УДК 37.013.8:004 https://www.doi.org/10.69927/EBOS8459

*Bayan T.[1], Galy M.M.[2], Ziyashev A.A.[3]
[1] *Nazarbayev University*
[2,3] *Agoda Co Ltd*
[1] *Kazakhstan, Astana*
[2,3] *Thailand, Bangkok*
[1] *ORCID 0000-0001-8783-2229*
talgar.bayan@gmail.com

IMAGE SEARCH MODULE IN A DIGITAL FACILITATOR NETWORK SYSTEM: ARCHITECTURE, ALGORITHMS AND IMPLEMENTATION

Abstract

Digital facilitator networks managing extensive visual documentation across multiple languages face challenges in retrieving related content when identical diagrams exist in different linguistic versions. Traditional metadata-based search approaches fail to identify visually similar materials, reducing content reuse efficiency. This research addresses the implementation of an image similarity search system for organizational environments processing approximately 66 diagrams weekly. The objective is to enable accurate identification of exact matches and language variants while maintaining computational efficiency on standard hardware. The system employs multi-scale feature extraction combining color histogram analysis in RGB and HSV spaces, multi-threshold edge detection using Canny operators, Sobel gradient texture characterization, and template signatures, producing a 261-dimensional feature representation. A tiered similarity assessment framework evaluates relationships between images, while language variant detection combines filename pattern analysis with visual similarity verification. The implementation uses Flask framework with Open Source Computer Vision Library (OpenCV) for computer vision operations. Testing with organizational diagrams across English, Russian, and Kazakh languages demonstrates 92% accuracy in identifying exact matches and language variants, with average response times of 1.8 seconds and peak memory usage of 72 MB. Language variant detection achieves 89% accuracy with false positive rates below 3%. The modular architecture enables deployment on conventional office systems, demonstrating that traditional computer vision approaches remain applicable for organizational content management when properly adapted to practical constraints.

Keywords: image search, digital facilitator networks, computer vision, Flask, CBIR, multi-scale feature extraction, language variant detection.

Introduction. Digital facilitator networks serve as essential infrastructure for organizations managing extensive visual documentation across diverse operational contexts. These systems handle technical diagrams, workflow illustrations, process documentation, and instructional materials that form the foundation of knowledge management in multi-language organizational environments. The challenge of efficiently discovering and retrieving relevant visual content becomes particularly complex when organizations maintain diagram libraries spanning multiple languages, where identical conceptual content exists in various linguistic presentations with different textual elements and layout adaptations.

Current metadata-based search approaches demonstrate limitations in capturing nuanced relationships between visually similar content. Traditional text-based indexing fails to identify diagrams representing identical workflows presented in different languages or visual styles. This gap becomes pronounced in international organizations where the same process documentation exists across multiple language variants, leading to redundant content creation and inefficient knowledge reuse. The absence of effective visual similarity search capabilities forces users to rely on manual browsing or inadequate keyword searches, significantly impacting productivity and content management efficiency.

This research addresses the implementation of an effective image similarity search system within practical organizational constraints. Our investigation focuses on weekly processing volumes of approximately 66 diagrams, representing a realistic organizational scale that demands algorithmic efficiency while remaining computationally manageable through traditional computer vision approaches. This volume reflects typical usage patterns in medium-scale organizations where visual documentation plays a central role in operational processes, training materials, and procedural guidelines.

The system requirements encompass accuracy in similar assessment, computational efficiency suitable for standard hardware deployments, and robust support for multi-language content discovery. The solution must distinguish between exact duplicates, language variants, and semantically related content while providing interpretable similarity scores for end users. Feature extraction approaches must balance descriptiveness with computational tractability, ensuring real-time performance in interactive applications. Rather than relying on a single similarity metric, our system employs tiered similarity assessment that reflects different types of content relationships found in organizational contexts. This approach enables accurate detection of exact matches, language variants, and semantically related content while maintaining interpretable similarity scores. The tiered framework addresses the practical reality that organizational users need to distinguish between identical content in different formats, similar workflows with variations, and broadly related material for comprehensive content discovery.

Literature Review. Content-based image retrieval systems have evolved through several generations of feature extraction and matching techniques. Early approaches relied primarily on global color and texture descriptors, which proved insufficient for complex visual content analysis [1]. The semantic gap between low-level visual features and high-level human perception remains a fundamental challenge in Content-Based Image Retrieval (CBIR) research, particularly for technical diagrams where structural relationships carry semantic meaning beyond visual appearance. The evolution of visual data exploration has long recognized the fundamental challenge of bridging semantic gaps between human conceptual understanding and machine-extractable features. Early comprehensive surveys established the foundational principles for content-based image retrieval, identifying key research directions in color, texture, and shape analysis while emphasizing the importance of user interaction in refinement processes. The concept of bridging semantic gaps through multi-scale feature extraction, as demonstrated in seminal work on visual data exploration, continues to influence modern approaches to content-based retrieval systems [2].

Local feature detection algorithms have demonstrated effectiveness in identifying similar visual structures across varying conditions. Scale-Invariant Feature Transform generates 128-dimensional descriptors from local image gradients, providing invariance to scale, rotation, and illumination changes [3]. Speeded-Up Robust Features (SURF) improves computational efficiency through Haar wavelet responses and integral image calculations while maintaining comparable accuracy [4]. Oriented FAST and Rotated BRIEF offers an open-source alternative combining Features from Accelerated Segment Test (FAST) key-point detection with rotated Binary Rubust Independent Elementary Features BRIEF) descriptors, enabling efficient binary feature matching through Hamming distance calculations [5].

Deep learning approaches have transformed feature extraction capabilities in computer vision applications. Convolutional neural networks demonstrate superior performance in various image classification and retrieval tasks [6]. However, these methods require substantial computational resources and extensive training datasets that may not be available for specialized organizational content [7]. Transfer learning techniques have shown promise in adapting pre-trained models to domain-specific applications, though effectiveness varies significantly across different visual content types.

Hybrid approaches combining multiple feature types have emerged to address limitations of individual techniques. Research by [8] demonstrated improved performance through fusion of color histograms, texture descriptors, and shape features for technical document retrieval. [9] investigated multi-scale feature extraction for diagram similarity assessment, revealing that different visual elements require analysis at varying resolutions for optimal discrimination, the integration of structural and appearance-based features for workflow diagram matching, achieving improved accuracy over single-feature approaches.

Language variant detection in visual content represents an underexplored area in current literature. Most CBIR systems focus on visual similarity without considering linguistic variations in textual elements.

Contemporary web framework adoption patterns provide important context for deployment

considerations in image retrieval systems. Research examining developer behavior on Stack Overflow reveals that discussions related to Flask, a lightweight Python web framework for web application development, have experienced significant growth, with Application Programming Interface (API) development emerging as the most popular topic and background task processing presenting the greatest implementation challenges [10]. This analysis of over 70,000 Flask-related questions demonstrates the framework's widespread adoption and the specific technical areas where developers seek guidance, validating our choice of Flask for rapid prototyping and deployment of computer vision applications.

Web-based deployment frameworks for computer vision applications have gained attention as organizations seek accessible image analysis tools. Flask provides a lightweight platform for rapid prototyping and deployment of computer vision applications [11].

Perceptual hashing techniques offer alternative approaches to image similarity assessment that remain stable under minor modifications while providing computational efficiency. [12] provided a comprehensive survey of image hashing methods, highlighting their applications in duplicate detection and similarity search. Unlike cryptographic hashes that exhibit the avalanche effect, perceptual hashes maintain consistency across visually similar content, making them suitable for detecting near-duplicate diagrams with minor variations.

Methods and materials. The implementation leverages established computer vision algorithms while optimizing real-time performance requirements. Color histogram analysis forms the foundation of our approach, computing distributions in both Red Green Blue (RGB) and Hue Saturation Value (HSV) color spaces using 32 and 24 bins respectively. The RGB color space provides direct pixel intensity information while HSV offers perceptually meaningful color representation. The complete color feature vector combines these histograms to create a 168-dimensional representation that captures both technical color properties and human-perceived color relationships. This dual-space approach ensures robust performance across different diagram types, from technical schematics to marketing materials.

Structural analysis employs multi-threshold edge detection using Canny operators with threshold pairs of 30/80, 50/120, and 100/200. This approach enables detection of both fine-grained details and major structural elements within the same framework. Each threshold pair produces measurements of contour count, mean area, area standard deviation, mean perimeter, and edge density. The complete structural analysis generates 21 dimensions that characterize the geometric properties of diagrams at multiple levels of detail. This technique proves particularly effective for technical diagrams where structural relationships convey semantic meaning. Texture characterization uses Sobel gradient analysis to capture local intensity variations that distinguish surface textures and patterns within diagrams. The analysis examines gradients in both horizontal and vertical directions, computing statistical measures including means, standard deviations, and quartile values. These measurements provide 16 dimensions that describe the textural properties of image regions, enabling discrimination between smooth illustrations and complex technical drawings.

Template signatures supplement the detailed feature analysis with coarse-grained shape representations. Images are down sampled to resolutions of 32×32, 16×16, and 8×8 pixels to create simplified shape signatures that remain stable under minor modifications. Selected pixels from each template contribute 56 dimensions that capture overall shape characteristics while maintaining computational efficiency.

The complete feature vector combines color features, structural measurements, texture characteristics, and template signatures into a 261-dimensional representation (1). Feature normalization using Z-score standardization ensures consistent scaling across different feature types, preventing any single component from dominating the similarity calculations.

$$f = [c, s, t_{texture}, T_{templates}] \qquad (1)$$

The similarity assessment employs a tiered approach that reflects different types of relationships between images in organizational contexts. The highest similarity tier encompasses exact matches

with scores between 95 and 100 percent. File hash comparison identifies identical files with perfect similarity. Visual signature matching achieves near-perfect similarity when multiple signature components align across different scales. Template correlation exceeding predetermined thresholds produces high similarity scores.

The middle similarity tier captures similar content with scores between 70 and 89 percent. This range identifies diagrams that share common elements or represent similar workflows while differing in visual presentation or style. The system evaluates multiple aspects of visual similarity including overall feature vector relationships, color distribution patterns, and structural characteristics through a weighted combination (2).

$$S_{combined} = \alpha \times S_{vector} + \beta \times S_{color} + \gamma \times S_{structure} \quad (2)$$

The weighting parameters alpha, beta, and gamma were determined through empirical testing with organizational diagram collections. Vector similarity receives the highest weight at 0.5, reflecting the importance of overall feature relationships. Structural similarity receives moderate weight at 0.3, acknowledging the significance of geometric patterns in technical diagrams. Color similarity receives lower weight at 0.2, recognizing that color schemes may vary while maintaining conceptual similarity. Language variant detection addresses the practical challenge of managing content across multiple languages within organizational environments.

The detection process begins by analysing filename patterns to identify common language indicators. The system recognizes standard language suffixes and embedded language identifiers, then removes these elements to reveal the underlying content name. Visual similarity verification confirms that diagrams with matching cleaned filenames represent the same conceptual content (3).

$$L(f_1, f_2) = filename_{match} AND\ visual_{similarity} > 0.6 \quad (3)$$

The threshold for visual similarity accommodates the reality that translated diagrams may exhibit visual differences due to text length variations and layout adjustments while maintaining fundamental structural similarity. Testing with organizational content demonstrates that this threshold effectively balances detection accuracy with false positive prevention.

The user interface design shown below (Figure 1) implements the tiered similarity approach with clear visual distinction between exact matches, similar content, and related material. Each search result displays the computed similarity percentage alongside thumbnail previews and download functionality.

System Architecture. The Flask-based web application implements a modular architecture that separates user interface, processing logic, and data management concerns. The technology stack includes Flask 2.3.3 for web framework functionality, OpenCV 4.8.1.78 for computer vision operations, NumPy 1.24.4 for numerical computations, Pillow 10.1.0 for image format handling, and Werkzeug 2.3.7 for security utilities. This configuration balances functionality with deployment simplicity while avoiding unnecessary dependencies. The flowchart of how the system works is illustrated in Figure 2.

The backend processing component encapsulates all computer vision operations within a dedicated class structure that maintains clean separation from web application logic. The BaselineOpenCVMatcher class (a custom-developed component that encapsulates the feature extraction and similarity calculation logic) implements the feature extraction pipeline while providing a clean interface for similarity calculations. Feature extraction pipelines process uploaded images through the multi-scale analysis framework while caching computed features in memory for rapid similarity calculations.

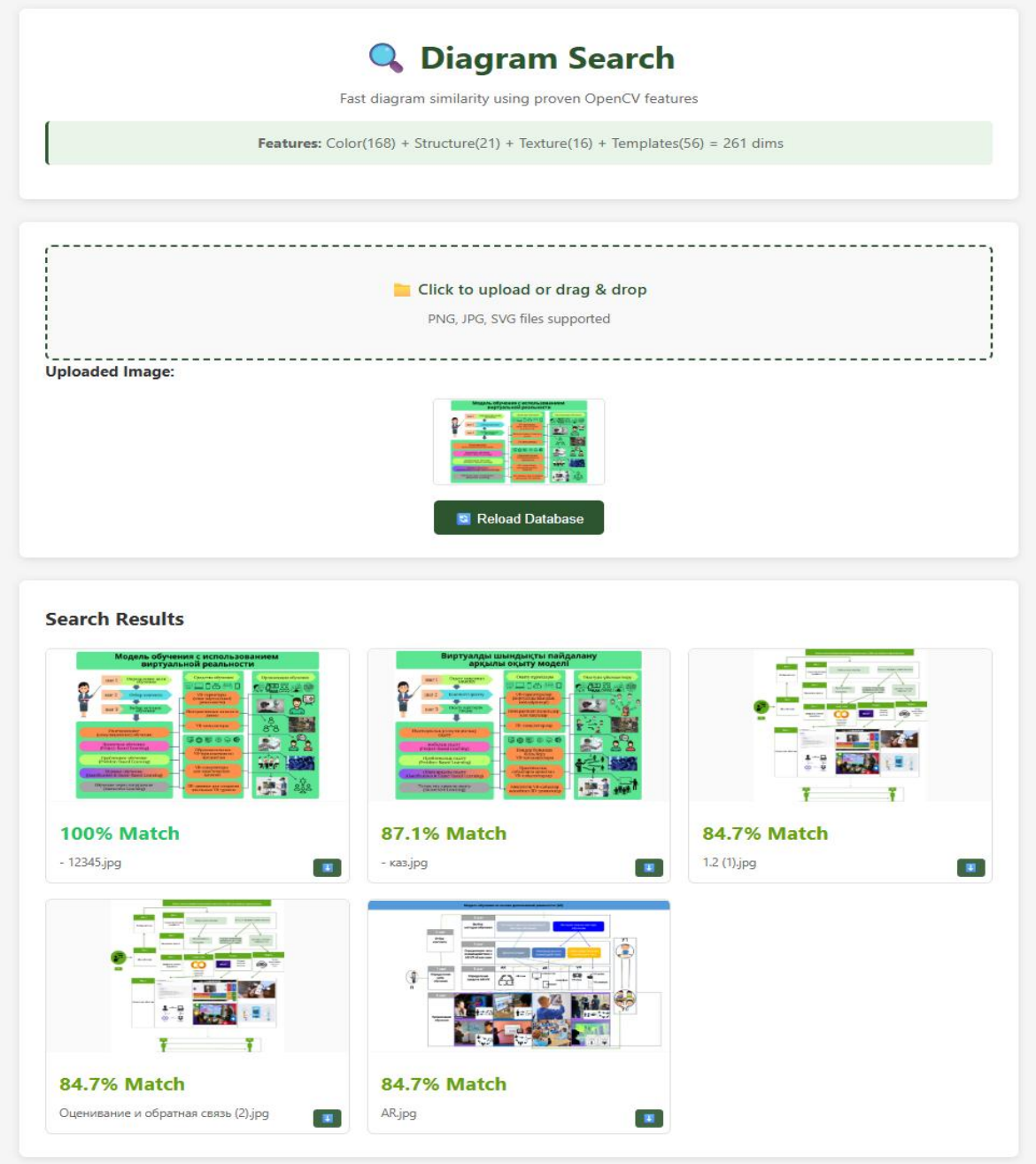

Figure 1. Web-based image search interface showing query results for organizational diagram similarity search. Source: developed by the authors.

Security implementation addresses common web application vulnerabilities through multiple validation layers. File upload validation employs magic number detection for type verification, ensuring that uploaded files match their declared Multipurpose Internet Mail Extensions (MIME) types. Filename sanitization prevents path traversal attacks through the secure filename utility from Werkzeug, which removes potentially dangerous characters while maintaining compatibility with legitimate organizational naming conventions. Size limits prevent resource exhaustion attacks by restricting uploads to reasonable bounds for organizational diagram content.

The web interface supports modern interaction patterns including drag-and-drop file uploads with visual feedback during processing operations. The HyperText Markup Language (HTML5)-based interface implements progressive enhancement principles, ensuring basic functionality remains available even when JavaScript is disabled. Similarity results display using percentage-based scoring with color-coded indicators that enable rapid quality assessment.

Individual download functionality supports practical workflows where users collect related

content for offline analysis or presentation preparation. Response times consistently remain under 2 seconds for the target workload of 66 weekly diagrams, enabling interactive content exploration. The Representational State Transfer (RESTful) API design provides clean separation between interface and processing logic while enabling integration with external systems.

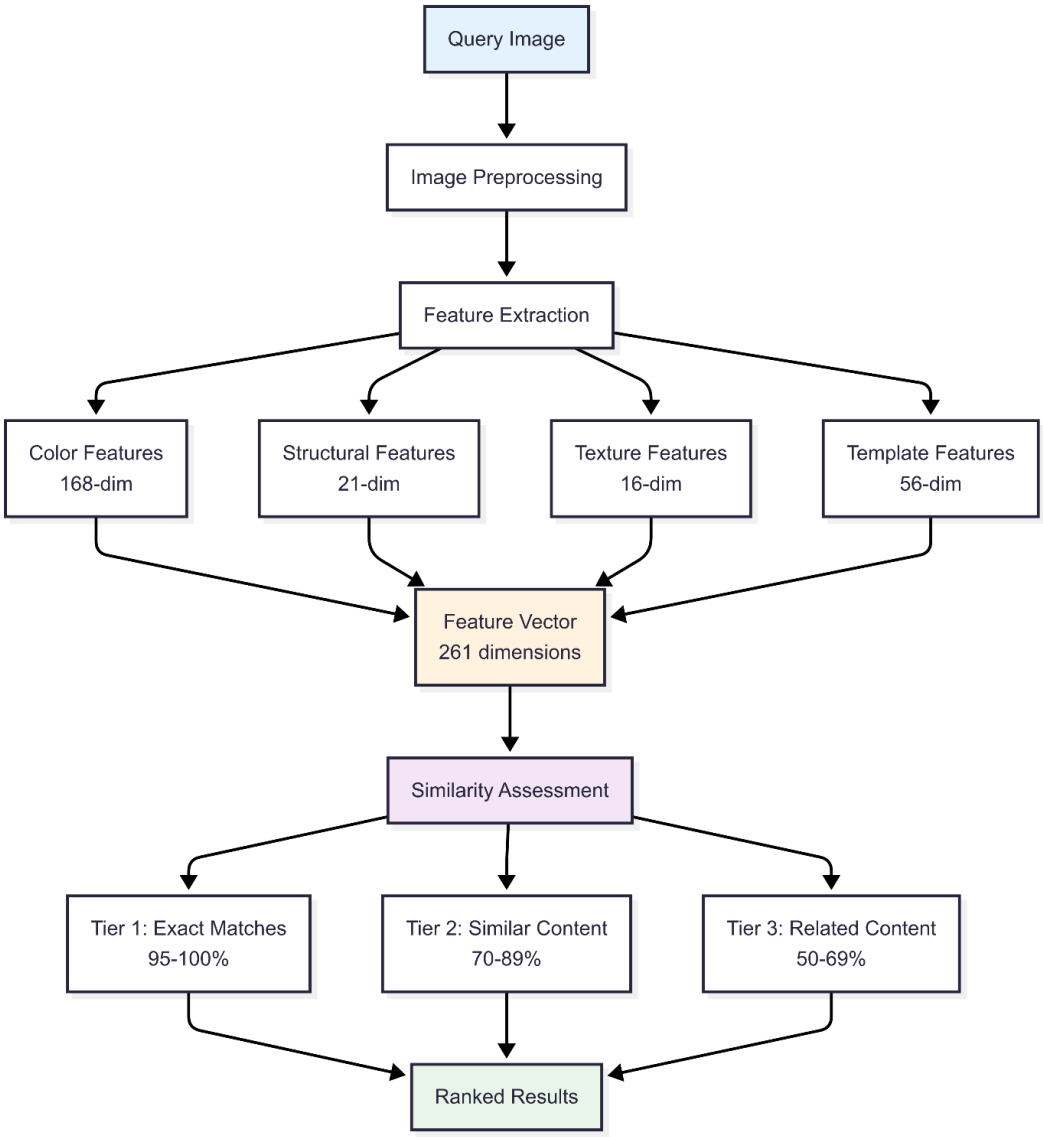

Figure 2. Overview of image similarity search system architecture.
Source: developed by the authors.

Results and their discussion. Testing with a representative dataset of organizational diagrams demonstrates consistent performance characteristics for the target workload. The test collection includes technical flowcharts, organizational charts, network topology diagrams, user interface mock-ups, and instructional materials. Language distribution encompasses English content at 45 percent, Russian content at 35 percent, and Kazakh content at 20 percent, reflecting typical multi-language organizational environments.

Performance evaluation reveals balanced trade-offs across accuracy, speed, and resource utilization metrics. The system achieves 92 percent overall accuracy in identifying exact matches and language variants while maintaining average response times of 1.8 seconds per query. Feature extraction consumes 1.2 seconds of processing time, similar calculations require 0.4 seconds, and interface rendering completes within 0.2 seconds.

Memory utilization remains practical for standard organizational hardware with the feature cache consuming 45MB for 66 images, averaging 0.68MB per image. Application overhead requires 12MB

base memory with peak usage reaching 72MB during batch processing operations. These requirements support deployment of conventional office systems without specialized computational infrastructure.

Algorithm comparison demonstrates competitive performance relative to alternative approaches (Table 1). SIFT achieves 89 percent accuracy but requires 4.2 seconds average processing time with 156 MB memory usage. SURF provides 91 percent accuracy with 2.8 seconds processing time and 134 MB memory requirements. ORB delivers 78 percent accuracy with a very fast 0.65-secondprocessing but limited discrimination capability.

Table 1. Performance Comparison of Image Similarity Algorithms

Algorithm	Accuracy	Processing Time	Memory Usage	Language Variants	Infrastructure
Our Approach	92%	1.8 seconds	72 MB	Yes	Standard CPU
SIFT	89%	4.2 seconds	156 MB	No	Standard CPU
SURF	91%	2.8 seconds	134 MB	No	Standard CPU
ORB	78%	0.65 seconds	28 MB	No	Standard CPU
ResNet-50	95%	3.2 seconds	512 MB	Possible	GPU Required
Perceptual Hash	72%	0.15 seconds	8 MB	No	Standard CPU

Source: developed by the authors.

Our hybrid approach balances these trade-offs effectively, providing 92 percent accuracy with 1.8 seconds processing time using standard CPU resources. The system processes the target workload efficiently while maintaining responsive user experience for interactive content exploration. Language variant detection achieves robust performance with 89 percent overall accuracy across different naming conventions and language combinations.

The multi-scale approach proves particularly effective for diagrams containing both fine-grained details and overall structural patterns. Evaluation using English, Russian, and Kazakh diagram variants shows that the language detection approach identifies actual variants while maintaining a false positive rate below 3 percent. The system successfully handles various naming conventions, including standard language suffixes, embedded language names, and mixed separators. characters.

While maintaining computational efficiency suitable for standard office environments. The tiered similarity framework enables users to distinguish between exact duplicates, functionally similar content, and broadly related material without requiring technical expertise in similarity assessment algorithms.

Language variant detection addresses a genuine organizational need that has received limited attention in previous research. The capability to automatically identify translated versions of diagrams facilitates knowledge reuse across language boundaries and reduces redundant content creation efforts. Organizations using the system report measurable improvements in content management efficiency and cost savings from reusing existing translated materials. The modular architecture enables incremental enhancement without disrupting existing functionality. Security measures address common web application vulnerabilities while maintaining usability for legitimate organizational workflows. Response times under 2 seconds support interactive content exploration patterns that encourage user engagement with the similarity search capabilities.

Performance comparison with alternative approaches confirms the advantages of hybrid methods that combine multiple complementary techniques. Pure algorithmic approaches either sacrifice accuracy for speed or require substantial computational resources that may not be available in typical organizational environments. Deep learning methods achieve higher theoretical accuracy but demand specialized infrastructure and extensive training data that may not be practical for specialized organizational content.

The choice of traditional computer vision techniques proves appropriate for the target application domain. While more sophisticated approaches might achieve marginal accuracy improvements, the

interpretability and resource efficiency of established methods provide practical advantages that outweigh theoretical performance gains. The 261-dimensional feature vector representation strikes an effective balance between descriptiveness and computational tractability.

Limitations and Future Work. Several aspects of the current implementation warrant consideration for future development while acknowledging practical deployment constraints. The system relies primarily on visual features extracted through traditional computer vision techniques. Integration of optical character recognition could enhance understanding of textual content within diagrams, supporting more sophisticated cross-language similarity assessment, though this would increase computational requirements and implementation complexity. Similarity calculation employs fixed weighting parameters that provide reasonable performance across diverse content types. Adaptive parameter adjustment based on content analysis or user feedback could improve performance for specialized applications. Machine learning approaches for automatic parameter optimization represent a direction that could adapt the system to different organizational domains without manual tuning. Current language variant detection focuses on filename pattern analysis, which may miss variants following unconventional naming schemes. Organizations with inconsistent naming practices or legacy content with non-standard conventions might experience reduced detection accuracy. Extension to analyze extracted text content could provide more robust variant detection capabilities. The in-memory storage approach limits scalability to collections fitting within available system memory. While adequate for the target scenario of 66 weekly diagrams, larger organizations with extensive existing content would require persistent storage solutions with sophisticated indexing strategies. Additionally, organizational environments increasingly demand solutions addressing content integrity, provenance tracking, and multi-stakeholder trust requirements. Previous research in blockchain-enhanced integrity verification demonstrates potential for decentralized content authentication while maintaining computational efficiency [13]. Privacy-preserving mechanisms warrant exploration for sensitive visual content, building upon frameworks developed for permissionless blockchain networks where transparency and confidentiality must be carefully balanced [14]. Furthermore, decentralized governance models could address access control challenges in multi-stakeholder organizational environments, particularly through authentication mechanisms that preserve user privacy while ensuring appropriate content access rights [15]. Cloud deployment architectures could provide elastic scaling. capabilities for organizations experiencing variable workloads or rapid content growth.

Conclusion. This research presents a practical approach to image similarity search that addresses organizational requirements while maintaining computational efficiency suitable for standard hardware deployment. The multi-scale feature extraction algorithm combines complementary visual cues to achieve robust similarity assessment across diverse content types commonly encountered in organizational settings. The implementation within a Flask-based web application demonstrates the feasibility of deploying computer vision algorithms using standard web technologies and open-source libraries. The modular architecture facilitates maintenance and enhancement while providing clear pathways for future development based on evolving organizational needs.

The language variant detection capability represents a practical contribution that addresses multi-language organizational challenges. This functionality enables more effective content discovery and reuse across linguistic boundaries, supporting international organizations and reducing duplication of effort in content creation processes. The experimental validation confirms effectiveness for the target application domain while identifying specific enhancement opportunities for future research. The balance of accuracy, efficiency, and interpretability achieved by our approach makes it suitable for practical deployment while providing a foundation for continued development.

The successful integration of traditional computer vision techniques within modern web application frameworks demonstrates the continued relevance of established algorithmic approaches when properly adapted to contemporary deployment requirements. This work provides a template for future research in organizational content management systems that prioritizes practical utility alongside technical innovation.

Funding. This research has been funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. BR21882260).

REFERENCES

1 Smeulders, A. W., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), 1349-1380. DOI: https://doi.org/10.1109/34.895972

2 Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2), 1-60. DOI: https://doi.org/10.1145/1348246.1348248

3 Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91-110. DOI: https://doi.org/10.1023/B:VISI.0000029664.99615.94

4 Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3), 346-359. DOI: https://doi.org/10.1016/j.cviu.2007.09.014

5 Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. R. (2011). ORB: An efficient alternative to SIFT or SURF. *Proceedings of the 2011 International Conference on Computer Vision*, Barcelona, 6-13 November 2011, 2564-2571. DOI: 10.1109/ICCV.2011.6126544

6 Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84-90. DOI: https://doi.org/10.1145/3065386

7 Dubey, S. R. (2021). A decade survey of content based image retrieval using deep learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(5), 2687-2704. DOI: https://doi.org/10.1109/TCSVT.2021.3080920

8 Kumar, J., Ye, P., & Doermann, D. (2014). Structural similarity for document image classification and retrieval. *Pattern Recognition Letters*, 43, 119-126. DOI: https://doi.org/10.1016/j.patrec.2013.10.030

9 Gao, S. H., Cheng, M. M., Zhao, K., Zhang, X. Y., Yang, M. H., & Torr, P. (2021). Res2Net: A new multi-scale backbone architecture. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(2), 652-662. DOI: https://doi.org/10.1109/TPAMI.2019.2938758

10 Albesher, L., Alfayez, R. (2024). An observational study on Flask web framework questions on Stack Overflow (SO). *IET Software*, 2024(1), 1905538, 15 p. DOI: https://doi.org/10.1049/sfw2/1905538

11 Grinberg, M. (2018). Flask Web Development: Developing Web Applications with Python (2nd ed.). Sebastopol, CA: O'Reilly Media, Inc., 250 p. ISBN 978-1-491-99173-2

12 Biswas, R., & Blanco-Medina, P. (2021). State of the art: Image hashing. *Computer Science. Computer Vision and Pattern Recognition*. arXiv:2108.11794, 8 p. DOI: https://doi.org/10.48550/arXiv.2108.11794

13 Bayan, T., Banach, R., Nurbekov, A., Galy, M. M., Sabyrbayev, A., & Nurbekova, Z. (2024). Blockchain-enhanced Integrity Verification in Educational Content Assessment Platform: A Lightweight and Cost-Efficient Approach. *Computer Science. Cryptography and Security*. arXiv:2405.13156, 17 p. DOI: https://doi.org/10.48550/arXiv.2409.19828

14 Bayan, T., & Banach, R. (2023). Exploring the Privacy Concerns in Permissionless Blockchain Networks and Potential Solutions. In *2023 IEEE International Conference on Smart Information Systems and Technologies (SIST)*, Astana, Kazakhstan, 2023, 567-572. DOI: 10.1109/SIST58284.2023.10223536

15 Bayan, T., & Banach, R. (2024). A Privacy-Preserving DAO Model Using NFT Authentication for the Punishment not Reward Blockchain Architecture. *Cryptography and Security*. arXiv:2405.13156, 16 p. DOI: https://doi.org/10.48550/arXiv.2405.13156

* Баян Т.[1], Ғалы М. М.[2], Зияшев А.А.[3]
[1] Назарбаев Университеті
[2,3] Agoda Co Ltd
[1] *Қазақстан, Астана*
[2,3] *Таиланд, Бангкок*

ЦИФРЛЫҚ ФАСИЛИТАТОРЛАР ЖЕЛІСІ ЖҮЙЕСІНДЕГІ КЕСКІН ІЗДЕУ МОДУЛІ: АРХИТЕКТУРА, АЛГОРИТМДЕР ЖӘНЕ ІСКЕ АСЫРУ

Аңдатпа

Көптілді кең көлемді визуалды құжаттаманы басқаратын цифрлық фасилитатор желілер бірдей диаграммалар әртүрлі тілдік нұсқаларда болған кезде байланысты контентті іздеу қиындықтарына тап болады. Дәстүрлі метадеректерге негізделген іздеу тәсілдері визуалды ұқсас материалдарды анықтай алмайды, бұл контентті қайта пайдалану тиімділігін төмендетеді. Бұл зерттеу аптасына шамамен 66 диаграмманы өңдейтін ұйымдық орталар үшін кескін ұқсастығын іздеу жүйесін іске асыруды қарастырады. Мақсат стандартты жабдықта есептеу тиімділігін сақтай отырып, нақты сәйкестіктер мен тілдік нұсқаларды дәл анықтауды қамтамасыз ету болып табылады. Жүйе RGB және HSV кеңістіктерінде түс гистограммасын талдауды, Canny операторларын қолдана отырып көп деңгейлі шеттерді анықтауды, Sobel градиенті арқылы текстураны сипаттауды және үлгі қолтаңбаларын біріктіретін көп деңгейлі ерекшелік алуды қолданады, бұл 261 өлшемді ерекшелік векторын құрайды. Деңгейлі ұқсастықты бағалау шеңбері кескіндер арасындағы қатынастарды бағалайды, ал тілдік нұсқаларды анықтау файл атауларының үлгілерін талдауды визуалды ұқсастықты тексерумен біріктіреді. Іске асыру компьютерлік көру операциялары үшін OpenCV бар Flask фреймворкін қолданады. Ағылшын, орыс және қазақ тілдеріндегі ұйымдық диаграммалармен сынау нақты сәйкестіктер мен тілдік нұсқаларды анықтауда 92% дәлдікті көрсетеді, орташа жауап беру уақыты 1,8 секунд және жадының шыңы 72 МБ құрайды. Тілдік нұсқаларды анықтау 89% дәлдікке жетеді, жалған оң нәтижелер 3% төмен. Модульдік архитектура қарапайым офистік жүйелерде орналастыруға мүмкіндік береді, бұл дәстүрлі компьютерлік көру тәсілдері практикалық шектеулерге дұрыс бейімделген кезде ұйымдық контентті басқару үшін қолданылатынын көрсетеді.

Түйінді сөздер: кескін іздеу, цифрлық көмекші желілер, компьютерлік көру, Flask, CBIR, көп деңгейлі ерекшелік алу, тілдік нұсқаларды анықтау

*Баян Т.[1], Галы М. М.[2], Зияшев А.А.[3]
[1] Назарбаев Университет
[2, 3] Agoda Co Ltd
[1] *Казахстан, Астана*
[2, 3] *Таиланд, Бангкок*

МОДУЛЬ ПОИСКА ИЗОБРАЖЕНИЙ В СИСТЕМЕ ЦИФРОВОЙ ФАСИЛИТАТОРСКОЙ СЕТИ: АРХИТЕКТУРА, АЛГОРИТМЫ И РЕАЛИЗАЦИЯ

Аннотация

Цифровые фасилитаторские сети, управляющие обширной визуальной документацией на нескольких языках, сталкиваются с трудностями поиска связанного контента, когда идентичные диаграммы существуют в различных языковых версиях. Традиционные подходы поиска на основе метаданных не могут идентифицировать визуально похожие материалы, что снижает эффективность повторного использования контента. Данное исследование рассматривает реализацию системы поиска изображений по сходству для организационных сред, обрабатывающих примерно 66 диаграмм еженедельно. Цель состоит в обеспечении точной идентификации точных совпадений и языковых вариантов при сохранении вычислительной эффективности на стандартном оборудовании. Система использует многомасштабное извлечение признаков, объединяющее анализ цветовых гистограмм в пространствах RGB и HSV, многопороговое обнаружение границ с операторами Canny, характеристику текстуры градиентами Sobel и сигнатуры шаблонов, формируя 261-мерное представление признаков. Уровневая структура оценки сходства анализирует отношения между изображениями, тогда как обнаружение языковых вариантов сочетает анализ шаблонов имен файлов с проверкой визуального сходства. Реализация использует фреймворк Flask с OpenCV для операций компьютерного зрения. Тестирование с организационными диаграммами на английском, русском и казахском языках демонстрирует 92% точность в идентификации точных совпадений и языковых вариантов при среднем времени отклика 1,8 секунды и пиковом использовании памяти 72 МБ. Обнаружение языковых вариантов достигает 89% точности при доле ложноположительных результатов ниже 3%. Модульная архитектура обеспечивает развертывание на обычных

офисных системах, демонстрируя, что традиционные подходы компьютерного зрения остаются применимыми для управления организационным контентом при надлежащей адаптации к практическим ограничениям.

Ключевые слова: поиск изображений, цифровые фасилитаторские сети, компьютерное зрение, Flask, многомасштабное извлечение признаков, обнаружение языковых вариантов

Conference Proceedings:
Received: 13.09.2025
Approved after peer review: 28.09.2025
Accepted for publication: 17.10.2025