Beyond Search: Interactive Image Exploration and Retrieval via a Unified Similarity Graph

Kai Uwe Barthel
Visual Computing Group
HTW Berlin
Berlin, Germany
0000-0001-6309-572X

Peter Eisert
Vision and Imaging Technologies
Fraunhofer HHI
Berlin, Germany
0000-0001-8378-4805

Konstantin Schall
Visual Computing Group
HTW Berlin
Berlin, Germany
0000-0003-3548-0537

Florian Tim Barthel
Vision and Imaging Technologies
Fraunhofer HHI
Berlin, Germany
0009-0004-7264-1672

Nico Hezel
Visual Computing Group
HTW Berlin
Berlin, Germany
0000-0002-3957-4672

Klaus Jung
Visual Computing Group
HTW Berlin
Berlin, Germany
0000-0002-3600-6848

***Abstract*—Navigu 2.0 is the latest iteration of a system designed for fast, scalable visual exploration of large, unstructured image collections. It combines an intuitive interface with an efficient architecture optimized for real-time user interaction. The new version introduces a significant architectural change to a single, unified similarity graph. Previously, three distinct graphs were used for visual appearance, semantic meaning, and color composition of the images. An improved CLIP-based encoder generates one joint feature vector per image, integrating visual and semantic similarities. This unified embedding, used in a Dynamic Exploration Graph, enables faster and more accurate search results for both text and image queries while simplifying the backend architecture. In response to a search query, the system retrieves a subgraph of related images and displays them as a visually sorted 2D grid. This grid is created using a novel version of the Fast Linear Assignment Sorting algorithm working solely on pairwise distances. Users can interact with this grid by dragging to explore neighboring regions of the graph or zooming to navigate between broad concepts and fine-grained similarities. These three enhancements provide a unified, efficient solution for cross-modal image search and extensive visual data exploration. The live demo is available at www.navigu.net.**

***Index Terms*—Image Graph, Exploration, Visualization, Image Retrieval, Similarity Search**

I. Technical Details

Motivation and System Description

Even with the progress made by visual deep learning models in image representation and similarity search, exploring large image collections is still challenging and time-consuming, often needing repeated queries and manual refinement. To accommodate human perceptual limitations, most image search systems display only a small subset of results, increasing the likelihood of not finding better matches. Grouping images by visual similarity reduces search time by allowing users to view more images at once [1], [2]. Dimensionality reduction methods such as t-SNE [3] and UMAP [4] project high-dimensional vectors into 2D, but they struggle with large datasets - often resulting in uneven and overlapping spacing, which limits their suitability for image layouts.

Fig. 1. A Navigu screenshot shows the result of a text search for ”balloon,” returning nine images representing different result clusters on the left; the hot air balloon image was selected.

In this demo, we present a system for efficient visual search and interactive exploration of large image collections (Fig. 1). Based on feature vector similarity, the system constructs a hierarchical graph where images are represented as vertices. This structure supports multi-level exploration, providing broad conceptual overviews at higher levels and fine-grained visual similarities at lower levels. A single feature vector per image supports both visual and textual search.

For any query, the closest match is identified, and a local subgraph of similar images is retrieved and arranged using a fast layout algorithm that aligns visual similarity with spatial proximity. User interaction relies on intuitive gestures: dragging reveals more related images (Fig. 2), zooming enables seamless transitions between high-level conceptual overviews and detailed examinations of fine-grained similarities, double-clicking starts a new search. Cached layouts allow to revisit past views, providing a map-like exploration.

The main contributions of this demo, going beyond the previous version of Navigu [5], are:

1. Improved joint feature vectors using a fine-tuned CLIP-based encoder enable unified visual and textual search.
2. A single similarity graph replaces the previous three graphs.
3. Improved sorting algorithm based on pairwise distances enables faster, similarity-preserving image layouts.

A demo video can be found at https://youtu.be/m-cLj5gTn9A.

Implementation

The system consists of a client-server web application. The front-end is a browser-based GUI rendered on a single HTML canvas, supporting real-time interaction and map-like navigation. The back-end handles session-based graph exploration and visually sorts images of a subgraph in a grid. A feature extraction module runs as a separate service, encoding all images into joint feature vectors. Users can initiate three types of queries: text-to-image, image-to-image with uploaded images, and in-graph queries. During navigation, the client transmits viewport and interaction data; the server responds by retrieving relevant images from the graph and updating the layout accordingly.

Feature Vectors are generated using *Multi-Caption-Image-Pairing* (MCIP) [6], an enhanced CLIP-based encoder optimized for cross-modal retrieval. It improves image retrieval quality while maintaining compatibility with the original text encoder for both text-to-image and image-to-image queries. Each image is encoded as a joint feature vector used for graph construction and layout.

The **Image Graph** is constructed using a new hierarchical version of the *Dynamic Exploration Graph* (DEG) [7]. Optimized for approximate nearest neighbor search in high-dimensional spaces, DEG guarantees full connectivity and efficient traversal while supporting dynamic updates.

Visual Image Sorting uses an improved *Fast Linear Assignment Sorting* (FLAS) method [2]. This new version uses pairwise distances, cutting computation time significantly by removing dependence on feature vector dimensionality while maintaining local similarity in the 2D layout. Retrieved subgraphs form coherent grids where existing images stay fixed and new results fill empty spots (see Fig. 2), ensuring smooth and stable navigation.

II. Relevance

This demo is highly relevant to the EUSIPCO community, bridging core topics like image processing, deep learning, optimization, and data retrieval. It offers a high-performance implementation that not only advances research in visual search and exploration but also holds potential for applications wherever large volumes of images need to be viewed, such as e-commerce, digital archives, and media management.

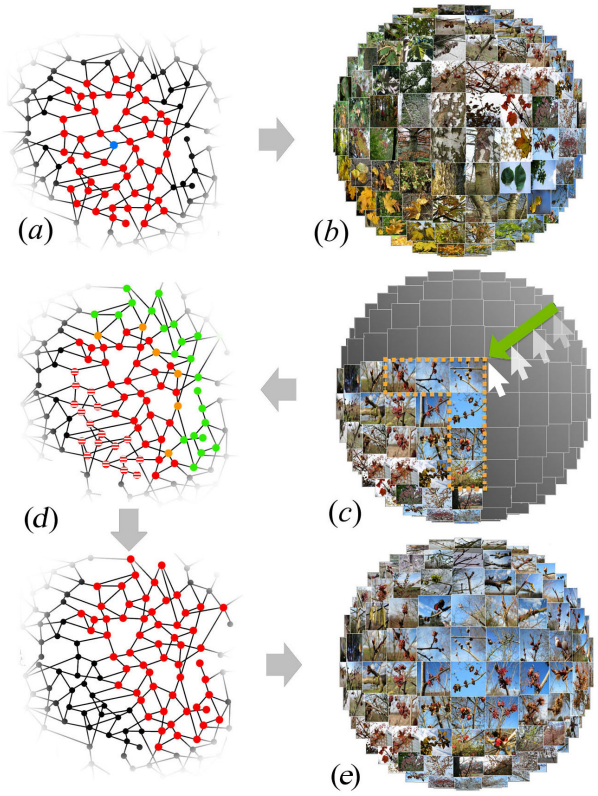

Fig. 2. The image collection is organized as a similarity graph with images as vertices. During a text or visual search, the closest match is identified, and related, connected images (red) are retrieved (a). These are shown on a 2D map for interactive exploration (b). Dragging the map causes some images to move out of view (c), triggering the system to fetch related images from the opposite border (d) and fill empty spaces with newly arranged results (e).

III. Logistics

We would need a table for our computer along with chairs, ideally a large monitor, power supply, and internet access, preferably via LAN. It would be nice to have a display panel for an explanatory poster.

References

[1] K. Schoeffmann and D. Ahlström, "Similarity-based visualization for image browsing revisited," in *ISM*. IEEE CS, 2011, pp. 422–427.

[2] K. U. Barthel, N. Hezel, K. Jung, and K. Schall, "Improved evaluation and generation of grid layouts using distance preservation quality and linear assignment sorting," *Computer Graphics Forum*, vol. 42, no. 1, pp. 261–276, 2023.

[3] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[4] L. McInnes and J. Healy, "Umap: Uniform manifold approximation and projection for dimension reduction." *CoRR*, 2018.

[5] K. U. Barthel, N. Hezel, K. Schall, and K. Jung, "navigu.net: Navigation in visual image graphs gets user-friendly," in *Proceedings of the 2023 ACM International Conference on Multimedia Retrieval*, ser. ICMR '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 654–658.

[6] K. Schall, K. U. Barthel, N. Hezel, and K. Jung, "Optimizing clip models for image retrieval with maintained joint-embedding alignment," in *Similarity Search and Applications*. Cham: Springer Nature Switzerland, 2024, pp. 97–110.

[7] N. Hezel, B. Schilling, K. Barthel, K. Schall, and K. Jung, *Adapting the Exploration Graph for High Throughput in Low Recall Regimes*, 10 2024, pp. 283–290.