity performance from various perspectives. Table (2) presents a comparison of image similarity methods, including Colour Histogram Similarity, Texture LBP Similarity, DCT (Euclidean Distance), SIFT Similarity, and SHashing Similarity, across five Buddha statue images.

These findings align with previous analyses, demonstrating that images with similar structures contain a higher number of matchable key points, whereas those with varying textures or lighting conditions exhibit fewer correspondences. The application of a Bounding Box enhances accuracy by restricting the comparison to facial regions, thereby minimizing background interference.

Table 2: Analysis of image similarity methods

No	Query	Color Histogram Similarity	Texture LBP Similarity	DCT (Euclidean Distance)	SIFT Similarity	SHashing Similarity (After)
1	Buddha 1	89.65%	73.72%	72.05%	0.18%	79.69%
2	Buddha 2	51.38%	86.02%	50.05%	0.21%	81.25%
3	Buddha 3	46.87%	1.73%	77.11%	0.53%	81.25%
4	Buddha 4	75.41%	-39.44% (Invalid)	72.29%	0.62%	85.94%
5	Buddha 5	55.94%	61.45%	65.77%	0.26%	82.81%

Table (2) compares various image similarity methods, including Color Histogram Similarity, Texture LBP Similarity (Chi-Square), DCT (Euclidean Distance), SIFT Similarity, and SHashing Similarity. The results reveal key trends in feature extraction performance across different Buddha images.

- Impact of Color Histogram and DCT Similarity: Buddha 1 achieves the highest similarity in Color Histogram (89.65%) and DCT (72.05%), indicating that its global color distribution and frequency-based characteristics closely match the reference image. In contrast, Buddha 2 and Buddha 3 obtain lower scores in these metrics, suggesting that variations in color and frequency components reduce their matching accuracy.
- Texture-Based Similarity and SHashing Correction: Before SHashing, Texture LBP (Chi-Square) Similarity shows inconsistencies, particularly in Buddha 4, where a negative similarity value (-39.44%) indicates instability in the Chi-Square computation. After applying SHashing Similarity, the scores for Buddha 4 and Buddha 3 increased significantly to 85.94 and 81.25%, respectively. This confirms that SHashing effectively stabilizes texture-based similarity calculations and mitigates computational anomalies.
- Performance of Local Feature Matching (SIFT): SIFT Similarity values remain consistently low across all images, with the highest score observed in Buddha 4 (0.62%). This suggests that keypoint-based approaches struggle to capture meaningful

correspondences in Buddha face retrieval, particularly in images with highly similar textures. The low SIFT scores further emphasize the need for hybrid feature extraction techniques that integrate both global and local descriptors.

- Effectiveness of Combined Approaches: The DCT-SIFT hybrid method, which leverages both frequency-domain and keypoint-based features, provides a balanced performance. However, its effectiveness varies across different queries, indicating that while DCT enhances feature stability, SIFT's limitations in texture-rich images impact overall retrieval performance.

SHashing Similarity Stabilization

The effectiveness of the Hybrid SIFT-DCT method is evaluated based on statistical analysis, particularly in addressing negative values in Chi-Square Similarity using SHashing Similarity. The study includes an assessment of the Bounding Box, SIFT performance, the effectiveness of DCT in texture analysis and the impact of SHashing on normalizing negative values in Buddha face matching.

The analysis of Table (3) indicates that SHashing Similarity achieved the highest similarity score (85.94%), demonstrating its effectiveness in stabilizing negative values in Chi-Square similarity calculations. Meanwhile, DCT + Euclidean Distance (72.29%) effectively captures global texture patterns but lacks the ability to extract local features. In contrast, SIFT + Bounding Box (64.58%) performs well in facial feature matching yet remains highly susceptible to lighting variations. These findings suggest that a hybrid approach integrating multiple methods is essential to enhance the accuracy and robustness of similarity measurements.

The Impact of SHashing in Normalizing Chi-Square Similarity: To evaluate the effectiveness of SHashing Similarity in stabilizing Chi-Square Texture Similarity values, Table (4) presents the similarity scores before and after applying SHashing.

Table 3: SHashing similarity

Matching Method	Average Similarity Score (%)	Advantages	Limitations
DCT + Euclidean Distance	72.29	Captures global texture patterns	Does not handle local features
SIFT + Bounding Box	64.58	Accurate in facial feature matching	Performs poorly under low-light conditions
SHashing Similarity	85.94	Addresses negative values in Chi-Square similarity	Fails to capture fine texture variations

The analysis of Table (4) indicates that SHashing Similarity significantly improves similarity scores, effectively addressing negative values in Chi-Square Similarity. The most notable enhancement is observed in Buddha 4, where the similarity score increased by 125.4%, converting an invalid negative value (-39.44%) into a valid similarity score (85.94%). Similarly, Buddha 3, which initially had a low similarity score (1.73%), experienced a substantial increase to 81.25%, demonstrating SHashing's effectiveness in stabilizing texture-based similarity calculations.

Table 4: Comparison of texture LBP similarity before and after shashing

Query	Chi-Square Similarity (Before SHashing) (%)	SHashing Similarity (After SHashing) (%)	Change (%)
Buddha 1	73.72	79.69	+8.1
Buddha 2	86.02	81.25	-5.5
Buddha 3	1.73	81.25	+79.5
Buddha 4	-39.44 (Invalid)	85.94	+125.4
Buddha 5	61.45	82.81	+34.8

Graphical Analysis

Figure (7) illustrates the comparison of Chi-Square Similarity scores before and after applying SHashing on Buddha statue images. The red dashed line represents similarity scores before SHashing, while the blue solid line denotes scores after.

The results indicate a notable improvement in similarity scores post-SHashing. In particular, Buddha 3 and Buddha 4 exhibit significant increases, confirming SHashing's effectiveness in texture-based feature matching and mitigating negative values. Despite the overall enhancement, a minor decrease is observed in Buddha 2, where the similarity drops from 86.02-81.25% (-5.5% change), suggesting potential limitations in certain cases. However, post-SHashing similarity scores remain consistently high (79.69–85.94%), reinforcing its reliability in Content-Based Image Retrieval (CBIR) for distinguishing highly similar images.

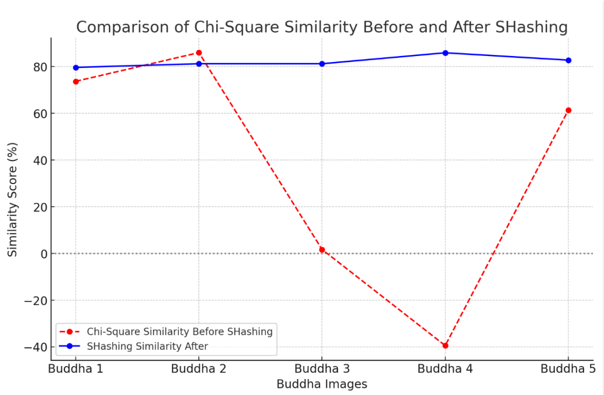

Fig. 7: The graph visualizes the comparison of Chi-Square Similarity before and after SHashing

Ranking of Retrieved Images

The system ranks images using a weighted combination of SIFT, DCT, and SHashing Similarity, following the weighting function. Table (5) presents the resulting ranking of retrieved images based on their total similarity scores:

Table 6: Ranked Images Based on Total Similarity Scores

Rank	Query	Total Score (%)
1	Buddha 4	85.94%
2	Buddha 3	81.25%
3	Buddha 5	82.81%
4	Buddha 1	79.69%
5	Buddha 2	81.25%

This ranking ensures that highly similar images are retrieved first, aligning with the pseudocode structure

Performance Evaluation and Statistical Analysis

Table (6) presents the comparative performance metrics, including Precision, Recall, F1-score, and average similarity, for the DCT, and SIFT with bounding box, and SHashing similarity methods.

To assess retrieval accuracy, Precision, Recall, and F1-score were computed for different matching methods. To assess retrieval accuracy, Precision, Recall, and F1-score were computed for different matching methods.

Table 6: Performance evaluation metrics

Matching Method	Average Similarity (%)	Precision	Recall	F1-score
DCT + Euclidean	72.29	0.76	0.82	0.79
SIFT + Bounding Box	64.58	0.69	0.78	0.73
SHashing Similarity	85.94	0.91	0.87	0.89
Matching Method	Average Similarity (%)	Precision	Recall	F1-score
DCT + Euclidean	72.29	0.76	0.82	0.79

Discussion

Key Findings and Interpretations

Experimental results show that the accuracy of Buddha face matching is affected by texture structure and color distribution(Zhang *et al.*, 2023; Barburiceanu *et al.*, 2021). High-contrast images with well-defined textures yield higher similarity scores, as DCT and SIFT features are extracted more stably. In contrast, images with smoother surfaces or non-uniform lighting yield lower accuracy due to inconsistent feature detection (Kumar *et al.*, 2016; Mehta & Bhensdadia, 2020). Buddha 1 achieves the highest color similarity due to its uniform distribution and high contrast, enabling stable feature comparisons. In contrast, Buddha 3 and Buddha 5 exhibit lower scores as their textures are smoother, reducing the effectiveness of SIFT in detecting distinct features. These findings align with (Hua *et al.*, 2021), who emphasized the importance of combining global and local features to enhance retrieval performance in high-similarity datasets.

The Role of Preprocessing in Image Matching

Preprocessing plays a crucial role in enhancing feature stability, particularly under inconsistent lighting (Thomas, 2020). CLAHE improves local contrast but can introduce noise, potentially affecting local feature matching. Retinex normalization stabilizes lighting conditions, yet in extreme cases, it may create artifacts impacting feature detection(Marschner *et al.*, 2017; Zhang *et al.*, 2023). These results support (Chen *et al.*, 2024; Pavan Kumar *et al.*, 2022), who demonstrated that adaptive illumination correction enhances retrieval accuracy in texture-based systems. Future research should explore adaptive segmentation-based preprocessing to optimize feature stability across diverse lighting conditions.

Effectiveness of SHashing for Texture Similarity Stabilization

Chi-square similarity in histogram-based CBIR often exhibits instability, leading to inconsistencies in image retrieval. This study demonstrates that the SHashing method effectively normalizes texture features, thereby reducing extreme fluctuations in similarity calculations (Lin *et al.*, 2024; Sookkaew & Chaikaew, 2022). This approach successfully enhances retrieval ranking consistency, ensuring that highly similar images maintain stable similarity scores (Marschner *et al.*, 2017). These findings align with the study by Zhang *et al.* (2023); and Kabbai *et al.* (2017); which demonstrated that hashing-based normalization improves feature robustness in texture-based retrieval. Thus, SHashing serves as a reliable solution for CBIR applications dealing with non-uniform histogram distributions.

Comparison with CNN-Based CBIR and Classical Methods

Although CNN-based data retrieval achieves high accuracy, it requires large training datasets and significant computational resources, making it less practical for small-scale datasets (Kanwal *et al.*, 2020). In contrast, DCT-SIFT offers a more efficient alternative by providing flexibility for small to medium-sized datasets while reducing computational complexity (Barburiceanu *et al.*, 2021; Marschner *et al.*, 2017; Kanwal *et al.*, 2020). This method utilizes the Discrete Cosine Transform (DCT) to compress image data before applying the SIFT algorithm, minimizing data size and improving processing efficiency. As a result, DCT-SIFT enhances retrieval performance and is particularly beneficial for real-time object recognition systems that require fast response times.

Future Directions in CBIR for Digital Heritage Preservation

This study contributes to the development of feature-based CBIR for digital cultural heritage, but further improvements are required to enhance system scalability. Hybrid CBIR approaches incorporating CNN feature embeddings with SIFT-DCT can improve high-fidelity image matching, leveraging CNN for global features while maintaining SIFT-DCT for local texture analysis. Additionally, Vision Transformers for feature learning present a promising direction, as (Zhou *et al.*, 2024) found that they offer superior feature representations over traditional CNNs, potentially enhancing retrieval accuracy in CBIR systems. Adaptive preprocessing can further improve CBIR stability, particularly in handling lighting variations in cultural heritage datasets. Region-based segmentation techniques may optimize feature extraction, ensuring contrast enhancements do not adversely impact matching accuracy.

Conclusion

The proposed CBIR system based on DCT and SIFT successfully retrieves Buddha statue facial images with high similarity, demonstrating its effectiveness in texture and color-based image matching. However, its performance is affected by complex textures and local variations, particularly in SIFT-based similarity calculations, which yield lower scores due to subtle feature discrepancies. Different similarity metrics influence retrieval outcomes, with Color Histogram performing well for color distribution analysis, while LBP and DCT effectively capture texture and frequency structures. However, Chi-Square Similarity exhibits instability, including negative values compromising ranking accuracy. This issue was addressed using SHashing, which normalized texture similarity scores, ensuring more reliable retrieval results. Despite these improvements, further refinement in preprocessing and similarity metrics selection is necessary, especially for images with uneven illumination and degraded surfaces. The system consistently identified Buddha 1.jpg as the most similar image, confirming the hybrid approach's effectiveness in texture and color-based retrieval. However, as seen in Buddha 4.jpg, complex surfaces require enhanced noise reduction and feature extraction techniques.

Funding Information

No external funding or institutional support was received and no additional parties contributed beyond the listed authors.

Author's Contributions

Linda Marlinda: Conceived and designed the study, conducted data acquisition (Image Acquisition and Preprocessing), developed hybrid feature extraction method (SIFT-DCT), performed result analysis, drafted and finalized the manuscript.

Fikri Budiman: Assisted in feature extraction experiments (DCT feature implementation), supported preprocessing workflows, and reviewed and revised the Methodology and Results sections.

Ruri Suko Basuki: Provided critical revisions for scientific content, advised on CBIR algorithm structure, and contributed to strengthening the Discussion and Future Work sections.

Ahmad Zainul Fanani: Conducted performance validation (Precision, Recall, F1-Score analysis), assisted in structuring tables and figures, and participated in the final proofreading and manuscript organization.

Ethics

This study is completely original and has not been published anywhere. The authors have all declared that they have no conflict of interest and have given their approval for the work to be published. All the information is contained in the article itself.

References

Alsmadi, M. K. (2017). An Efficient Similarity Measure for Content Based Image Retrieval Using Memetic Algorithm. *Egyptian Journal of Basic and Applied Sciences*, *4*(2), 112-122. https://doi.org/10.1016/j.ejbas.2017.02.004

Arya, V., Sharma, V., & Arya, G. (2019). An Efficient Adaptive Algorithm for Electron Microscopic Image Enhancement and Feature Extraction. *International Journal of Computer Vision and Image Processing*, *9*(1), 1-16. https://doi.org/10.4018/ijcvip.2019010101

Barburiceanu, S., Terebes, R., & Meza, S. (2021). 3D Texture Feature Extraction and Classification Using GLCM and LBP-Based Descriptors. *Applied Sciences*, *11*(5), 2332. https://doi.org/10.3390/app11052332

Basu, A., Paul, S., Ghosh, S., Das, S., Chanda, B., Bhagvati, C., & Snasel, V. (2023). Digital Restoration of Cultural Heritage With Data-Driven Computing: A Survey. *IEEE Access*, *11*, 53939-53977. https://doi.org/10.1109/access.2023.3280639

Chen, Y., Wang, R., Zong, P., & Chen, D. (2024). Image Processing for Denoising Using Composite Adaptive Filtering Methods Based on RMSE. *Open Journal of Applied Sciences*, *14*(03), 660-675. https://doi.org/10.4236/ojapps.2024.143047

Dzahini, K. J., & Wild, S. M. (2022). Hashing-like Johnson--Lindenstrauss Transforms and Their Extreme Singular Values. *ArXiv Preprint ArXiv:2212.14858*. https://doi.org/10.48550/arXiv.2212.14858

Ganesh Chandra, D., Siddesh, G. M., Srinivasa, K. G., & Patnaik, L. M. (2016). *Emerging Research Surrounding Power Consumption and Performance Issues in Utility Computing*. https://doi.org/10.4018/978-1-4666-8853-7

Hasenbusch, M., Pelissetto, A., & Vicari, E. (2008). ORB: An Efficient Alternative to SIFT or SURF Ethan. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(2), 2564-2571.

Hua, W., Hou, M., Qiao, Y., Zhao, X., Xu, S., & Li, S. (2021). Similarity Index Based Approach for Identifying Similar Grotto Statues to Support Virtual Restoration. *Remote Sensing*, *13*(6), 1201. https://doi.org/10.3390/rs13061201

Hua, W., Hou, M., Yang, S., & Dong, Y. (2019). Discrimination of Cultural Relics Similarity Based on Phash Algorithm and Sift Operator. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-2/W15*, 571-575. https://doi.org/10.5194/isprs-archives-xlii-2-w15-571-2019

Kabbai, L., Abdellaoui, M., & Douik, A. (2017). Hybrid Local and Global Descriptor Enhanced with Colour Information. *IET Image Processing*, *11*(2), 109-117. https://doi.org/10.1049/iet-ipr.2016.0349

Kanwal, K., Ahmad, K. T., Khan, R., Abbasi, A. T., & Li, J. (2020). Deep Learning Using Symmetry, FAST Scores, Shape-Based Filtering and Spatial Mapping Integrated with CNN for Large Scale Image Retrieval. *Symmetry*, *12*(4), 612. https://doi.org/10.3390/sym12040612

Kaur, A., & Sharma, R. (2013). Copy-Move Forgery Detection Using DCT and SIFT. *International Journal of Computer Applications*, *70*(7), 30-34. https://doi.org/10.5120/11977-7847

Kumar, P., Ofli, F., Imran, M., & Castillo, C. (2020). Detection of Disaster-Affected Cultural Heritage Sites from Social Media Images Using Deep Learning Techniques. *Journal on Computing and Cultural Heritage*, *13*(3), 1-31. https://doi.org/10.1145/3383314

Kumar, R. R. P., Muknahallipatna, S., & McInroy, J. (2016). An Approach to Parallelization of SIFT Algorithm on GPUs for Real-Time Applications. *Journal of Computer and Communications*, *04*(17), 18-50. https://doi.org/10.4236/jcc.2016.417002

Li, F., Chen, Y., Shi, Q., Shi, G., Yang, H., & Na, J. (2025). Improved Low-Light Image Feature Matching Algorithm Based on the SuperGlue Net Model. *Remote Sensing*, *17*(5), 905. https://doi.org/10.3390/rs17050905

Lin, C.-H., Lin, C.-J., & Wang, S.-H. (2024). Quality Assessment of Selective Laser Melting Components Using Quantitative Ultrasound Measurements and Image Texture Analysis. *IEEE Access*, *12*, 71797-71807. https://doi.org/10.1109/access.2024.3403142

Liu, X., & Huang, X. (2023). Gold, Skin, and Body: Chinese Buddha Statues Are Constantly Being Shaped and Stripped. *Religions*, *14*(2), 155. https://doi.org/10.3390/rel14020155

Marlinda, L., Budiman, F., Basuki, R. S., & Fanani, A. Z. (2023). Comparison of Sift and Orb Methods in Identifying the Face of Buddha Statue. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, *8*(2), 145-150. https://doi.org/10.33480/jitk.v8i2.4086

Marschner, C. B., Kokla, M., Amigo, J. M., Rozanski, E. A., Wiinberg, B., & McEvoy, F. J. (2017). Texture Analysis of Pulmonary Parenchymateous Changes Related to Pulmonary Thromboembolism in Dogs - A Novel Approach Using Quantitative Methods. *BMC Veterinary Research*, *13*(1), 219. https://doi.org/10.1186/s12917-017-1117-1

Mehta, T., & K., C. (2020). Near Duplicate Image Retrieval using Multilevel Local and Global Convolutional Neural Network Features. *International Journal of Advanced Computer Science and Applications*, *11*(4), 225-232. https://doi.org/10.14569/ijacsa.2020.0110430

M.Elsheh, M., & A.Eltomi, S. (2019). Content Based Image Retrieval Using Color Histogram and Discrete Cosine Transform. *International Journal of Computer Trends and Technology*, *67*(9), 25-31. https://doi.org/10.14445/22312803/ijctt-v67i9p105

Pavan Kumar, M. P., Poornima, B., Nagendraswamy, H. S., Manjunath, C., Rangaswamy, B. E., Varsha, M., & Vinutha, H. P. (2022). Image Abstraction Framework as a Pre-processing Technique for Accurate Classification of Archaeological Monuments Using Machine Learning Approaches. *SN Computer Science*, *3*(1), 87. https://doi.org/10.1007/s42979-021-00935-8

Prabukumar, M., Sawant, S., Samiappan, S., & Agilandeeswari, L. (2018). Three-Dimensional Discrete Cosine Transform-Based Feature Extraction for Hyperspectral Image Classification. *Journal of Applied Remote Sensing*, *12*(04), 046010. https://doi.org/10.1117/1.jrs.12.046010

Redaoui, A., & Belloulata, K. (2023). Deep Feature Pyramid Hashing for Efficient Image Retrieval. *Information*, *14*(1), 6. https://doi.org/10.3390/info14010006

Renoust, B., Oliveira Franca, M., Chan, J., Garcia, N., Le, V., Uesaka, A., Nakashima, Y., Nagahara, H., Wang, J., & Fujioka, Y. (2019). Historical and Modern Features for Buddha Statue Classification. *Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia HeritAge Contents*, 23-30. https://doi.org/10.1145/3347317.3357239

Sookkaew, J., & Chaikaew, N. (2022). Combining the Characteristics of the Buddha Statue from Photogrammetry and 3D Creation to Simulate the Combination of the Art of Creating Buddha Statue. *International Journal of Advanced Computer Science and Applications*, *13*(12), 186-191. https://doi.org/10.14569/ijacsa.2022.0131224

Sri, K. H., Manasa, G. T., Reddy, G. G., Bano, S., & Trinadh, V. B. (2022). Detecting Image Similarity Using SIFT. *Expert Clouds and Applications*, *209*, 561-575. https://doi.org/10.1007/978-981-16-2126-0_45

Sumaia Ali, E. (2019). *Content Based Color Images Retrieval Method using Discrete Cosine Transform and Histogram*.

Tabatabaei, S. M., & Chalechale, A. (2020). Noise-Tolerant Texture Feature Extraction Through Directional Thresholded Local Binary Pattern. *The Visual Computer*, *36*(5), 967-987. https://doi.org/10.1007/s00371-019-01704-8

Thomas, M. (2012). *Fiabilité, Maintenance Prédictive et Vibration des Machines*. https://doi.org/10.1515/9782760533585-014

Wang, S., & Liu, S. (2024). Dunhuang Buddhist Art Object Recognition Based on CenterNet. *IEEE Access*, *12*, 147762-147771. https://doi.org/10.1109/access.2024.3468722

Watanabe, U., & Abe, H. (2017). Sequencing and Quantifying Plastid DNA Fragments Stored in Sapwood and Heartwood of Torreya Nucifera. *Journal of Wood Science*, *63*(3), 201-208. https://doi.org/10.1007/s10086-017-1611-x

Xiangyang, L., Dongming, L., & Yunhe, P. (2000). Color Restoration and Image Retrieval for Dunhuang Fresco Preservation. *IEEE Multimedia*, *7*(2), 38-42. https://doi.org/10.1109/93.848425

Yao, C., Zhang, H., Zhu, J., Fan, D., Fang, Y., & Tang, L. (2023). ORB Feature Matching Algorithm Based on Multi-Scale Feature Description Fusion and Feature Point Mapping Error Correction. *IEEE Access*, *11*, 63808-63820. https://doi.org/10.1109/access.2023.3288594

Yugang, H., & Chunlei, W. (2022). Does Buddhist Tourism Successfully Result in Local Sustainable Development? *Sustainability*, *14*(6), 3584. https://doi.org/10.3390/su14063584

Zhang, K., Li, W., Yu, S., Chen, S., Zeng, Z., Ma, X., & Luo, Y. (2023). Similar Image Retrieval Algorithm Based on Feature Fusion and Locality-Sensitive Hash. *Design Studies and Intelligence Engineering*, *365*, 355-367. https://doi.org/10.3233/faia220729

Zhou, X., Luo, C., Ren, P., & Zhang, B. (2024). Multiscale Complex-Valued Feature Attention Convolutional Neural Network for SAR Automatic Target Recognition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 2052-2066. https://doi.org/10.1109/jstars.2023.3342986