

MARFE and ALSPH: a high-performance framework for scalable and efficient image deduplication

Rahul Shah^{1,2}, Ashok Kumar Shrivastava¹, Vivek Thapa³

¹Department of Computer Science and Engineering, Amity University Madhya Pradesh, Gwalior, India

²Faculty of Science and Technology, The ICFAI University Sikkim, Gangtok, India

³Department of Computer Applications, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Gangtok, India

Article Info

Article history:

Received Jul 11, 2025

Revised Aug 11, 2026

Accepted Aug 20, 2026

Keywords:

Adaptive locality-sensitive perceptual hashing
Feature extraction
Image deduplication
Multi-scale adaptive region-based feature extraction
Near-duplicate detection
Perceptual hashing
Scalability

ABSTRACT

Image deduplication has become increasingly important with the rapid growth of digital image repositories in cloud storage, multimedia systems, and large-scale visual databases. Existing approaches face a trade-off between detection accuracy and computational efficiency. Deep learning methods provide high robustness but require substantial computational resources, while traditional perceptual hashing and handcrafted features offer faster processing but have limited resilience to image transformations. To address this issue, this paper proposes an efficient image deduplication framework integrating multi-scale adaptive region-based feature extraction (MARFE) and adaptive locality-sensitive perceptual hashing (ALSPH). MARFE adaptively selects information-rich image regions to preserve discriminative visual characteristics while reducing redundant computations. ALSPH generates compact binary fingerprints through covariance-guided projection learning for efficient and robust similarity matching. The framework was evaluated on benchmark and web-scale image collections, with additional scalability analysis performed at dataset sizes of up to 10 million images. The framework achieved an F1-score of 0.979, processing speed of 187 images per second, and 49.3% storage savings, demonstrating a good balance between accuracy, efficiency, and scalability.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Rahul Shah

Department of Computer Science and Engineering, Amity University Madhya Pradesh

Gwalior, India

Email: sahrahul77@gmail.com

1. INTRODUCTION

The exponential growth of digital media has created massive volumes of image data, posing significant challenges for storage, management, and retrieval. A fundamental task is image deduplication, which identifies and removes redundant images to reduce storage requirements, improve content-based image retrieval (CBIR), and optimize cloud storage systems [1]-[3]. Image deduplication encompasses both exact duplicates, which are identical copies, and near-duplicates, which result from transformations such as resizing, cropping, rotation, illumination changes, or compression.

Traditional approaches based on color histograms, texture descriptors, and perceptual hashing are computationally efficient but often fail to detect transformed near-duplicate images reliably [4]-[6]. In contrast, deep learning methods, particularly convolutional neural network (CNN)-based models, provide higher robustness by learning semantic image representations but require substantial computational resources for training and inference, limiting their applicability in large-scale real-time environments [7].

Consequently, achieving an effective balance between detection accuracy, computational efficiency, and scalability remains an open research challenge.

To address this limitation, this paper proposes an efficient hybrid framework integrating multi-scale adaptive region-based feature extraction (MARFE) and adaptive locality-sensitive perceptual hashing (ALSPH). MARFE adaptively selects information-rich image regions to generate compact and discriminative feature representations, while ALSPH produces adaptive binary fingerprints for efficient similarity matching using data-driven projections [8], [9]. The proposed framework is designed to improve robustness against common image transformations while maintaining low computational overhead for large-scale image repositories [10].

The contributions of this work are: i) an efficient image deduplication framework integrating the proposed MARFE feature extraction and ALSPH hashing techniques; ii) a lightweight adaptive region-selection strategy for robust multi-scale feature extraction; iii) a data-adaptive perceptual hashing algorithm that generates compact and discriminative binary fingerprints for efficient similarity matching; and iv) comprehensive evaluation on benchmark and web-scale datasets containing up to 10 million images, demonstrating improved accuracy, processing efficiency, and storage savings compared with representative existing methods.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents the proposed MARFE-ALSPH framework. Section 4 describes the experimental setup and section 5 discusses the experimental results. Finally, section 6 concludes the paper and outlines future research directions.

2. RELATED WORK

Image deduplication has evolved from traditional handcrafted descriptors to deep learning-based image representations. Existing methods can be broadly categorized into exact duplicate detection, which identifies identical files using cryptographic hashes (e.g., MD5 and SHA-1), and near-duplicate detection, which identifies visually similar images affected by transformations such as resizing, cropping, rotation, illumination changes, and compression [1], [6]. This work focuses on near-duplicate image detection.

2.1. Traditional and perceptual hashing method

Traditional feature descriptors such as scale invariant feature transform (SIFT) [4], histograms of oriented gradients (HOG) [5], and local binary pattern (LBP) [11] provide interpretable image representations but are computationally expensive and sensitive to complex transformations. Perceptual hashing methods, including aHash, pHash, and dHash [6], generate compact fingerprints for efficient similarity search; however, their robustness decreases under significant geometric and photometric variations. Recent large-scale image deduplication studies have further highlighted the need for scalable and transformation-robust solutions [12]–[14].

2.2. Deep learning for image similarity

Deep learning approaches, including CNNs, residual networks (ResNet) [15], Siamese networks [16], deep ranking methods [17], FaceNet (facial recognition network) [18], simple framework for contrastive learning of visual representations (SimCLR) [19], and vision transformers (ViT) [20], [21], have substantially improved near-duplicate image detection by learning semantic feature representations. Deep hashing methods further integrate learned embeddings into compact binary codes [12], [22]. Although these approaches achieve high detection accuracy, they require considerable computational resources, memory, and graphics processing unit (GPU) acceleration, limiting their practicality for large-scale image repositories.

2.3. Comparative analysis

Table 1 summarizes representative traditional, deep learning, and hybrid image deduplication methods with respect to robustness, computational complexity, and scalability. Traditional methods provide fast retrieval but limited robustness, whereas deep learning methods improve semantic understanding at the expense of computational cost. Hybrid approaches offer improved trade-offs but often remain computationally demanding for web-scale deployment.

2.4. Research gap

Recent hybrid methods combine deep feature extraction with perceptual hashing to improve duplicate detection accuracy [12], [13], [22]. Despite these advances, existing approaches still struggle to balance transformation robustness, computational efficiency, and scalability. Most deep learning methods require expensive training and inference, while lightweight hashing techniques remain vulnerable to complex image transformations. To address these limitations, the proposed framework integrates MARFE, which

performs lightweight multi-scale adaptive feature extraction, with ALSPH, which generates adaptive binary fingerprints using data-driven projections without supervised training. This combination provides an efficient balance between detection accuracy, computational cost, and scalability for large-scale image deduplication.

Table 1. Comparative analysis of image deduplication methods

Category	Representative methods	Strengths	Limitations
Traditional feature-based	SIFT [4], HOG [5], LBP [11], and locality-sensitive hashing (LSH) [8]	Fast, interpretable, and low-resource	Poor robustness to transformations; high false positives
Deep learning-based	ResNet [15], FaceNet [18], SimCLR [19], ViT [20], [21]	High accuracy; captures semantic features	Computationally expensive; large storage; scalability issues
Hybrid approaches	Hamming embedding [23], perceptual hashing [6], and double hashing [13]	Reduced storage; improved speed; moderately robust	Limited scalability; weaker semantic awareness than deep embeddings
Proposed MARFE+ALSPH	Proposed work	Combines region-based features with adaptive hashing; high F1-score; fast and scalable	Requires minimum dataset size for ALSPH; higher memory than pHash

3. PROPOSED MARFE AND ALSPH FRAMEWORK

The proposed framework integrates MARFE and ALSPH into a modular image deduplication pipeline (Figure 1). Input images are resized and normalized before MARFE extracts discriminative multi-scale regional features. ALSPH subsequently transforms these features into compact binary fingerprints using covariance-guided adaptive projections. Candidate images are retrieved through LSH Forest indexing and verified using Hamming distance. Compared with conventional global indexing methods, the proposed framework improves computational efficiency while maintaining robustness against common image transformations.

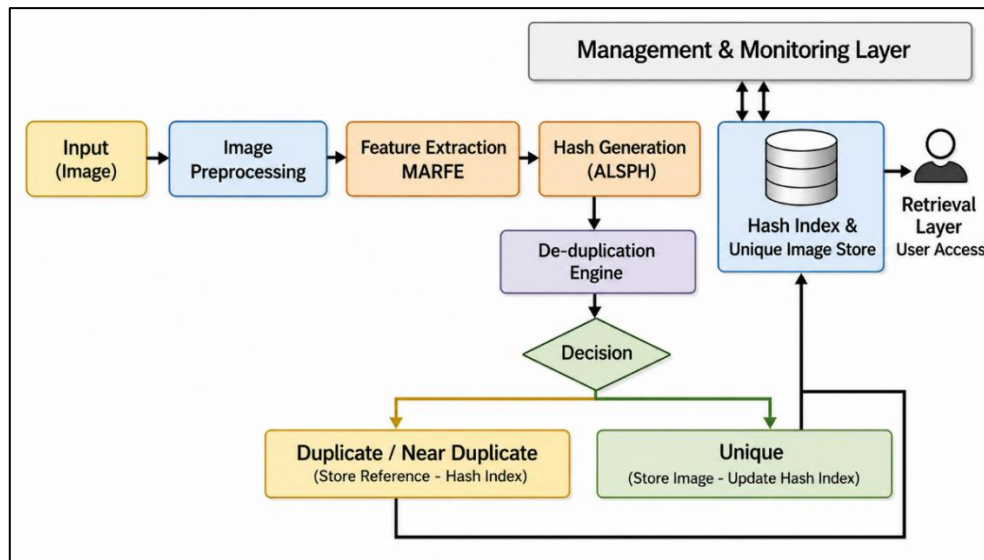


Figure 1. System architecture of the proposed image deduplication framework, integrating MARFE and ALSPH

3.1. Multi-scale adaptive region-based feature extraction

MARFE partitions each pre-processed image into 64 non-overlapping blocks and computes the average gradient magnitude of each block using Sobel operators. The blocks with the highest gradient responses are selected as structurally informative regions for feature extraction. Color moments, LBP [11], and discrete cosine transform (DCT) coefficients are extracted and concatenated into a high-dimensional feature vector v_{high} . Principal component analysis (PCA) [24] projects the feature vector into a compact representation;

$$v_{MARFE} = W v_{high}$$

where W denotes the learned projection matrix. The resulting feature vector is L2-normalized before hashing. Gradient magnitude is computed as:

$$M(x, y) = \sqrt{(I * G_x)^2 + (I * G_y)^2}$$

where G_x and G_y are Sobel filters.

Sensitivity analysis showed that adaptive region selection improves robustness against illumination variation, background clutter, and moderate geometric transformations while reducing computational cost. Empirical evaluation indicated that partitioning each image into 64 blocks provided the best trade-off between accuracy and efficiency. This adaptive strategy allocates computation to structurally informative regions while suppressing low-information background areas.

Algorithm 1. MARFE - multi-scale adaptive region-based feature extraction

Input: Image I

Output: Feature vector v_{MARFE}

1. Resize and normalize the input image.
2. Partition the image into 64 non-overlapping blocks.
3. Compute Sobel gradient magnitude for each block.
4. Select blocks with the highest average gradient response.
5. Extract Color Moments, LBP and DCT features.
6. Concatenate descriptors into v_{high} .
7. Apply PCA for dimensionality reduction.
8. Normalize the feature vector.
9. Output v_{MARFE} .

3.2. Adaptive locality-sensitive perceptual hashing

ALSPH converts MARFE feature vectors into compact binary fingerprints using covariance-guided adaptive projections. Unlike conventional LSH that employs fixed or random projections, ALSPH learns projection vectors from the statistical characteristics of the MARFE feature space. For a set of projected feature vectors $\{v_{MARFE}^{(i)}\}_{i=1}^N$, the covariance matrix is computed as:

$$C_{MARFE} = \frac{1}{N} \sum_{i=1}^N (v_{MARFE}^{(i)} - \mu)(v_{MARFE}^{(i)} - \mu)^T$$

where μ is the mean feature vector.

Eigen-decomposition of C_{MARFE} yields the principal projection vectors. Each feature vector is projected and converted into an L -bit binary hash according to:

$$h_i = \begin{cases} 1, & v_{MARFE} \cdot p_i \geq 0 \\ 0, & otherwise \end{cases}$$

where p_i represents the selected projection vector.

This covariance-guided projection preserves the intrinsic variance structure of the feature space, improving similarity preservation, reducing hash collisions, and generating compact transformation-robust fingerprints suitable for scalable image deduplication. Batch-wise adaptation enables stable projection learning while maintaining computational efficiency.

Algorithm 2. ALSPH - adaptive locality-sensitive perceptual hashing

Input: MARFE feature vectors

Output: Binary hash code

1. Compute covariance matrix of MARFE feature vectors.
2. Perform eigen-decomposition.
3. Select principal projection vectors.
4. Project feature vectors.
5. Apply binary quantization.
6. Generate final L -bit hash.

3.3. Deduplication engine

The deduplication engine performs duplicate identification using the binary fingerprints generated by MARFE and ALSPH. A LSH-based indexing approach is employed for efficient approximate nearest-neighbour search in high-dimensional feature spaces [25]. For each incoming image, the generated fingerprint is queried against the index to retrieve candidate duplicates. A Hamming distance comparison is then performed on the candidate set to confirm duplicate matches. Unique images are stored using a content-addressable storage mechanism, whereas duplicate images are represented through lightweight reference pointers, thereby reducing storage requirements while maintaining efficient retrieval [26].

3.4. Theoretical framework and mathematical formulation

The proposed MARFE-ALSPH framework combines adaptive feature selection with data-driven projection learning to achieve robust and scalable image deduplication. MARFE identifies perceptually significant image regions based on gradient magnitude, preserving discriminative structural information while reducing redundant feature extraction in high-dimensional feature spaces [27]. The extracted descriptors are transformed into a compact representation using PCA, defined as:

$$v_{\text{MARFE}} = W^T v_{\text{high}}$$

where v_{high} denotes the aggregated feature vector and W represents the PCA projection matrix [24].

ALSPH learns projection vectors from the covariance structure of the MARFE feature space, allowing statistically similar images to generate similar binary fingerprints. The binary hash is obtained as:

$$h_i = \begin{cases} 1 & \text{if } v_{\text{marfe}} \cdot p_i \geq 0 \\ 0 & \text{otherwise} \end{cases}, \quad i = 1, \dots, L$$

where p_i denotes the i^{th} adaptive projection vector. This covariance-guided projection preserves the intrinsic variance of the feature space while reducing hash collisions and improving robustness to common image transformations.

4. EXPERIMENTAL SETUP

To ensure a systematic and reproducible evaluation of the proposed MARFE-ALSPH framework, experiments were conducted using a standardized implementation and evaluation protocol on benchmark and large-scale image datasets. The evaluation was designed to assess duplicate detection accuracy, computational efficiency, scalability, and storage optimization under realistic operating conditions. The framework was implemented using Python 3.9 with OpenCV, NumPy, and PyTorch 1.9.0 on Ubuntu 20.04 LTS. Experiments were performed on a workstation equipped with an Intel Xeon multi-core processor, 128 GB RAM, and an NVIDIA Tesla V100 GPU for preprocessing and baseline deep learning comparisons. Parallel processing tools were employed to improve large-scale data processing efficiency, while a structured database system was used to maintain experimental logs and metadata.

4.1. Datasets

The proposed framework was evaluated using benchmark and large-scale image datasets containing exact and near-duplicate images with variations in viewpoint, illumination, scaling, compression, and rotation. The datasets include INRIA Holidays, Oxford Buildings, UKBench, YFCC100M, and NDID, providing diverse visual content and large-scale retrieval scenarios. The complete dataset summary is provided in Table 2.

Table 2. Summary of key datasets used for evaluation

Dataset name	No. of images	Key characteristics and purpose
INRIA holidays [23]	1,491	High-quality photos with realistic transformations for near-duplicate detection.
Oxford buildings [28]	5,062	Architectural images with viewpoint and illumination changes.
UKBench [29]	10,200	Object-centric images in groups of four for similarity evaluation.
YFCC100M Subset [30]	1,000,000	Web-scale dataset with diverse real-world content.
NDID	10,000	Tests near-duplicate detection with cropping, resizing, and transformations.

4.1.1. Pre-processing

All images were resized to 256×256 pixels, converted to the Lab colour space, and normalized before feature extraction. MARFE extracted multi-scale regional descriptors, while ALSPH generated adaptive binary fingerprints using batch-wise covariance estimation.

4.2. Evaluation metrics

Performance was evaluated using precision, recall, F1-score, deduplication rate (DR), processing throughput (images/s), latency (ms/image), and storage saving (%). Each experiment was repeated multiple times, and the reported results represent the average performance under identical experimental conditions.

4.3. Comparative analysis

The performance of the proposed MARFE-ALSPH framework was compared with representative state-of-the-art image deduplication methods, including SIFT+LSH, DeepDup, VisualDNADedupe, and ImgDedupNet. The comparison focused on duplicate detection accuracy, processing speed, storage savings, and scalability. The results demonstrate that the proposed framework achieves an effective balance between detection accuracy, computational efficiency, and scalability for large-scale image deduplication.

The comparative results demonstrate that the proposed MARFE-ALSPH framework achieves the highest overall performance among the evaluated methods. The framework obtains an F1-score of 0.979, a processing speed of 187 images/s, and 49.3% storage savings, while also achieving the highest scalability score (0.92). Compared with traditional feature-based methods such as SIFT+LSH and recent deep learning-based approaches including DeepDup, VisualDNADedupe, and ImgDedupNet, the proposed framework provides a better balance between duplicate detection accuracy, computational efficiency, storage optimization, and scalability, making it suitable for large-scale image deduplication applications.

4.4. Ablation study

To evaluate the contribution of the proposed components, ablation experiments were conducted by analysing the performance of MARFE feature extraction and ALSPH hashing individually and in combination. The results demonstrate that MARFE improves the discriminative quality of image features, while ALSPH enhances the robustness and efficiency of binary fingerprint generation. The combined MARFE-ALSPH framework consistently achieves the highest duplicate detection performance, confirming the complementary nature of the two proposed components.

4.5. Scalability and limitations

Although the proposed MARFE-ALSPH framework demonstrates strong performance for large-scale image deduplication, several limitations remain. Performance may decrease under extreme geometric transformations and complex overlay-based modifications such as logos and watermarks. In addition, the effectiveness of ALSPH depends on representative covariance estimation during adaptive projection learning, while computational performance is influenced by the underlying hardware configuration. Future work will focus on improving robustness to complex transformations and extending the framework to video and multimodal deduplication applications.

5. RESULTS AND DISCUSSION

This section presents the experimental results of the proposed MARFE-ALSPH framework. The evaluation focuses on duplicate detection accuracy, computational efficiency, storage optimization, and scalability using benchmark and large-scale image datasets. The proposed framework is compared with representative state-of-the-art image deduplication methods to demonstrate its effectiveness under realistic operating conditions.

5.1. Overall performance

The comparative performance of the proposed MARFE-ALSPH framework is summarized in Table 3. The proposed framework achieved the highest F1-score (0.979), processing speed (187 images/s), storage savings (49.3%), and scalability score (0.92) among the evaluated methods. Compared with SIFT+LSH, DeepDup, VisualDNADedupe, and ImgDedupNet, the proposed framework provides a balanced trade-off between duplicate detection accuracy, computational efficiency, storage optimization, and scalability. These results demonstrate that the combination of adaptive region-based feature extraction and covariance-guided perceptual hashing effectively improves duplicate detection while maintaining efficient large-scale processing.

Table 3. MARFE+ALSPH performance comparison with state-of-the-art methods

Method	F1-score	Processing speed (img/s)	Storage savings (%)	Scalability score
SIFT+LSH [4], [8]	0.945	132	41.7	0.85
DeepDup	0.968	95	47.2	0.79
VisualDNADedupe	0.953	156	44.6	0.88
ImgDedupNet	0.971	108	46.8	0.83
Proposed MARFE+ALSPH	0.979	187	49.3	0.92

5.2. Transformation robustness

Figure 2 compares the robustness of the proposed ALSPH hashing algorithm with conventional perceptual hashing methods, including pHash, aHash, and dHash, under common image transformations. The proposed ALSPH consistently achieves the highest performance across rotation, scaling, compression, brightness variation, and cropping. The improvement is particularly evident under geometric and photometric transformations, demonstrating that the adaptive covariance-guided hashing strategy generates more stable and discriminative binary fingerprints than conventional perceptual hashing techniques. These results confirm that ALSPH provides improved robustness while maintaining computational efficiency for practical image deduplication applications.

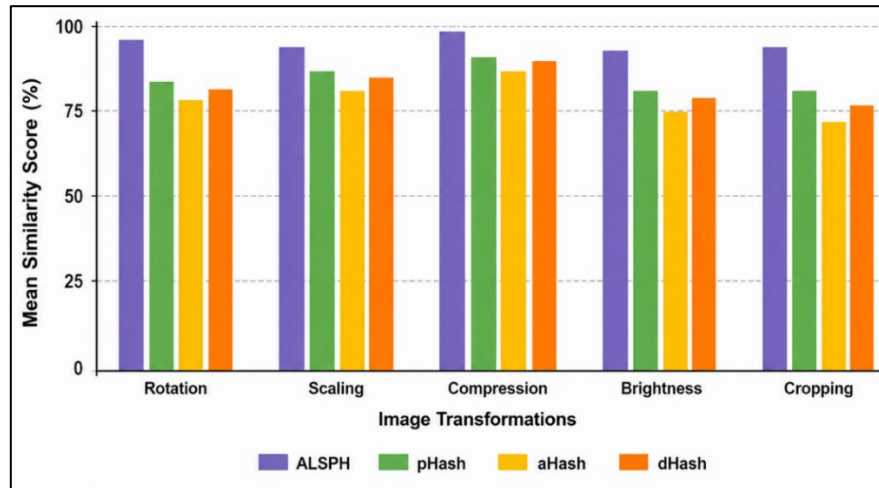


Figure 2. Robustness of ALSPH to image transformations compared to other hashing methods

5.3. Scalability analysis

The scalability of the proposed framework was evaluated using datasets ranging from 10,000 to 10 million images. As illustrated in Figure 3, the processing time increases with dataset size, with a more pronounced increase at larger dataset scales, while the processing throughput remains relatively stable. The F1-score decreases only marginally from 0.981 to 0.975, while the processing speed decreases from 195 to 179 images/s because of the increased indexing and memory overhead associated with larger datasets. Furthermore, storage savings remain nearly constant, varying from 49.7% to 48.8%, indicating consistent storage optimization across different dataset sizes. These results demonstrate that the proposed framework scales efficiently while maintaining robust duplicate detection performance.

5.4. Discussion

The experimental results demonstrate that the proposed MARFE-ALSPH framework successfully addresses the trade-off between duplicate detection accuracy and computational efficiency in large-scale image deduplication. MARFE improves feature representation by focusing on structurally informative image regions, while ALSPH enhances perceptual hashing through adaptive covariance-guided projection learning. The complementary integration of these components enables accurate duplicate detection while reducing computational overhead and storage requirements. Compared with existing image deduplication approaches that focus on efficient visual indexing, retrieval, and resource optimization [31], the proposed framework achieves competitive detection performance with significantly improved processing efficiency and scalability. These characteristics make the framework suitable for practical deployment in applications such as cloud storage systems, digital asset management, multimedia repositories, and large-scale image retrieval.

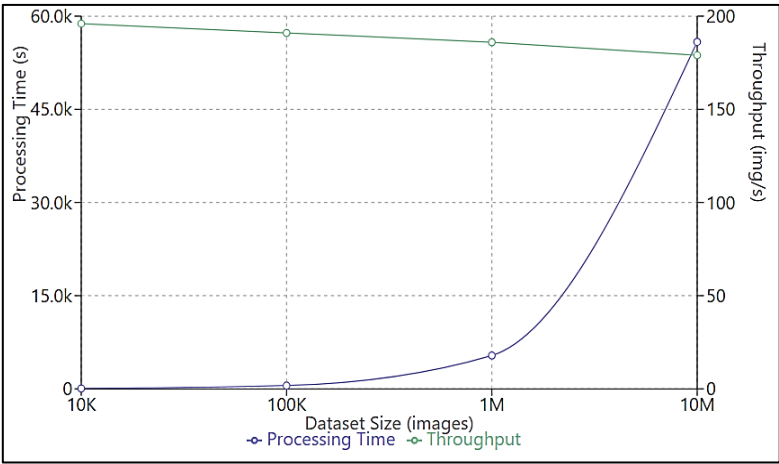


Figure 3. MARFE+ALSPH engine processing speed and scalability

6. CONCLUSION

This paper introduced an efficient and scalable image deduplication framework based on the novel MARFE feature extraction and ALSPH hashing techniques. By intelligently balancing accuracy and computational efficiency, the proposed system addresses a critical challenge in large-scale data management. Experimental results provide strong evidence of its superior performance, achieving a state-of-the-art F1-score of 0.979 while demonstrating significantly faster processing and greater storage efficiency compared to existing methods. The adaptive, data-driven design of its components ensures robustness and versatility across a wide range of image types and transformations.

Future work could explore extending this framework to other media types like video, further optimizing the algorithms for real-time stream processing, and investigating the use of hardware acceleration for even greater performance. The MARFE+ALSPH system represents a significant step forward, offering a practical, high-performance solution for managing the ever-growing repositories of digital images in real-world applications.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Rahul Shah	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Ashok Kumar	✓	✓		✓	✓		✓			✓		✓		
Shrivastava														
Vivek Thapa		✓	✓	✓	✓	✓	✓	✓		✓	✓			

C : Conceptualization	I : Investigation	Vi : Visualization
M : Methodology	R : Resources	Su : Supervision
So : Software	D : Data Curation	P : Project administration
Va : Validation	O : Writing - Original Draft	Fu : Funding acquisition
Fo : Formal analysis	E : Writing - Review & Editing	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available on request from the corresponding author, [Rahul Shah – sahratul77@gmail.com] upon reasonable request.

REFERENCES

- [1] W. Xia *et al.*, “A comprehensive study of the past, present, and future of data deduplication,” in *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1681–1710, 2016, doi: 10.1109/JPROC.2016.2571298.
- [2] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-based image retrieval at the end of the early years,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000, doi: 10.1109/34.895972.
- [3] R. Datta, D. Joshi, J. Li, and J. Z. Wang, “Image retrieval: Ideas, influences, and trends of the new age,” *ACM Computing Surveys*, vol. 40, no. 2, pp. 1–60, Apr. 2008, doi: 10.1145/1348246.1348248.
- [4] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/B:VISI.0000029664.99615.94.
- [5] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, San Diego, CA, USA: IEEE, 2005, pp. 886–893, doi: 10.1109/CVPR.2005.177.
- [6] X. Chen, C. Qin, and P. Ji, “Perceptual image hashing using block truncation coding and local binary pattern,” in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, Hong Kong, China, 2015, pp. 856–859, doi: 10.1109/APSIPA.2015.7415392.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Advances in Neural Information Processing Systems*, vol. 25, 2012.
- [8] A. Gionis, P. Indyk, and R. Motwani, “Similarity Search in High Dimensions via Hashing,” in *VLDB 1999 - Proceedings of the 25th International Conference on Very Large Data Bases*, 1999, pp. 518–529.
- [9] H. Jégou, F. Perronnin, M. Douze, J. Sánchez, P. Pérez, and C. Schmid, “Aggregating local image descriptors into compact codes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1704–1716, 2012, doi: 10.1109/TPAMI.2011.235.
- [10] B. Cao, A. Araujo, and J. Sim, “Unifying Deep Local and Global Features for Image Search,” in *European Conference on Computer Vision*, pp. 726–743, 2020, doi: 10.1007/978-3-030-58565-5_43.
- [11] T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, Jul. 2002, doi: 10.1109/TPAMI.2002.1017623.
- [12] Y. Jakhar and M. D. Borah, “Effective near-duplicate image detection using perceptual hashing and deep learning,” *Information Processing and Management*, vol. 62, no. 4, p. 104086, Jul. 2025, doi: 10.1016/j.ipm.2025.104086.
- [13] P. Subudhi and K. Kumari, “A fast and efficient large-scale near duplicate image retrieval system using double perceptual hashing,” *Signal, Image and Video Processing*, vol. 18, no. 12, pp. 8565–8575, Dec. 2024, doi: 10.1007/s11760-024-03490-w.
- [14] Y. K. Adimoolam, C. Poullis, and M. Averkiou, “Data Leakage Detection and De-duplication in Large Scale Geospatial Image Datasets,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026, pp. 72–81.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- [16] R. Hadsell, S. Chopra, and Y. LeCun, “Dimensionality reduction by learning an invariant mapping,” in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, IEEE, 2006, pp. 1735–1742, doi: 10.1109/CVPR.2006.100.
- [17] J. Wang *et al.*, “Learning fine-grained image similarity with deep ranking,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1386–1393, 2014, doi: 10.1109/CVPR.2014.180.
- [18] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A unified embedding for face recognition and clustering,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2015, pp. 815–823, doi: 10.1109/CVPR.2015.7298682.
- [19] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A Simple Framework for Contrastive Learning of Visual Representations,” in *International Conference on Machine Learning*, 2020, pp. 1597–1607.
- [20] A. Dosovitskiy *et al.*, “An Image Is Worth 16×16 Words: Transformers for Image Recognition at Scale,” in *International Conference on Learning Representations (ICLR)*, 2021.
- [21] M. Caron *et al.*, “Emerging Properties in Self-Supervised Vision Transformers,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9630–9640, 2021, doi: 10.1109/ICCV48922.2021.00951.
- [22] K. Lin, H. F. Yang, J. H. Hsiao, and C. S. Chen, “Deep learning of binary hash codes for fast image retrieval,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 27–35, doi: 10.1109/CVPRW.2015.7301269.
- [23] H. Jégou, M. Douze, and C. Schmid, “Hamming embedding and weak geometric consistency for large scale image search,” in *European Conference on Computer Vision (ECCV 2008)*, Berlin, Heidelberg: Springer, 2008, pp. 304–317, doi: 10.1007/978-3-540-88682-2_24.
- [24] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer, p. 488, 2002, doi: 10.1007/b98835.
- [25] P. Indyk, *Nearest neighbors in high-dimensional spaces*, in *Handbook of Discrete and Computational Geometry*, 2nd ed., J. E. Goodman and J. O’Rourke, Eds. Boca Raton, FL, USA: Chapman & Hall/CRC, 2004, pp. 877–892, doi: 10.1201/9781420035315.ch39.
- [26] L. Morra and F. Lamberti, “Benchmarking unsupervised near-duplicate image detection,” *Expert Systems with Applications*, vol. 135, pp. 313–326, Nov. 2019, doi: 10.1016/j.eswa.2019.05.002.
- [27] J. H. Friedman, “On bias, variance, 0/1-loss, and the curse-of-dimensionality,” *Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 55–77, Mar. 1997, doi: 10.1023/A:1009778005914.
- [28] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, “Object retrieval with large vocabularies and fast spatial matching,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, 2007, pp. 1–8, doi: 10.1109/CVPR.2007.383172.

- [29] D. Nistér and H. Stewénus, "Scalable recognition with a vocabulary tree," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, New York, NY, USA, 2006, pp. 2161–2168, doi: 10.1109/CVPR.2006.264.
- [30] B. Thomee *et al.*, "YFCC100M: The new data in multimedia research," *Communications of the ACM*, vol. 59, no. 2, pp. 64–73, 2016, doi: 10.1145/2812802.
- [31] M. M. M. Rahman, D. Biswas, X. Chen, and J. Tešić, "Image deduplication using efficient visual indexing and retrieval: optimizing storage, time and energy for deep neural network training," *Signal, Image and Video Processing*, vol. 18, no. 12, pp. 9495–9503, Dec. 2024, doi: 10.1007/s11760-024-03562-x.

BIOGRAPHIES OF AUTHORS



Mr. Rahul Shah    is working as an Assistant Professor in the School of Information Technology at ICFAI University, Sikkim. He holds a Master of Computer Applications (MCA) degree from Shri Ramaswamy Memorial University, Sikkim, obtained in 2017, and is currently pursuing his Ph.D. at the Amity School of Engineering and Technology, Amity University Madhya Pradesh, Gwalior. He has more than 9 years of teaching experience and has authored 9 Scopus-indexed and 5 Web of Science (WoS)-indexed research papers in the fields of computer science and emerging technologies. He has also actively participated in numerous international and national conferences. His research interests include image processing, artificial intelligence, machine learning, and cloud computing. He can be contacted at email: sahrul77@gmail.com.



Dr. Ashok Kumar Shrivastava    working as an Associate Professor and academic coordinator in the Department of Computer Science and Engineering at Amity School of Engineering and Technology, Amity University Madhya Pradesh, Gwalior. He graduated in Mathematics from Science College Gwalior. He has done Master of Science in CS branch from Jawaji University, Gwalior. He secured M.Tech. CSE from Uttar Pradesh Technical University, Lucknow. He earned his Ph.D. in Image Processing from NIU, Greater Noida. He is in teaching profession for more than 21 years. He has published 14 Scopus indexed, 2 SCI and 13 Google Scholar papers in national and international journals, conference and symposiums. He has published 4 patents, 3 copyrights, and 2 books. His main area of interest includes noise detection and removal, image enhancement, and machine learning. He can be contacted at email: ashok79.shrivastava@gmail.com.



Mr. Vivek Thapa    works as an Assistant Professor in the Department of Computer Applications at the Sikkim Manipal Institute of Technology, Sikkim Manipal University. He holds a Master of Computer Applications (MCA) degree from the same institute, obtained in 2016. He has two years of industry experience and over seven years of teaching experience at multiple reputed universities. His research interests span computer vision, natural language processing, application development, cyber security, and algorithmic analysis, reflecting his passion for exploring cutting-edge technologies. He can be contacted at email: vivek.t@smit.smu.edu.in.