

AN IMPROVED HYBRID CLUSTERING ALGORITHM FOR SEMANTIC IMAGE MINING: DESIGN AND PERFORMANCE EVALUATION

Varalakshmi Sreenivasula & Shailendra Kumar

*Department of Computer Science and Engineering, School of Engineering and Technology, K K University, Nalanda,
Bihar, India*

ABSTRACT

The exponential growth of digital image repositories has intensified the need for clustering techniques that can organize visual content according to underlying semantic structure rather than raw pixel similarity alone. Classical clustering algorithms such as K-Means, Fuzzy C-Means (FCM), DBSCAN, and Agglomerative Hierarchical Clustering each present well-documented shortcomings when applied to high-dimensional, noisy, and unevenly distributed image feature spaces. This paper proposes an Improved Hybrid Clustering Algorithm (IHCA) for semantic image mining that combines density-aware initialization, fuzzy membership refinement, and a metaheuristic cluster-optimization stage to overcome the sensitivity to initialization, fixed-parameter rigidity, and poor handling of arbitrarily shaped clusters that limit existing methods. The proposed pipeline integrates deep convolutional feature extraction with handcrafted color-texture descriptors, applies dimensionality reduction to mitigate the curse of dimensionality, and fuses DBSCAN-based density estimation with a Fuzzy-K-Means core whose centroids are refined through a Particle Swarm Optimization (PSO) search. The hybrid model is evaluated on three benchmark image collections (Corel-1000, Caltech-101 subset, and a custom social-media image corpus) against four baseline algorithms. Experimental results, measured using Accuracy, Precision, Recall, F1-Score, Silhouette Coefficient, Davies-Bouldin Index, and execution time, show that IHCA achieves an average accuracy improvement of 9.4 percent and a 21 percent reduction in the Davies-Bouldin Index relative to the strongest baseline, while incurring a modest and predictable increase in execution time. These findings validate the hybrid design as a practical foundation for semantic image mining applications such as content-based image retrieval, automated tagging, and large-scale visual archive organization.

KEYWORDS: *Semantic Image Mining, Hybrid Clustering, Feature Extraction, Cluster Optimization, Particle Swarm Optimization, Content-Based Image Retrieval, Davies-Bouldin Index, Silhouette Coefficient*

Article History

Received: 10 Aug 2023 | Revised: 16 Aug 2023 | Accepted: 20 Aug 2023
