International Journal Of Engineering Research And Development
e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 21, Issue 1 (January, 2025), PP 291-295

A Review on Recent Trends on Content-based Image Retrieval for Remote Sensing Images

Nidhi Prajapati, Naimish Patel
[1]*Computer Engineering Department*
Silver Oak University, Ahmedabad, INDIA
Corresponding Author: Nidhi Prajapati

Abstract
Content-based image retrieval (CBIR) systems have emerged as a cornerstone in managing and analyzing remote sensing data, particularly for Land Use and Land Cover (LULC) datasets. The increasing availability of high-resolution satellite imagery necessitates efficient and precise indexing, searching, and retrieving relevant data to support environmental monitoring, urban planning, and disaster management. This survey comprehensively overviews state-of-the-art CBIR methodologies tailored for LULC applications. We examine advancements in feature extraction techniques, ranging from traditional texture-based descriptors to deep learning approaches, and compare their effectiveness on prominent benchmark datasets, including UC Merced and BigEarthNet. Challenges such as dataset heterogeneity, scalability, annotation bottlenecks, and retrieval accuracy are critically evaluated. Additionally, the paper discusses current trends, including multimodal data integration and active learning, and highlights emerging opportunities for explainable AI and hybrid feature models. The synthesis of findings offers actionable insights and proposes novel directions to advance CBIR systems in remote sensing.

Date of Submission: 14-01-2025 Date of acceptance: 28-01-2025

I. INTRODUCTION

The rapid advancement of remote sensing technologies, coupled with the widespread availability of satellite and aerial imagery, has significantly enhanced our ability to monitor and analyze Earth's surface. This unprecedented growth has led to the generation of vast amounts of geospatial data, often organized into Land Use and Land Cover (LULC) datasets. These datasets serve as invaluable resources for numerous global applications, ranging from urban development and environmental conservation to disaster management and agricultural planning. Despite their potential, such data's sheer volume and complexity present significant challenges in storage, indexing, and retrieval, making traditional text-based search techniques inadequate. Consequently, Content-Based Image Retrieval (CBIR) systems have emerged as indispensable tools, enabling the efficient querying of remote-sensing images based on their visual content rather than textual metadata

The CBIR systems leverage features extracted from the visual properties of images, such as color, texture, shape, or higher-order attributes, to perform similarity-based retrieval. Early CBIR techniques focused on handcrafted feature extraction methods, including Local Binary Patterns (LBP), Scale-Invariant Feature Transform (SIFT), and texture descriptors derived from Gray-Level Co-Occurrence Matrices (GLCM). While effective for simple retrieval tasks, these methods often struggle with the high intra-class variability and multi-spectral nature of remote sensing imagery. Furthermore, they require domain expertise and significant manual intervention for effective deployment.

The advent of deep learning has revolutionized the CBIR field, enabling automated feature learning and significantly enhancing retrieval accuracy and scalability. Pre-trained Convolutional Neural Networks (CNNs) and transfer learning approaches have shown exceptional performance in capturing spatial and spectral intricacies of LULC datasets. These advancements have opened up new possibilities for large-scale image retrieval, including the integration of multiple modalities such as hyperspectral and LiDAR data. However, despite these advancements, challenges such as high computational costs, the requirement for extensive labeled data, and issues related to model interpretability and generalization remain unresolved.

Another critical factor in the development of CBIR systems for remote sensing is the diversity and complexity of datasets. Benchmark datasets such as UC Merced, BigEarthNet, and Sentinel images vary widely in resolution, coverage, and data modalities, making it challenging to design universally applicable retrieval systems. These datasets often face issues of class imbalance, noise, and insufficient annotation, further complicating the retrieval process. As a result, researchers have explored various strategies, including active learning, relevance feedback mechanisms, and hybrid feature models, to address these limitations.

In this survey, we provide a comprehensive review of CBIR methodologies for remote sensing images, focusing on their application to LULC datasets. The survey explores the evolution of feature extraction techniques, from traditional descriptors to cutting-edge deep learning approaches. It also examines prominent datasets and benchmarks, along with the metrics used to evaluate retrieval performance. In addition, we discuss emerging trends and persistent challenges, such as the integration of explainable artificial intelligence (XAI) and the need for efficient hybrid retrieval models. By identifying research gaps and opportunities, we propose novel approaches to improve CBIR systems for remote sensing applications.

This work contributes to the understanding of how CBIR techniques can effectively support the management of LULC datasets and paves the way for advancements in intelligent remote sensing systems. By addressing existing challenges and leveraging modern computational techniques, CBIR systems can become more accurate, efficient, and accessible, unlocking the full potential of remote sensing imagery for global benefit.

1. Content-Based Image Retrieval (CBIR) for Remote Sensing

Content-Based Image Retrieval (CBIR) systems have emerged as indispensable tools for efficiently managing and analyzing the growing volume of remote sensing data. These systems index and retrieve images based on their visual content, such as textures, shapes, colors, or more abstract patterns, making them far more effective than traditional metadata-based approaches for large and diverse datasets like LULC (Land Use and Land Cover).

Remote sensing presents unique challenges for CBIR due to the complex characteristics of the imagery. These datasets include multi-spectral, multi-temporal, and high-resolution images with substantial heterogeneity. CBIR systems for remote sensing must account for inter-class similarity (e.g., different types of vegetation) and intra-class variability (e.g., varying appearances of urban areas). This has driven the development of specialized algorithms that focus on feature extraction, similarity matching, and adaptive indexing methods.

Initial CBIR systems relied on handcrafted features, such as texture (e.g., LBP, GLCM) and spectral characteristics. However, with advancements in machine learning, modern systems are predominantly deep learning-based, leveraging Convolutional Neural Networks (CNNs) to extract robust multi-scale features. These advancements have broadened CBIR's capabilities, enabling rapid and precise retrieval for various geospatial applications, such as disaster response, urban development, and ecological studies.

2. Trends in Feature Representation

Feature representation is a critical component of CBIR systems, dictating image retrieval processes' efficiency, accuracy, and scalability. Over the years, advancements in feature representation have transformed CBIR systems, evolving from simple handcrafted techniques to sophisticated automated and hybrid approaches capable of handling the diverse complexities of remote sensing imagery.

2.1. Handcrafted Features

Handcrafted features were the foundation of early CBIR systems and remain relevant in specific scenarios where interpretability and computational simplicity are important. These features primarily focus on capturing the low-level visual attributes of an image, such as texture, shape, and color.

- **Texture Descriptors**: Techniques like the Gray-Level Co-occurrence Matrix (GLCM) and Local Binary Patterns (LBP) extract spatial patterns that describe the arrangement of pixel intensities. These methods are particularly useful for distinguishing between homogeneous classes, such as water bodies or agricultural fields.[1]
- **Shape Descriptors**: Features derived from edges and contours, often using algorithms like Canny edge detection, play a role in identifying structural elements such as roads, buildings, and urban layouts.[2]
- **Spectral Features**: In remote sensing, spectral data across multiple bands is critical. Features like band ratios and vegetation indices (e.g., NDVI) are commonly used to classify LULC categories.

Despite their utility, handcrafted features are limited by their inability to capture complex spatial and spectral relationships, especially in heterogeneous or high-resolution datasets.

2.2. Mid-Level Features

The advent of mid-level features marked a shift from low-level pixel-based analysis to more abstract representations that encode patterns across regions or objects within an image. One widely used approach in this category is the **Bag-of-Visual-Words (BoVW)** model.[3]

- **Bag-of-Visual-Words**: BoVW represents images as collections of visual words, analogous to how text documents are represented in natural language processing. Texture or gradient-based features are quantized into a visual vocabulary, and images are described using histograms of these words. This method proved effective for scene classification tasks in LULC datasets.

- **Spatial Pyramid Matching**: An extension of BoVW, spatial pyramid matching incorporates spatial information into the visual word histograms, enhancing the system's ability to distinguish between scenes with similar textures but differing spatial arrangements.[4]

Mid-level approaches, while improving retrieval accuracy, often rely on carefully designed dictionaries and fail to generalize well across datasets with varying resolutions or acquisition conditions.

2.3. Deep Learning-Based Features

Deep learning has revolutionized feature representation in CBIR, offering automated methods to learn hierarchical representations directly from image data. Convolutional Neural Networks (CNNs) have become the dominant choice for feature extraction, with models pre-trained on large-scale datasets like ImageNet often fine-tuned for remote sensing applications.

- **Hierarchical Feature Learning**: CNNs extract features at multiple levels of abstraction. Early layers capture basic patterns like edges and corners, while deeper layers encode complex structures, enabling the identification of high-level semantic concepts.[5]
- **Transfer Learning**: Given the scarcity of labeled LULC datasets, transfer learning is widely adopted, where models pre-trained on generic datasets are fine-tuned on domain-specific imagery. This significantly reduces computational costs and training time while improving retrieval performance.
- **Self-Supervised Learning**: Emerging methods in self-supervised learning allow models to learn features without extensive labeled data. Techniques like contrastive learning use inherent image properties (e.g., spatial coherence) to create pseudo-labels, making them highly suitable for large-scale remote sensing datasets.[6]

  To summarize, Deep learning-based features are highly effective but come with challenges, including high computational demands, a need for large datasets, and limited interpretability.

2.4. Hybrid Approaches

Hybrid methods combine the strengths of handcrafted and deep learning-based features to achieve robust performance across diverse remote sensing scenarios.

- **Feature Fusion**: Handcrafted features, such as texture and spectral indices, are combined with deep learning features to enhance the representation. For instance, Local Binary Patterns (LBP) can complement CNN-derived features by capturing fine-grained textures that CNNs may overlook.
- **Multimodal Integration**: Hybrid systems can integrate data from multiple sources, such as combining LiDAR data with spectral imagery, or temporal data with spatial features. This approach is particularly beneficial for LULC applications where data from diverse sensors is available.[7]

Hybrid approaches mitigate the weaknesses of individual methods, providing a balance between computational efficiency, generalization, and interpretability.

2.5. Contextual and Semantic Features

Feature representation is a critical component of CBIR systems, dictating the efficiency, accuracy, and scalability of image retrieval processes. Over the years, advancements in feature representation have transformed CBIR systems, evolving from simple handcrafted techniques to sophisticated automated and hybrid approaches capable of handling the diverse complexities of remote sensing imagery.

4. Recent Trends in CBIR

Content-Based Image Retrieval (CBIR) for remote sensing has witnessed transformative developments in recent years, driven by advancements in artificial intelligence and innovations in leveraging diverse data types. These trends are reshaping how LULC datasets are managed and analyzed, improving retrieval accuracy, scalability, and efficiency. This section categorizes the recent trends into two core areas: advancements in machine learning techniques and multimodal integration.

To provide a comprehensive view of recent advances, the table below summarizes selected key studies in CBIR for remote sensing applications:

Paper	Objective	Dataset	Methodology	Performance	Loop Holes
[1]	Transfer learning for LULC image retrieval	High-resolution LULC	Fine-tuned deep CNN	High precision-recall curves	High Computational costs
[2]	Fusion of morphological & LBP features	UC Merced LULC	Combined low-level features	85% precision	Limited scalability

[3]	Active learning for annotation optimization	UC Merced LULC	Deep metric learning	90% accuracy with reduced costs	Dependent on labeled datasets
[4]	Scene retrieval using co-occurrence matrices	Land-cover maps	Label Co-occurrence Matrix (LCM)	92% precision	Requires complete datasets
[5]	Morphological bag-of-words for geospatial CBIR	UC Merced LULC	Texture-based bag-of-words	87% mAP	Texture-biased, spectral info ignored
[6]	Neural networks for LULC datasets	Heterogeneous remote	CNN architecture for multispectral	89% precision	Computationally expensive
[7]	Local invariant feature extraction for geospatial	UC Merced LULC	SIFT-based feature extraction	86% precision	Lacks semantic feature analysis
[8]	Improved texture and color descriptors for CBIR	UC Merced LULC	Multi-feature fusion	88% accuracy	Struggles with dataset scalability

4.1 Advancements in Machine Learning Techniques

Machine learning, particularly deep learning, has become the backbone of modern CBIR systems for remote sensing imagery. Convolutional Neural Networks (CNNs) are widely utilized for automated feature extraction, capturing hierarchical representations of images that range from simple patterns to complex semantic information. Transfer learning further enhances these models, where pre-trained architectures such as ResNet and VGGNet are adapted for LULC datasets, significantly reducing training time and resources. For example, Alzu'bi et al. (2019) demonstrated the effectiveness of fine-tuning CNNs for aerial imagery, achieving superior precision-recall curves compared to traditional handcrafted methods.[9][11]

Deep metric learning has also emerged as a prominent trend, focusing on learning embedding spaces where similar images are close together and dissimilar ones are far apart. Siamese networks and triplet-loss mechanisms are frequently applied in this context. These methods improve similarity matching, especially in challenging datasets with high inter-class similarity, such as urban versus suburban areas.

A growing area of interest is **self-supervised learning**, which reduces reliance on annotated datasets. Models using techniques like contrastive learning exploit intrinsic properties of images—such as spatial coherence or temporal relationships—for feature extraction. These approaches address one of the most pressing limitations in CBIR, i.e., the scarcity of labeled data for large-scale remote sensing imagery.

Additionally, hybrid methods combining deep learning with traditional feature descriptors (e.g., Local Binary Patterns or GLCM) have gained traction.[8][12] By integrating complementary strengths, these approaches achieve a balance between interpretability and accuracy. These advancements in machine learning are enabling CBIR systems to adapt better to complex and large-scale LULC datasets, but challenges such as computational efficiency and generalization across datasets remain critical areas for further exploration.

4.2 Multimodal Data Integration

A significant leap in CBIR for remote sensing is the incorporation of multimodal data, which combines diverse types of information to enhance retrieval performance. Remote sensing data often spans multiple modalities, including spectral bands (multispectral and hyperspectral), spatial features, temporal dimensions, and even metadata. One key area of multimodal integration involves **spectral-spatial fusion**, where spectral features captured across multiple wavelengths are combined with spatial characteristics derived from CNN-based analysis. This fusion approach allows CBIR systems to distinguish between LULC classes with subtle differences, such as varying types of vegetation or water bodies.[10]

Metadata integration is another valuable trend. Incorporating auxiliary information such as acquisition date, geographic location, or elevation data enhances retrieval performance, particularly in dynamic scenarios where temporal or spatial context is essential. For example, multimodal CBIR systems can detect and retrieve images depicting flood-affected regions based on a combination of visual and temporal patterns.Moreover, the temporal dimension has added a new layer of intelligence to CBIR systems. Incorporating time-series data enables systems to capture and analyze changes in land use, such as urban expansion or deforestation. This capability is critical for monitoring dynamic environments and planning interventions.

While multimodal systems promise greater adaptability and accuracy, their development introduces challenges. Data preprocessing and alignment are often computationally expensive, and designing architectures that effectively merge these modalities remains a complex task. However, as multimodal integration matures, it offers exciting opportunities for robust and context-aware CBIR systems tailored to geospatial applications. These studies illustrate the interplay of novel feature engineering techniques, dataset diversity, and algorithmic efficiency. While many approaches achieve high retrieval accuracy, scalability and computational costs remain critical areas for improvement.

Challenges and Opportunities

5.1 Challenges

- **Data Diversity**: Remote sensing datasets are heterogeneous, combining multi-resolution, multi-temporal, and multi-spectral data, which complicates feature consistency.
- **Computational Costs**: Deep learning models demand significant computational resources, which may limit practical applicability, particularly for real-time retrieval tasks.
- **Limited Annotations**: Label scarcity in LULC datasets hinders supervised learning approaches, necessitating innovative semi-supervised or self-supervised methods.
- **Inter-class Similarity**: Fine-grained distinctions between classes (e.g., cropland vs. pasture) challenge conventional CBIR systems.

5.2 Opportunities

- **Explainable AI (XAI)**: Enhancing CBIR systems with explainability will boost user confidence by providing insights into why specific images were retrieved.
- **Hybrid Techniques**: Combining traditional descriptors with deep learning features can balance efficiency and accuracy.
- **Big Data Integration**: Incorporating cloud computing and distributed systems to handle large-scale geospatial datasets effectively.
- **Automated Annotation**: Leveraging self-supervised or weakly-supervised techniques to generate large-scale annotated data cost-effectively.

II. Conclusion and Future Scope

This survey underscores the significant advancements in CBIR systems for remote sensing imagery, with a focus on LULC applications. The evolution from handcrafted features to deep learning models has remarkably improved the retrieval accuracy and scalability of these systems. However, critical gaps remain in handling dataset diversity, annotation bottlenecks, and computational efficiency.

Future research should focus on developing hybrid feature models, integrating explainable AI to enhance system transparency, and employing big data analytics for scalable solutions. The exploration of multimodal CBIR—incorporating spatial, spectral, and temporal data—could revolutionize image retrieval in remote sensing. Additionally, the adoption of novel self-supervised learning approaches could mitigate the challenges posed by limited labeled data, paving the way for a new generation of CBIR systems that are intelligent, efficient, and adaptable to complex LULC tasks.

REFERENCES

[1]. Alzu'bi, A., Amira, A., & Ramzan, N. (2019). Learning transfer using deep convolutional features for remote sensing image retrieval. *IAENG International Journal of Computer Science*, *46*(4), 637-644.

[2]. Jain, Swati, and Shailee Patel. "Satellite Image Retrieval by Fusion of Morphological and LBP Features."

[3]. Hoxha, Genc, Gencer Sumbul, Julia Henkel, Lars Möllenbrok, and Begüm Demir. "Annotation Cost-Efficient Active Learning for Deep Metric Learning Driven Remote Sensing Image Retrieval." *IEEE Transactions on Geoscience and Remote Sensing* (2024).

[4]. Liu, Jun, Bin Luo, Qianqing Qin, and Guopeng Yang. "Alike scene retrieval from land-cover products based on the label co-occurrence matrix (LCM)." *Remote Sensing* 9, no. 9 (2017): 912.

[5]. Aptoula, Erchan. "Bag of morphological words for content-based geographical retrieval." In *2014 12th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pp. 1-5. IEEE, 2014.

[6]. Alshehri, Mohammed. "A content-based image retrieval method using neural network-based prediction technique." *Arabian journal for science and engineering* 45, no. 4 (2020): 2957-2973.

[7]. Yang, Yi, and Shawn Newsam. "Geographic image retrieval using local invariant features." *IEEE transactions on geoscience and remote sensing* 51, no. 2 (2012): 818-832.

[8]. Shao, Zhenfeng, Weixun Zhou, Lei Zhang, and Jihu Hou. "Improved color texture descriptors for remote sensing image retrieval." *Journal of applied remote sensing* 8, no. 1 (2014): 083584-083584.

[9]. Jain, Swati, Tanish Zaveri, Kinjal Prajapati, and Shailee Patel. "Deep learning feature map for content based image retrieval system for remote sensing application." *International Journal of Image Mining* 2, no. 1 (2016): 1-11.

[10]. Zhou, W., Guan, H., Li, Z., Shao, Z., & Delavar, M. R. (2023). Remote sensing image retrieval in the past decade: Achievements, challenges, and future directions. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16, 1447-1473.

[11]. Pandey, P. C., Koutsias, N., Petropoulos, G. P., Srivastava, P. K., & Ben Dor, E. (2021). Land use/land cover in view of earth observation: Data sources, input dimensions, and classifiers—A review of the state of the art. Geocarto International, 36(9), 957-988.

[12]. Papoutsis, I., Bountos, N.I., Zavras, A., Michail, D. and Tryfonopoulos, C., 2023. Benchmarking and scaling of deep learning models for land cover image classification. ISPRS Journal of Photogrammetry and Remote Sensing, 195, pp.250-268