Tracing Semantic Provenance: Multimodal Fuzzy Linkography for Long-Term AI-Assisted Visual Ideation

WEN-FAN WANG, National Taiwan University, Taiwan
YI-TING CHIU, National Taiwan University, Taiwan
MAX KREMINSKI, Cornell University, USA

AI-assisted visual ideation tools enable the rapid generation of artifacts, yet standard history tools fail to capture the semantic evolution of a user's intent across mixed modalities (text and images). This paper proposes Multi-modal Fuzzy Linkography as a method to analyze creative activity traces captured from GenTune [2], a system that supports "traceable prompts" and precise element control. By defining design moves that encompass text edits, full image generation, and region-of-interest (ROI) selection, we construct a semantic graph of the design process. We apply this method to a dataset of long-term GenTune usage to identify patterns of fixation, divergence, and convergence that are invisible in traditional linear histories.

CCS Concepts: • **Human-centered computing → Interactive systems and tools**; **User centered design**.

Additional Key Words and Phrases: Linkography, Creativity Support Tool, Visual Exploration

ACM Reference Format:
Wen-Fan Wang, Yi-Ting Chiu, and Max Kreminski. 2026. Tracing Semantic Provenance: Multimodal Fuzzy Linkography for Long-Term AI-Assisted Visual Ideation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–April 17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 3 pages.

1 introduction

Recent advances in human–AI co-creation in the visual domain, such as AIdeation [3] and GenTune [2], support expressive visual exploration and enable designers to manipulate elements within generated images. However, although users can now control individual artifacts, they often lack semantic provenance that reveals how their intentions evolve over time. Existing provenance tools typically track parent–child lineage or present artifacts in chronological order, but they fail to capture semantic relationships across temporally distant design moves.

To address this gap, we propose Multi-modal Fuzzy Linkography, inspired by Fuzzy Linkography [1], which uses computational models to estimate semantic similarity between design moves. Unlike prior approaches that rely on manual coding or text-only analysis, our method leverages embedding models to automatically compute semantic connections across heterogeneous interaction types—including text prompts, generated images, and specific visual regions—captured from real-world GenTune sessions.

Authors' Contact Information: Wen-Fan Wang, vann@cmlab.csie.ntu.edu.tw, National Taiwan University, Taipei, Taiwan; Yi-Ting Chiu, R13922018@csie.ntu.edu.tw, National Taiwan University, Taipei, Taiwan; Max Kreminski, maxkreminski@gmail.com, Cornell University, Ithaca, New York, USA.

2 Data Source: The GenTune Logs

Our analysis leverages interaction logs from GenTune, specifically data collected from long-term users of the system. GenTune's architecture is uniquely update-rich for linkography because it captures granular "moves" beyond simple text prompting. The system logs user actions including:

- **User Instruction:** Writing or revising high-level instructions for image generation (e.g., "Design a teenager's room").
- **Element Refinement:** Selecting specific visual elements or Regions of Interest (ROIs) and requesting targeted revisions (e.g., selecting the "vintage music corner" and adding a guitar).
- **Reference Integration:** Incorporating external references to combine with or revise selected elements or ROIs.

We map these rich image and prompt data logs into a linkography structure to analyze the "shape" of the design process.

3 Methodology: Multi-modal Fuzzy Linkography

To operationalize these logs, we extend the concept of Fuzzy Linkography to support a multi-modal definition of "design moves."

3.1 Defining the Multi-modal Design Move

In traditional linkography, a move is often a single verbal utterance or sketch. To reflect the reality of AI-assisted visual exploration, we conceptualize a "design move" as a composite unit of user intent. A move in our framework is not just a text prompt; it encapsulates the state of the canvas at that moment. This includes Text Nodes (user prompts), Image Nodes (generated outputs), and crucially, ROI Nodes (specific visual regions selected by the user) . By treating a pinned visual element (e.g., a specific lighting effect) as a discrete node, we can trace how a specific visual concept persists or evolves throughout a session.

3.2 Computing Semantic Links

The core of our methodology is measuring the semantic distance between these diverse moves. We map all design elements, whether text descriptions or visual pixels, into a shared high-dimensional embedding space. This allows us to compute a "fuzzy" link strength between any two moves, regardless of their modality. For example, a text prompt generated in Move 5 can be semantically linked to a visual region selected in Move 50, revealing a recurrence of intent that a linear history would miss.

3.3 Addressing the Modality Gap

A significant challenge in multi-modal analysis is that the statistical distribution of similarity scores often differs between modalities (e.g., text-to-text vs. image-to-image). To address this "modality gap," we apply statistical normalization to the raw similarity scores. This ensures that the definition of a "strong link" is consistent across the graph, allowing us to build a unified network where text and visual nodes connect seamlessly based on their semantic content rather than their data format.

4 Analysis Goals

By applying this multi-modal linkography to the GenTune logs, we aim to transform raw interaction history into a graph of semantic relationships. This allows us to identify specific cognitive patterns in AI-assisted workflows:

- Fixation: Identifying "saturation webs" where a user creates many tightly interlinked moves without significant semantic deviation.
- Divergence without Integration: Identifying widely scattered exploration paths that fail to converge.
- Late-Stage Convergence: Detecting when distant ideas (e.g., a "guitar" from Move 3 and "lighting" from Move 20) are successfully synthesized into a final artifact.

5 Conclusion

We present Multi-modal Fuzzy Linkography as a lens for understanding the complex, non-linear pathways of AI-assisted visual exploration. By rigorously defining multi-modal moves and normalizing semantic similarity across the GenTune logs, we can reveal the hidden structure of creative ideation, paving the way for tools that not only generate content but help users make sense of their own process.

References

[1] Amy Smith, Barrett R Anderson, Jasmine Tan Otto, Isaac Karth, Yuqian Sun, John Joon Young Chung, Melissa Roemmele, and Max Kreminski. 2025. Fuzzy Linkography: Automatic Graphical Summarization of Creative Activity Traces. In *Proceedings of the 2025 Conference on Creativity and Cognition*. 637–650.

[2] Wen-Fan Wang, Ting-Ying Lee, Chien-Ting Lu, Che-Wei Hsu, Nil Ponsa Campany, Yu Chen, Mike Y Chen, and Bing-Yu Chen. 2025. GenTune: Toward Traceable Prompts to Improve Controllability of Image Refinement in Environment Design. *arXiv preprint arXiv:2508.15227* (2025).

[3] Wen-Fan Wang, Chien-Ting Lu, Nil Ponsa Campanyà, Bing-Yu Chen, and Mike Y Chen. 2025. AIdeation: Designing a Human-AI Collaborative Ideation System for Concept Designers. *arXiv preprint arXiv:2502.14747* (2025).