Annotated Bibliography

Leslie Osei-Anane

January 29, 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 targets automatic discovery of high-quality match cuts, where two shots align in framing, composition, or action to create a smooth transition. The authors present a modular pipeline that generates candidate shot pairs and ranks them using visual, audio, and audio-visual descriptors, and they compare classification and metric-learning models on a labeled dataset of roughly 20,000 pairs. A key contribution is a scalable, reproducible benchmark with released data, learned embeddings, and evaluation metrics for match-cut retrieval. As a peer-reviewed WACV 2023 publication, the work is credible and methodologically transparent, with controlled experiments and publicly released artifacts that support verification and reuse. It is relevant because it operationalizes "smooth visual transition" into measurable features and evaluation protocols that can be adopted directly in new systems. This connects to the MatchCut Compiler concept, which aims to automatically score candidate match cuts using motion, shape, framing, and semantic similarity, and to provide concise explanations ("receipts") and ranked alternatives for editors.

[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 large library of interactive objects, and randomizes layouts, materials, and lighting to scale scene diversity. The paper contributes both the ProcTHOR generation pipeline and a complementary artist-designed evaluation set (ArchitecTHOR), and it reports 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 peer-reviewed NeurIPS 2022 contribution with extensive empirical evidence and a clear methodological advance in scalable environment creation, which is central to embodied AI research. It also clarifies trade-offs between procedural controllability and realism that matter when selecting generation settings. This connects to the Holodeck-like concept by providing a concrete example of deterministic compilation from structured scene specifications into simulated environments and by outlining evaluation protocols for 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 presents a system that converts a natural-language prompt into an interactive 3D environment for embodied AI. The approach uses a large language model to draft object inventories 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 preference studies that outperform procedural baselines on residential scenes, and demonstrations that agents can be trained in the generated environments. This is a credible source because it is a peer-reviewed CVPR 2024 paper with a clear methodology, extensive evaluations, and transparent comparisons to baselines, which strengthens confidence in the reported gains. It is relevant to research on language-guided environment generation and controllable simulation. This connects to the Holodeck-like prompt-to-world/spec generation concept by showing how constraints can be explicitly represented, solved deterministically, and evaluated in simulated environments to assess downstream agent performance.