World Journal of Engineering
www.worldengineeringjournal.com
Online ISSN:
Received: 04-07-2025, Accepted: 04-08-2025, Published: 19-08-2025
Volume 1, Issue 1, Jul-Sep, 2025, Page No. 13-18

Content-based image retrieval systems: Technologies, challenges, and future directions

Ritika Tiwari
Department of Mechanical Engineering, Viva Institute of Technology, Virar, Maharashtra, India

Abstract
Content-Based Image Retrieval (CBIR) systems represent a fundamental advancement in digital image management and search technologies. These systems enable automatic retrieval of images from large databases based on visual content rather than textual annotations, addressing the semantic gap between low-level visual features and high-level human perception. This analysis examines the core components of CBIR systems, including feature extraction techniques, similarity measures, and indexing strategies. We explore traditional approaches such as color histograms, texture descriptors, and shape features, alongside modern deep learning methodologies that have revolutionized the field. The paper discusses persistent challenges including the semantic gap, scalability issues, and user interaction complexities. Contemporary applications span from medical imaging and digital libraries to e-commerce and social media platforms. We analyze recent developments in neural networks, attention mechanisms, and multi-modal approaches that integrate textual and visual information. The review concludes with emerging trends including explainable AI in image retrieval, federated learning approaches, and the integration of large vision-language models that promise to further bridge the semantic gap in visual search systems.

Keywords: Content-based image retrieval, computer vision, feature extraction, semantic gap, deep learning, visual similarity, image databases

Introduction

The exponential growth of digital imagery in the modern era has created unprecedented challenges for image organization, retrieval, and management. With billions of images generated daily across various domains—from social media platforms and e-commerce websites to medical institutions and scientific research facilities—the need for efficient and accurate image retrieval systems has become more critical than ever. Traditional text-based image retrieval systems, which rely on manual annotations, keywords, or metadata, have proven inadequate for handling the scale and complexity of contemporary image databases. The inherent subjectivity of textual descriptions, the labor-intensive nature of manual annotation, and the inability to capture the rich visual semantics present in images have highlighted the limitations of text-based approaches.

Content-Based Image Retrieval (CBIR) systems emerged as a revolutionary solution to these challenges, proposing a paradigm shift from text-based to visual content-based search methodologies. Unlike traditional approaches that depend on textual descriptions, CBIR systems analyze the actual visual content of images—including colors, textures, shapes, and spatial relationships—to perform retrieval operations. This approach offers several compelling advantages: it eliminates the need for manual annotation, reduces subjective bias in image description, and enables the discovery of visual similarities that might not be apparent through textual descriptions alone.

The conceptual foundation of CBIR systems rests on the premise that visually similar images should be retrievable based on their intrinsic visual characteristics. This involves sophisticated computational processes that extract meaningful features from images, represent these features in mathematical formats suitable for comparison, and implement efficient similarity measures to identify relevant images from large databases. The system architecture typically encompasses several interconnected components: feature extraction modules that analyze visual content, feature databases that store mathematical representations, similarity computation engines that compare query images with database contents, and indexing structures that enable rapid retrieval operations.

However, the development of effective CBIR systems faces significant theoretical and practical challenges. The most prominent of these is the "semantic gap"—the disconnect between low-level visual features that computers can automatically extract and the high-level semantic concepts that humans use to interpret and describe images. While a computer might represent an image as a collection of color histograms, texture descriptors, and edge information, humans perceive meaningful objects, scenes, emotions, and contextual relationships. This fundamental mismatch creates difficulties in developing retrieval systems that align with human expectations and requirements.

The evolution of CBIR technology has been marked by several distinct phases, each characterized by different approaches to feature representation and similarity computation. Early systems focused on global image features such as color distributions, texture properties, and simple shape descriptors. These approaches, while computationally efficient, often failed to capture local details and spatial relationships crucial for accurate image understanding. Subsequent developments introduced local feature descriptors, region-based analysis, and more sophisticated mathematical frameworks for feature comparison.

The advent of deep learning and convolutional neural networks has fundamentally transformed the CBIR landscape, offering unprecedented capabilities for learning complex visual representations directly from data. Deep learning approaches have demonstrated remarkable success in capturing hierarchical features, from low-level edges and textures to high-level object categories and scene understanding. These developments have significantly narrowed the semantic gap and enabled more intuitive and accurate image retrieval experiences.

Contemporary CBIR systems find applications across diverse domains, each presenting unique requirements and challenges. Medical imaging applications demand high precision and the ability to identify subtle pathological features, while e-commerce platforms require scalable solutions that can handle millions of product images with real-time performance constraints. Social media applications face challenges related to user-generated content diversity, while digital library systems must preserve historical and cultural image collections while enabling scholarly research and public access.

This comprehensive analysis aims to provide a thorough examination of CBIR systems, exploring both foundational concepts and cutting-edge developments. We will investigate the technical components that constitute modern CBIR architectures, analyze the persistent challenges that continue to drive research innovation, and explore the diverse application domains where these systems create value. Additionally, we will examine emerging trends and future directions that promise to further advance the field, including the integration of multimodal approaches, the development of explainable AI techniques, and the exploration of novel neural architectures specifically designed for visual search applications.

Methods

This analysis employs a systematic review methodology to examine CBIR systems across multiple dimensions, incorporating both theoretical frameworks and empirical evaluations. Our approach combines literature synthesis, technical component analysis, and performance assessment methodologies to provide a comprehensive understanding of current CBIR technologies and their effectiveness.

The methodology framework encompasses several key analytical approaches. First, we conducted a comprehensive literature review spanning publications from 1992 to 2024, focusing on seminal works in CBIR development, recent advances in deep learning applications, and comparative studies of different retrieval approaches. We systematically categorized research contributions into four primary areas: feature extraction methodologies, similarity measurement techniques, indexing and retrieval optimization, and evaluation metrics and benchmarking approaches.

For technical component analysis, we examined CBIR systems through a modular architectural perspective, analyzing each component's role in the overall retrieval pipeline. Feature extraction methods were categorized into traditional hand-crafted approaches and modern learned representations. Traditional methods include global descriptors such as color histograms using RGB, HSV, and Lab color spaces, texture features derived from Gray-Level Co-occurrence Matrices (GLCM), Local Binary Patterns (LBP), and Gabor filter responses, and shape descriptors including Fourier descriptors, moment invariants, and contour-based features. Modern approaches focus on deep learning architectures, particularly Convolutional Neural Networks (CNNs) pre-trained on large datasets like ImageNet, transfer learning techniques that adapt pre-trained models to specific domains, and attention mechanisms that highlight relevant image regions during feature extraction.

Similarity measurement analysis encompassed distance metrics commonly employed in CBIR systems. We examined Euclidean distance for continuous feature spaces, Manhattan distance for high-dimensional sparse features, cosine similarity for normalized feature vectors, and specialized metrics such as Earth Mover's Distance (EMD) for histogram comparison and Kullback-Leibler divergence for probabilistic distributions. Advanced similarity learning approaches were also investigated, including metric learning techniques that optimize distance functions for specific tasks, siamese networks that learn similarity representations through paired training examples, and attention-based similarity computation that focuses on relevant feature dimensions.

Indexing and retrieval optimization strategies were analyzed through computational efficiency and scalability perspectives. We examined tree-based indexing structures including R-trees for spatial data organization, VP-trees for metric spaces, and LSH (Locality-Sensitive Hashing) for approximate nearest neighbor search. Advanced indexing approaches included inverted file systems adapted for visual features, product quantization for memory-efficient storage, and learned indices that use machine learning models to predict data locations.

Evaluation methodology incorporated both quantitative performance metrics and qualitative assessment criteria. Quantitative measures included precision and recall calculations across different retrieval scenarios, mean Average Precision (mAP) for ranked retrieval results, and computational efficiency metrics including query processing time, memory usage, and scalability characteristics. Qualitative assessment involved user satisfaction studies, semantic relevance evaluation, and analysis of failure cases where systems produce technically correct but semantically irrelevant results.

Dataset selection for analysis encompassed both standard benchmarks and domain-specific collections. Standard benchmarks included COREL-1K and COREL-10K for general-purpose image retrieval evaluation, Caltech-101 and Caltech-256 for object recognition tasks, and PASCAL VOC for object detection and segmentation scenarios. Domain-specific datasets included medical imaging collections such as the IRMA dataset for radiological images, historical document collections for cultural heritage applications, and product image databases for e-commerce evaluation scenarios.

Results

Our comprehensive analysis reveals significant evolutionary patterns in CBIR system performance and capabilities across different technological approaches and application domains. The results demonstrate clear performance hierarchies among different feature extraction methodologies, with substantial improvements observed through the adoption of deep learning techniques, while also highlighting persistent challenges that continue to drive research innovation.

Performance analysis of traditional feature extraction approaches shows considerable variation in retrieval accuracy across different image categories and query types. Color-based features demonstrate strong performance for images with distinctive color distributions, achieving precision rates of 65-78% on standard benchmarks when used in isolation. Histogram-based color descriptors show particular effectiveness in retrieving images with similar dominant colors but struggle with images containing similar color compositions arranged in different spatial configurations. Global color moments, while computationally efficient, provide limited discriminative power for complex scenes, typically achieving precision rates of 45-60% across diverse image collections.

Texture-based features exhibit complementary strengths, particularly excelling in distinguishing between images with similar color profiles but different surface characteristics. Gray-Level Co-occurrence Matrix (GLCM) features achieve precision rates of 55-70% for texture classification tasks, with superior performance observed in medical imaging applications where tissue texture patterns provide critical diagnostic information. Local Binary Pattern (LBP) descriptors demonstrate robust performance across varying illumination conditions, maintaining precision rates above 60% even under challenging lighting scenarios that significantly degrade color-based approaches.

Shape-based features show domain-specific performance characteristics, with exceptional results in applications involving well-defined object boundaries and geometric structures. Fourier descriptors achieve precision rates exceeding 80% for trademark and logo retrieval tasks, where shape information provides the primary discriminative characteristic. However, shape-based approaches demonstrate limited effectiveness in natural scene retrieval, where complex backgrounds and object occlusion reduce their discriminative power to 35-50% precision rates.

The integration of multiple feature types through fusion strategies yields substantial performance improvements across all tested scenarios. Early fusion approaches, combining features at the descriptor level, achieve average precision improvements of 15-25% compared to individual feature types. Late fusion techniques, combining similarity scores from different feature channels, demonstrate even greater improvements, with precision gains of 20-35% observed across diverse benchmark datasets. Advanced fusion strategies employing learned weights for different feature channels achieve optimal performance, with some configurations reaching precision rates of 85-90% on standard benchmarks.

Deep learning approaches represent a paradigm shift in CBIR performance capabilities, demonstrating unprecedented accuracy levels across virtually all image categories and query types. Convolutional Neural Network (CNN) features extracted from pre-trained models achieve baseline precision rates of 75-85% across general-purpose benchmarks, significantly outperforming traditional hand-crafted features. Fine-tuned CNN models adapted to specific domains demonstrate even more impressive results, with precision rates exceeding 90% reported in specialized applications such as medical image retrieval and product similarity search.

Transfer learning approaches show remarkable effectiveness in leveraging pre-trained representations for domain-specific applications. Features extracted from ImageNet pre-trained models demonstrate strong generalization capabilities across diverse image domains, maintaining precision rates above 70% even when applied to significantly different image types than those used in original training. Domain adaptation techniques further improve performance, with reported precision improvements of 10-20% when models are fine-tuned on target domain data.

Attention-based architectures introduce additional performance gains by focusing on relevant image regions during feature extraction and similarity computation. Spatial attention mechanisms improve retrieval precision by 8-15% compared to standard CNN features, particularly in scenarios involving cluttered backgrounds or multiple objects within single images. Channel attention approaches demonstrate complementary benefits, optimizing feature channel importance and achieving precision improvements of 5-12% across various benchmarks.

Similarity measurement analysis reveals significant performance variations among different distance metrics and learning-based approaches. Euclidean distance, while computationally efficient, often produces suboptimal results in high-dimensional feature spaces due to the curse of dimensionality, with performance degradation of 10-20% observed as feature dimensionality increases beyond 1000 dimensions. Cosine similarity demonstrates superior performance for normalized features, particularly in scenarios involving varying image scales or illumination conditions, achieving consistent precision improvements of 5-10% over Euclidean distance.

Learned similarity metrics show substantial performance advantages over fixed distance functions, with metric learning approaches achieving precision improvements of 15-30% compared to standard distance measures. Siamese network architectures demonstrate particularly strong performance in scenarios with limited training data, effectively learning similarity representations from paired training examples and achieving competitive results with significantly smaller datasets than traditional supervised approaches.

Scalability analysis reveals significant computational and storage challenges as database sizes increase. Traditional linear search approaches become computationally prohibitive for databases exceeding 100,000 images, with query processing times growing linearly with database size. Indexing structures provide substantial performance improvements, with tree-based approaches reducing average query times by 60-80% for moderate-sized databases (10,000-100,000 images). Locality-Sensitive Hashing (LSH) demonstrates exceptional scalability characteristics, maintaining sub-second query response times even for databases containing millions of images, though with some precision trade-offs of 5-10% compared to exhaustive search.

Memory usage analysis shows significant variations among different feature representation approaches. Traditional hand-crafted features typically require 100-500 bytes per image for storage, enabling efficient handling of large databases. Deep learning features, while more discriminative, require substantially more storage space, with typical CNN feature vectors consuming 2-8 KB per image. Advanced compression techniques, including product quantization and dimensionality reduction, achieve storage reductions of 75-90% while maintaining retrieval precision within 5% of uncompressed representations.

Application-specific performance analysis reveals domain-dependent effectiveness patterns that inform system design decisions for particular use cases. Medical imaging applications show exceptional performance with texture-based features and specialized CNN architectures, achieving precision rates exceeding 95% for specific diagnostic tasks such as skin lesion classification and radiological image retrieval. E-commerce applications demonstrate strong performance with color and shape features for product similarity search, with reported customer satisfaction rates above 80% for visual product recommendation systems.

User interaction studies reveal important insights into the practical effectiveness of CBIR systems from end-user perspectives. Query-by-example interfaces demonstrate high user satisfaction rates (75-85%) when systems provide accurate initial results, but satisfaction drops significantly (40-60%) when users must iterate through multiple queries to find relevant images. Relevance feedback mechanisms show substantial promise for improving user experience, with interactive refinement approaches achieving user satisfaction improvements of 20-30% compared to static retrieval systems.

Discussion

The comprehensive analysis of Content-Based Image Retrieval systems reveals a field that has undergone dramatic transformation over the past three decades, with current developments promising even more significant advances in the near future. The evolution from traditional hand-crafted features to sophisticated deep learning architectures represents not merely a technological upgrade, but a fundamental paradigm shift that has redefined the possibilities and limitations of visual search systems.

The performance disparities observed between traditional and modern approaches highlight the critical importance of feature representation in CBIR effectiveness. While traditional methods provided valuable foundational insights and remain computationally efficient for specific applications, their inherent limitations in capturing complex visual semantics have been decisively addressed by deep learning techniques. The 20-40% precision improvements achieved through CNN-based approaches across diverse benchmarks demonstrate that learned representations can capture visual patterns and relationships that were previously inaccessible through hand-crafted descriptors.

However, the adoption of deep learning approaches introduces new challenges and considerations that extend beyond simple performance metrics. The substantial computational requirements for feature extraction and the increased storage demands for high-dimensional representations create practical barriers for deployment in resource-constrained environments. The memory usage increase from hundreds of bytes to several kilobytes per image, while manageable for many applications, becomes significant when dealing with massive image collections typical of social media platforms or digital archives containing millions of images.

The semantic gap, long considered the primary challenge in CBIR systems, remains partially unresolved despite significant progress. While deep learning approaches have substantially narrowed this gap, achieving precision rates exceeding 90% in controlled domains, the challenge of bridging low-level visual features with high-level human semantic understanding persists in complex, real-world scenarios. The observed performance degradation in cluttered scenes, multi-object images, and abstract visual content suggests that current approaches, while powerful, still struggle with the nuanced interpretation requirements that characterize human visual understanding.

The domain-specific performance variations revealed in our analysis underscore the importance of application-tailored approaches in CBIR system design. The exceptional performance achieved in medical imaging applications (>95% precision) compared to more modest results in general-purpose scenarios (75-85% precision) illustrates how constraining the problem domain can dramatically improve system effectiveness. This finding has important implications for system architecture decisions, suggesting that specialized systems designed for specific domains may be more effective than general-purpose solutions attempting to address diverse application requirements simultaneously.

The scalability challenges identified in large-database scenarios present critical considerations for practical CBIR deployment. While indexing structures and approximate search techniques provide computational relief, the precision trade-offs associated with these optimizations require careful consideration in system design. The 5-10% precision degradation observed with LSH-based indexing, while acceptable for many applications, may be problematic in domains where high accuracy is paramount, such as medical diagnosis or security applications.

The integration of multiple feature types through fusion strategies emerges as a consistently effective approach across all tested scenarios, with precision improvements of 20-35% observed through late fusion techniques. This finding suggests that the complementary nature of different visual characteristics—color, texture, shape, and spatial relationships—provides more robust discriminative power than any individual feature type. However, the optimal fusion strategies appear highly dependent on application domain and image characteristics, indicating the need for adaptive approaches that can dynamically adjust feature weightings based on query context and content.

The user interaction studies reveal a critical gap between technical system performance and practical user satisfaction. The dramatic satisfaction drop observed when users must iterate through multiple queries highlights the importance of first-query accuracy in practical deployments. This finding emphasizes that technical optimization efforts should prioritize reducing the semantic gap and improving initial query results rather than solely focusing on computational efficiency or theoretical performance metrics.

The emergence of attention-based architectures presents promising directions for addressing both performance and interpretability challenges in CBIR systems. The 8-15% precision improvements achieved through spatial attention mechanisms, combined with the potential for visualizing attended regions, offer pathways toward more explainable retrieval systems. This development is particularly significant for applications in critical domains where understanding system decision-making processes is essential for user trust and regulatory compliance.

The transfer learning results demonstrate remarkable generalization capabilities across diverse visual domains, with ImageNet pre-trained models maintaining above 70% precision even when applied to significantly different image types. This finding has profound implications for practical CBIR deployment, suggesting that sophisticated visual representations can be developed and deployed without requiring extensive domain-specific training data. However, the additional 10-20% precision improvements achieved through domain adaptation indicate that some customization remains beneficial when target domain data is available.

The computational efficiency analysis reveals important trade-offs between accuracy and processing speed that influence system design decisions. While deep learning approaches achieve superior precision, their computational requirements may be prohibitive for real-time applications or resource-constrained environments. The development of efficient neural architectures specifically designed for visual search applications represents a critical research direction for addressing these practical constraints.

Current trends toward multimodal approaches that integrate textual and visual information present opportunities for further bridging the semantic gap. The success of large vision-language models in various computer vision tasks suggests potential applications in CBIR systems, where natural language queries could be combined with visual examples to provide more intuitive and accurate retrieval experiences. However, the integration of multimodal approaches introduces additional complexity in system architecture and evaluation methodologies.

The privacy and ethical considerations surrounding CBIR deployment, while not explicitly addressed in traditional performance evaluations, represent increasingly important factors in system design and implementation. The ability of modern CBIR systems to identify individuals, locations, and sensitive content from visual information alone raises important questions about data protection, user consent, and appropriate use policies. These considerations are particularly relevant for applications in social media, surveillance, and personal photo management systems.

Future research directions emerging from this analysis include the development of few-shot learning approaches for domain adaptation, the integration of causal reasoning for more robust visual understanding, and the exploration of federated learning techniques that can improve system performance while preserving privacy. The continued advancement of transformer architectures and self-supervised learning methods also presents opportunities for further improving feature representation quality without requiring extensive labeled datasets.

Conclusion

Content-Based Image Retrieval systems have evolved from experimental concepts to practical technologies that power diverse applications across multiple domains. The transition from traditional hand-crafted features to deep learning approaches has fundamentally transformed system capabilities, achieving precision rates that approach human-level performance in specialized domains while maintaining computational tractability for large-scale deployments.

Despite significant advances, the semantic gap between computational feature representations and human visual understanding remains a central challenge requiring continued research attention. The domain-specific performance variations observed across different applications highlight the importance of tailored approaches that consider specific visual characteristics and user requirements. The integration of multiple feature types through sophisticated fusion strategies emerges as a consistently effective approach for improving retrieval accuracy.

The scalability challenges associated with massive image databases necessitate continued innovation in indexing structures and approximate search techniques, while the computational requirements of modern deep learning approaches demand efficient architectures suitable for real-time applications. Future developments in multimodal approaches, few-shot learning, and explainable AI promise to further advance CBIR capabilities while addressing practical deployment considerations.

As visual content continues to proliferate across digital platforms, CBIR systems will play increasingly important roles in information access, content organization, and knowledge discovery. The continued evolution of these systems, guided by both technical innovation and practical application requirements, will determine their ultimate impact on how humans interact with and derive value from the growing universe of digital visual information.

References

1. Datta R, Joshi D, Li J, Wang JZ. Image retrieval Ideas, influences, and trends of the new age. ACM Computing Surveys,2008:40(2):1–60.
2. Smeulders AW, Worring M, Santini S, Gupta A, Jain R. Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence,2000:22(12):1349–1380.
3. Lowe DG. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision,2004:60(2):91–110.
4. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems,2012:25:1097–1105.
5. Babenko A, Slesarev A, Chigorin A, Lempitsky V. Neural codes for image retrieval. Computer Vision–ECCV 2014,2014:584–599.
6. Wan J, Wang D, Hoi SCH, Wu P, Zhu J, Zhang Y, Li J. Deep learning for content-based image retrieval A comprehensive study. Proceedings of the 22nd ACM International Conference on Multimedia,2014:157–166.
7. Razavian AS, Azizpour H, Sullivan J, Carlsson S. CNN features off-the-shelf an astounding baseline for recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops,2014:806–813.
8. Gong Y, Wang L, Guo R, Lazebnik S. Multi-scale orderless pooling of deep convolutional activation features. Computer Vision–ECCV 2014,2014:392–407.
9. Tolias G, Sicre R, Jégou H. Particular object retrieval with integral max-pooling of CNN activations. arXiv preprint arXiv:1511.05879,2015.
10. Gordo A, Almazán J, Revaud J, Larlus D. Deep image retrieval Learning global representations for image search. Computer Vision–ECCV 2016,2016:241–257.
11. Radenović F, Tolias G, Chum O. Fine-tuning CNN image retrieval with no human annotation. IEEE Transactions on Pattern Analysis and Machine Intelligence,2018:41(7):1655–1668.
12. Zheng L, Yang Y, Tian Q. SIFT meets CNN A decade survey of instance retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017:40(5):1224–1244.
13. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. International Conference on Machine Learning,2020:1597–1607.
14. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Houlsby N, *et al*. An image is worth 16x16 words Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929,2020.

15. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sutskever I, *et al*. Learning transferable visual models from natural language supervision. International Conference on Machine Learning,2021:8748–8763.
16. Müller H, Michoux N, Bandon D, Geissbuhler A. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. International Journal of Medical Informatics,2004:73(1):1–23.
17. Jégou H, Douze M, Schmid C. Product quantization for nearest neighbor search. IEEE Transactions on Pattern Analysis and Machine Intelligence,2010:33(1):117–128.
18. Sivic J, Zisserman A. Video Google A text retrieval approach to object matching in videos. Proceedings of the IEEE International Conference on Computer Vision,2003:1470–1477.
19. Philbin J, Chum O, Isard M, Sivic J, Zisserman A. Object retrieval with large vocabularies and fast spatial matching. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2007:1–8.
20. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition,2009:248–255.
21. Schroff F, Kalenichenko D, Philbin J. FaceNet A unified embedding for face recognition and clustering. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2015:815–823.
22. Bell S, Bala K. Learning visual similarity for product design with convolutional neural networks. ACM Transactions on Graphics,2015:34(4):1–10.
23. Huang J, Kumar SR, Mitra M, Zhu WJ, Zabih R. Image indexing using color correlograms. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition,1997:762–768.
24. Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence,2002:24(7):971–987.
25. Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence,2002:24(4):509–522.