

# Annotated Bibliography

Leslie Osei-Anane

January 28, 2026

## References

- [1] Boris Chen, Amir Ziai, Rebecca S. Tucker, and Yuchen Xie. Match cutting: Finding cuts with smooth visual transitions. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2115–2125, January 2023.

This paper tackles the practical problem of finding high-quality match cuts, where two shots are connected by similar framing, composition, or action. The authors introduce a modular pipeline that filters and ranks candidate shot pairs at scale using visual, audio, and audio-visual features, and they evaluate classification and metric-learning approaches on a labeled dataset of about 20,000 shot pairs. The main contribution is a scalable system for match-cut discovery along with a released dataset, embeddings, and code, which makes the problem measurable and reproducible. This is a strong scholarly source because it is a peer-reviewed WACV 2023 paper with a clear experimental setup and public resources, which allows the community to verify results and build on the benchmarks. It is also highly relevant to my project because it operationalizes “smooth visual transition” with concrete feature choices and evaluation procedures. For my MatchCut Compiler idea, this paper provides concrete signals and metrics for automatically scoring candidate match cuts by motion, shape, framing, and semantic similarity; I can extend their approach to produce “receipts” (explanations) and rank alternatives.

- [2] Shixing Chen, Xiaohan Nie, David Fan, Dongqing Zhang, Vimal Bhat, and Raffay Hamid. Shot contrastive self-supervised learning for scene boundary detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9796–9805, June 2021.
- [3] Matt Deitke, Eli VanderBilt, Alvaro Herrasti, Luca Weihs, Kiana Ehsani, Jordi Salvador, Winson Han, Eric Kolve, Aniruddha Kembhavi, and Roozbeh Mottaghi. Procthor: Large-scale embodied ai using procedural generation. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

ProcTHOR introduces a procedural generation framework for creating large numbers of diverse, fully interactive indoor environments for embodied AI. The

system samples floorplans, populates them with a sizable library of interactive objects, and supports randomized layouts, materials, and lighting to scale scene diversity. The paper contributes both the ProcTHOR generation pipeline and a complementary artist-designed test set (ArchitecTHOR), and it shows that training on large procedurally generated collections improves performance and generalization across multiple embodied AI benchmarks. This is a strong scholarly source because it is a NeurIPS 2022 peer-reviewed contribution with extensive empirical evidence and a clear methodological contribution to scalable environment creation, which is central to embodied AI research. For my Holodeck-like idea, ProcTHOR offers a concrete example of deterministic compilation from structured specs into simulated environments and a template for evaluating those environments in embodied tasks.

- [4] Alejandro Pardo, Fabian Caba Heilbron, Juan León Alcázar, Ali Thabet, and Bernard Ghanem. Moviecuts: A new dataset and benchmark for cut type recognition. In *Computer Vision – ECCV 2022*, volume 13667 of *Lecture Notes in Computer Science*, pages 668–685. Springer, 2022.
- [5] Yue Yang, Fan-Yun Sun, Luca Weihs, Eli VanderBilt, Alvaro Herrasti, Winson Han, Jiajun Wu, Nick Haber, Ranjay Krishna, Lingjie Liu, Chris Callison-Burch, Mark Yatskar, Aniruddha Kembhavi, and Christopher Clark. Holodeck: Language guided generation of 3d embodied ai environments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16227–16237, June 2024.

Holodeck proposes a system that converts a natural-language prompt into a 3D interactive environment for embodied AI. The method uses a large language model to draft object lists and spatial relations, then optimizes object placement to satisfy those constraints while populating scenes with assets from Objaverse. The paper contributes a full prompt-to-scene pipeline, human evaluations showing preference over procedural baselines in residential scenes, and demonstrations of training agents in novel, automatically generated environments. This is a credible and relevant source because it is a peer-reviewed CVPR 2024 paper with detailed methodology, large-scale evaluation, and clear empirical comparisons, making it a strong foundation for research on language-guided environment generation. For my Holodeck-like idea, the paper directly validates prompt-to-world/spec generation and suggests how to make the compilation deterministic by explicitly emitting constraints and solving them, then evaluating in simulated environments.