PerforMagic: Coordinating Bodily Performance and Camera Movement in AI Video Creation

Yufeng Zeng
Computational Media and Arts Thrust
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, Guangdong, China
yzeng683@connect.hkust-gz.edu.cn

Troy TianYu Lin
Computational Media and Arts Thrust
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, Guangdong, China
tlin932@connect.hkust-gz.edu.cn

Shuyue Li
Computational Media and Arts Thrust
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, Guangdong, China
sli688@connect.hkust-gz.edu.cn

Yue Hong
Robotics and Autonomous Systems Thrust
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, Guangdong, China
yhong252@connect.hkust-gz.edu.cn

Anca-Simona Horvath
Computational Media and Arts Thrust
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, Guangdong, China
ancahorvath@hkust-gz.edu.cn

Chen Liang*
Computational Media and Arts Thrust
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, Guangdong, China
lliangchenc@163.com

Figure 1: With PerforMagic, even solo creators without professional equipment or collaborators can direct cinematic camera movements through self-filmed input. The workflow involves: a) providing the story as input, after which the system suggests suitable camera movements and provides filming guidance; b) re-filming the performance in a controllable and immersive virtual filming environment; and c) generating an AI-enhanced video that refines the shot with emotional and stylistic expression.

*Corresponding author.

CHI EA '26, Barcelona, Spain

ACM ISBN 979-8-4007-2281-3/26/04
https://doi.org/10.1145/3772363.3798782

Abstract

Performance-driven AI video creation allows users to express ideas through bodily action, but it also requires performers to coordinate embodied performance with camera movement. For solo creators, managing this coordination in real time can be cognitively demanding and limits cinematic exploration. We present PerforMagic, an interactive system that explores how coordination between performance and camera movement can be supported through a re-filming interaction. Users first record their performance without explicitly

controlling the camera. The system then reconstructs the performance in a virtual environment, enabling users to reason about and explore camera movements in relation to their embodied actions. Through a user study, participants reported clearer relationships between performance and camera movement, less pressure during performance, and broader exploration of camera techniques. Our work contributes an interaction concept for coordinating embodied action and cinematography in performance-driven AI video creation and offers qualitative insights into how staged interaction supports expressive control.

CCS Concepts

• **Human-centered computing** → **Interactive systems and tools**; **Interaction design**.

Keywords

AIGC, AI Video Creation, Authoring Tool, Personalized Cinematic Authoring

ACM Reference Format:

Yufeng Zeng, Troy TianYu Lin, Shuyue Li, Yue Hong, Anca-Simona Horvath, and Chen Liang. 2026. PerforMagic: Coordinating Bodily Performance and Camera Movement in AI Video Creation. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3772363.3798782

1 INTRODUCTION AND BACKGROUND

Recent generative AI video tools have enabled creators to incorporate bodily performance into video generation workflows, allowing expressive intent to be articulated through enacted movement rather than abstract textual description [13, 32]. By recording performance videos as conditional inputs, users can transform their actions into videos featuring different characters, scenes, or visual styles. Film theory has long emphasized that cinematic meaning emerges not from action alone, but from how camera movement frames, emphasizes, and interprets embodied motion [7–10].

For non-professional users, however, coordinating performance and camera movement remains difficult in practice. When filming themselves, users must perform while anticipating how camera movement will shape the final result, often under fixed camera setups and limited feedback. Prior work shows that such concurrent demands increase cognitive load and creative pressure, particularly for novices with limited cinematographic knowledge [27, 30]. As a result, camera decisions are often simplified or avoided altogether.

Advances in accessible motion capture and pose estimation have lowered the barrier to performance-driven creation by enabling motion input through commodity RGB cameras [16, 32]. Hybrid workflows further combine automatic reconstruction with manual refinement to correct imperfect motion estimates [19]. Building on these techniques, embodied animation and puppeteering systems map bodily actions to virtual characters or objects, supporting expressive control through movement [12, 15, 18, 25, 29, 31]. Across this body of work, bodily performance primarily serves as input for motion generation, while camera behavior is typically fixed or treated as secondary.

In parallel, research on virtual cinematography has explored algorithmic camera planning and re-rendering along novel viewpoints [6, 14, 35]. Recent generative video models further demonstrate increasing capability in synthesizing diverse camera movements [6, 11, 24], including approaches that integrate camera parameters or high-level instructions into the generation process [17, 24]. End-to-end video generation tools enable rapid creation from text, images, or short clips [1–3], while controllable generation techniques condition synthesis on pose or motion signals [36]. Despite these advances, camera intentions are typically specified in advance or embedded implicitly in generation pipelines, offering little support for revisiting camera decisions after performance, particularly for exploratory and early-stage creation scenarios.

When user-recorded performance is used as the primary input, achieving desired camera effects often depends on the original footage already containing appropriate camera movement. For self-filming users relying on static cameras, this creates a mismatch between embodied expression and cinematographic control, forcing users to anticipate camera outcomes during performance without effective means to revise them afterward.

Together, this body of work reveals a gap not in generation capability, but in how performers are supported in reasoning about camera movement in relation to their own actions.

To inform the design of such a workflow, we conducted a formative study with six participants involving self-filming and camera movement planning. The findings revealed difficulties in reasoning about camera movement during performance, heightened pressure when managing both simultaneously, and limited opportunities to explore camera techniques in relation to one's own actions.

Based on these insights, we present *PerforMagic*, an interactive system that supports coordination between embodied performance and camera movement through a re-filming workflow. By allowing users to perform first and design camera movements afterward in relation to reconstructed performance, PerforMagic reframes camera authoring as a post-performance creative activity. We evaluated the system through a within-subjects study with nine participants, demonstrating improved alignment between performance and camera movement, reduced pressure during performance, and increased exploration of camera techniques.

2 PRELIMINARY STUDY

We conducted a formative study with six non-professional creators experienced in AI video generation to explore challenges in performance-driven self-filming. Participants were asked to record short self-filmed performances intended for AI-based video creation and reflect on how they planned and executed camera movements in relation to their performance.

Based on this study, we identified a design space for performance-driven AI video creation in which systems should (1) reduce the cognitive and physical burden of coordinating performance and filming during self-recording, and (2) support users in developing and refining camera ideas by providing visually grounded guidance and enabling exploration of camera movement in relation to embodied performance after recording. PerforMagic is designed to explore this space through a re-filming workflow.

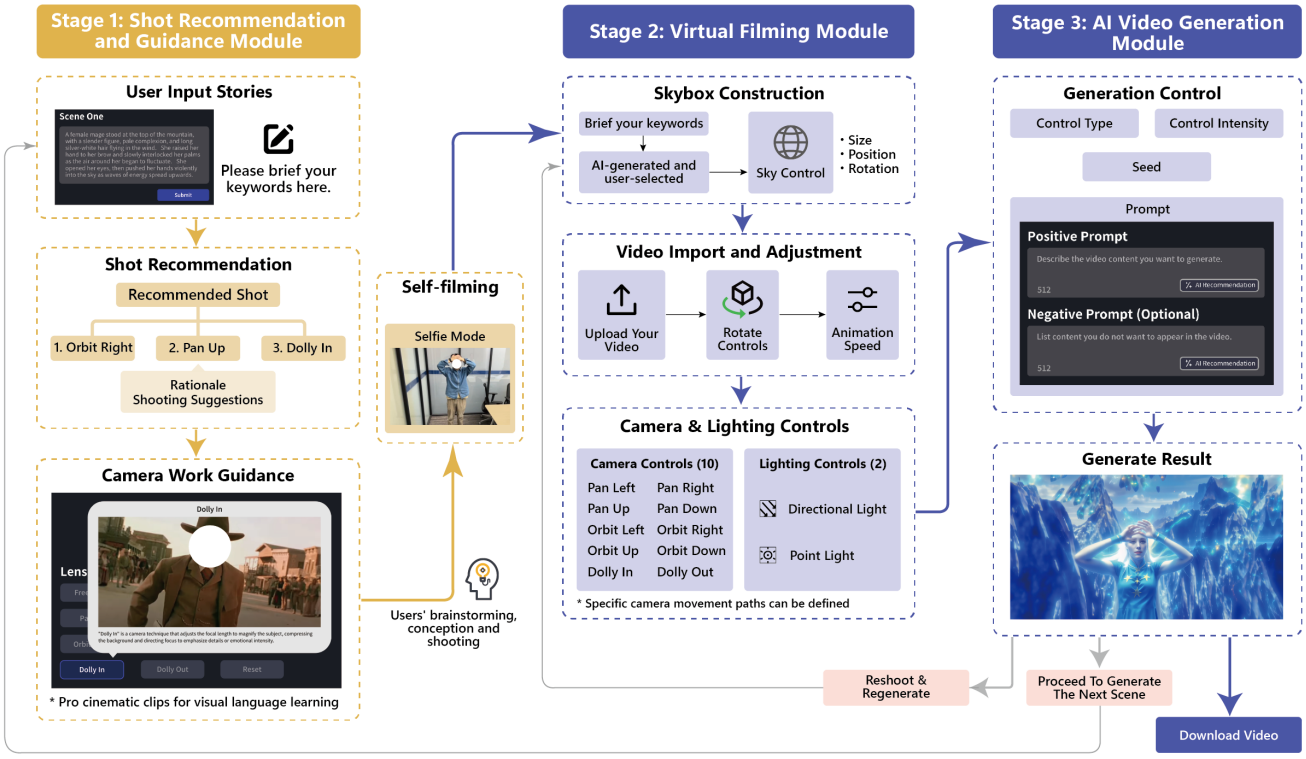

Figure 2: The overall user workflow of PerforMagic, an interactive AI video creation system structured into three stages.

3 PerforMagic OVERVIEW

PerforMagic provides an authoring workflow that takes recorded performance as input, reconstructs it as an animated representation, supports camera movement design through virtual re-filming, and generates AI videos that combine performance and camera intentions, as shown in Fig. 2.

3.1 Shot Recommendation and Guidance

PerforMagic supports early-stage shot planning through a card-based storyboard interface that externalizes narrative structure and camera intention. Each storyboard card represents a scene and requires a short plot description authored by the user. Based on this description, the system suggests three suitable camera movements selected from a predefined set of cinematic shot types. Rather than prescribing a single solution, PerforMagic presents multiple plausible camera interpretations of the same scene, encouraging users to reflect on how different camera movements shape narrative emphasis and emotional tone.

For each suggested shot, the system provides a brief expressive rationale and concise shooting guidance, along with expert-curated film clips that visually demonstrate the intended cinematic effect. These visual references ground abstract camera concepts in concrete examples, helping users develop an intuitive understanding of camera language without requiring prior cinematography expertise. Technically, shot recommendations are generated using an instruction-based prompting strategy with the Qwen large language model [5], which interprets narrative descriptions and aligns them with appropriate camera movements. Storyboard cards and selected shots persist across later re-filming and generation stages, ensuring continuity of narrative and camera intent throughout the workflow.

3.2 Virtual Re-filming

PerforMagic provides a virtual re-filming environment that allows users to explore camera movement and composition after completing their physical performance. Recorded performances are reconstructed as a manipulable three-dimensional animation and placed into a virtual scene. This setup enables users to revisit their embodied action from a third-person perspective and experiment with alternative cinematographic interpretations without re-performing.

Users begin by configuring the virtual environment through a panoramic skybox generated with the Photon text-to-image model [23], along with adjustable lighting to establish scene atmosphere. The user's performance is reconstructed using a monocular motion capture pipeline that combines GVHMR for full-body motion [26] and HaMeR for detailed hand motion [22]. These outputs are unified through the SMPL-X body model [21], producing a single animation that can be replayed, reoriented, and temporally adjusted within a web-based three-dimensional interface implemented using Three.js and WebGL.

Camera authoring is supported through multiple interaction modes, including predefined camera presets for rapid setup, keyframe-based planning relative to selected body joints, and a free-camera mode for exploratory framing. Once the camera configuration is finalized, users can preview and export virtual shots for downstream AI video generation, ensuring continuity across the re-filming workflow.

3.3 Video Generation

PerforMagic generates final videos by combining the re-filmed animation with narrative descriptions and selected camera movements. Generation is implemented using CogVideoX [33] with ControlNet conditioning, guided by full-body pose maps extracted using DW Pose [34]. An additional LoRA module is applied to improve visual quality, allowing users to iteratively refine outputs while preserving prior creative intent.

4 PerforMagic EVALUATION

We conducted a within-subjects user study with nine participants (five female, four male; aged 22–29), recruited through an internal social media platform at a local university. All participants had prior experience using AI-based video generation tools, which ensured familiarity with common generation workflows and minimized training overhead.

Participants completed a performance-driven video creation task using three workflows: PerforMagic and two baseline modes from Runway, Video-to-Video and Keyframe Animation. These two baseline modes were selected to represent two distinct forms of performance-driven AI video creation that differ in how bodily action is involved. Video-to-Video requires users to perform a continuous action that is directly transformed into generated video, whereas Keyframe Animation relies on a small number of discrete poses, resulting in a non-continuous mode of bodily participation. Together, the baselines capture contrasting ways in which body movement is incorporated into AI video generation.

Across all workflows, participants worked with the same set of short, single-character scripts. These scripts were designed to foreground bodily performance and camera movement planning while maintaining a consistent narrative context. For each workflow, participants self-filmed their performance and generated a final video using the corresponding tool. The order of workflows was counterbalanced using a Latin square design to mitigate ordering effects.

After completing each workflow, participants reflected on their creation process in post-task semi-structured interviews. The interviews focused on how participants reasoned about camera movement, how they coordinated performance with camera design, and what constraints they experienced under each workflow. Interview transcripts were analyzed to identify recurring patterns in participants' strategies, difficulties, and experiential differences across the three approaches.

4.1 System Guidance and the Limits of Instruction

Participants described that PerforMagic's system guidance provided concrete starting points for reasoning about camera movement, helping them articulate initial cinematic ideas from narrative intent. As P8 noted, *"Some of these clips felt familiar to me, so they made me want to imitate them. When I imitate, I have a reference in mind."* This form of guidance reduced ideation pressure and enabled users to engage with camera language earlier in the creative process. At the same time, participants often treated model-generated suggestions as authoritative, a pattern also observed in prior work on novice interaction with generative systems [30]. While such reliance can accelerate early exploration, it may also limit deeper understanding of when and why specific camera movements are appropriate. These observations suggest that guidance in AI video tools may benefit from emphasizing interpretation and comparison rather than instruction, supporting gradual sense-making rather than one-off knowledge transfer.

4.2 Performance and Filming for Immersive Engagement

By separating bodily performance from camera authoring, participants described a different experience of enactment during video generation. Participants reported that deferring cinematographic decisions allowed them to concentrate more fully on embodied action, without continuously monitoring filming outcomes. As P9 explained, while performing in V2V mode, her focus gradually moved from *"how to perform well"* to *"how to achieve the intended camera movement."* This staging resonates with prior findings that sustained engagement and active involvement strengthen users' sense of agency in creative systems [28, 37]. Reconstructing performance as an editable representation further reframed filming as an interpretive activity applied after action rather than a constraint imposed during recording. As a result, participants described revisiting and coordinating performance and camera movement more deliberately across stages of video generation.

4.3 Camera Control and Exploration in Video Generation

PerforMagic reframed camera control as an explicit stage within performance-driven video generation rather than an implicit outcome of prompt-based synthesis. Participants described experiencing less uncertainty about how camera movements would be interpreted by the generative model when re-filming reconstructed performances, making trial and error feel less costly. For example, P6 initially planned to use the suggested orbit-up shot. After reviewing the reconstructed performance, she realized that the buildup phase, when the character lowered their head, created stronger visual tension. While adjusting the camera from a closer view to a gradual pull-back, she sensed this shift and revised her plan. This predictability encouraged users to explore alternative camera movements and refine expressive intent through comparison and iteration. Such hands-on control aligns with prior work showing that direct manipulation supports learning and creative engagement in generative systems [4]. In contrast, workflows that compress camera decisions into generation prompts prioritize efficiency but restrict opportunities for reflective exploration and incremental refinement [20, 28]. As P1 described, in keyframe mode her intended camera move from a frontal to a side view was replaced by

the character rotating its head, which differed from her original intention.

5 LIMITATIONS AND FUTURE WORK

PerforMagic currently supports camera authoring at a coarse level, which was sufficient for exploring coordination between performance and filming, but limits more nuanced cinematographic expression. Extending the system with visual trajectory editing or embodied camera input could allow users to reason about camera movement with greater precision. In addition, our studies focused on individual, short-term use; future work should examine how performance-driven re-filming supports longer-term learning, expert practice, and collaborative video creation.

6 CONCLUSION

This work examined camera movement as a distinct design space within performance-driven AI video generation, focusing on how users coordinate bodily action and filming decisions. PerforMagic explores this space by allowing users to perform first and reason about camera movement afterward, supported by system guidance and a re-filming workflow. Our observations indicate that this staged interaction shaped how participants approached camera expression after performance, enabling more deliberate reflection on the relationship between movement and cinematography. Rather than optimizing generation efficiency alone, PerforMagic highlights the potential of structuring creative workflows to invite reflection, exploration, and expressive control in AI video creation.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No. 62502411.

References

[1] 2025. *KlingAI.* https://app.klingai.com/global/ Accessed: 2025-09-11.

[2] 2025. *OpenAI Sora Explore Videos.* https://sora.chatgpt.com/explore/videos Accessed: 2025-09-11.

[3] 2025. *Runway.* https://runwayml.com/ Accessed: 2025-09-11.

[4] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300233

[5] Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, Binyuan Hui, Luo Ji, Mei Li, Junyang Lin, Runji Lin, Dayiheng Liu, Gao Liu, Chengqiang Lu, Keming Lu, Jianxin Ma, Rui Men, Xingzhang Ren, Xuancheng Ren, Chuanqi Tan, Sinan Tan, Jianhong Tu, Peng Wang, Shijie Wang, Wei Wang, Shengguang Wu, Benfeng Xu, Jin Xu, An Yang, Hao Yang, Jian Yang, Shusheng Yang, Yang Yao, Bowen Yu, Hongyi Yuan, Zheng Yuan, Jianwei Zhang, Xingxuan Zhang, Yichang Zhang, Zhenru Zhang, Chang Zhou, Jingren Zhou, Xiaohuan Zhou, and Tianhang Zhu. 2023. Qwen Technical Report. *arXiv preprint arXiv:2309.16609* (2023).

[6] Jianhong Bai, Menghan Xia, Xiao Fu, Xintao Wang, Lianrui Mu, Jinwen Cao, Zuozhu Liu, Haoji Hu, Xiang Bai, Pengfei Wan, et al. 2025. ReCamMaster: Camera-Controlled Generative Rendering from A Single Video. *arXiv preprint arXiv:2503.11647* (2025).

[7] Andre Bazin and Hugh Gray. 1971. *What Is Cinema, Volume II.* University of California Press, Berkeley, CA.

[8] David Bordwell. 2005. *Figures Traced in Light: On Cinematic Staging.* University of California Press, Berkeley, CA. https://books.google.co.kr/books?id=T2aWxNkBP5EC

[9] Blain Brown. 2016. *Cinematography: Theory and Practice: Image Making for Cinematographers and Directors* (3rd ed.). Routledge, New York. 1–421 pages. doi:10.4324/9781315667829

[10] Gilles Deleuze. 1986. *Cinema 1: The Movement-Image.* University of Minnesota Press, Minneapolis.

[11] Yufan Deng, Ruida Wang, Yuhao Zhang, Yu-Wing Tai, and Chi-Keung Tang. 2025. DragVideo: Interactive Drag-Style Video Editing. In *Computer Vision – ECCV 2024*, Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (Eds.). Springer Nature Switzerland, Cham, 183–199.

[12] Andrew Feng, Samuel Shin, and Youngwoo Yoon. 2022. A Tool for Extracting 3D Avatar-Ready Gesture Animations from Monocular Videos. In *Proceedings of the 15th ACM SIGGRAPH Conference on Motion, Interaction and Games* (Guanajuato, Mexico) *(MIG '22)*. Association for Computing Machinery, New York, NY, USA, Article 7, 7 pages. doi:10.1145/3561975.3562953

[13] Brett A. Halperin, Diana Flores Ruíz, and Daniela K. Rosner. 2025. Underground AI? Critical Approaches to Generative Cinema through Amateur Filmmaking. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1141, 18 pages. doi:10.1145/3706598.3713342

[14] Rui He, Huaxin Wei, and Ying Cao. 2024. An Interactive System for Supporting Creative Exploration of Cinematic Composition Designs. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 78, 15 pages. doi:10.1145/3654777.3676393

[15] Ching-Wen Hung, Ruei-Che Chang, Hong-Sheng Chen, Chung Han Liang, Liwei Chan, and Bing-Yu Chen. 2022. Puppeteer: Exploring Intuitive Hand Gestures and Upper-Body Postures for Manipulating Human Avatar Actions. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology* (Tsukuba, Japan) *(VRST '22)*. Association for Computing Machinery, New York, NY, USA, Article 13, 11 pages. doi:10.1145/3562939.3565609

[16] Dong-Hyun Hwang, Kohei Aso, Ye Yuan, Kris Kitani, and Hideki Koike. 2020. MonoEye: Multimodal Human Motion Capture System Using A Single Ultra-Wide Fisheye Camera. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 98–111. doi:10.1145/3379337.3415856

[17] Yunxin Li, Haoyuan Shi, Baotian Hu, Longyue Wang, Jiashun Zhu, Jinyi Xu, Zhen Zhao, and Min Zhang. 2024. Anim-Director: A Large Multimodal Model Powered Agent for Controllable Animation Video Generation. In *SIGGRAPH Asia 2024 Conference Papers* (Tokyo, Japan) *(SA '24)*. Association for Computing Machinery, New York, NY, USA, Article 34, 11 pages. doi:10.1145/3680528.3687688

[18] Jingyuan Liu, Hongbo Fu, and Chiew-Lan Tai. 2020. PoseTween: Pose-driven Tween Animation. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) *(UIST '20)*. Association for Computing Machinery, New York, NY, USA, 791–804. doi:10.1145/3379337.3415822

[19] Jingyuan Liu, Li-Yi Wei, Ariel Shamir, and Takeo Igarashi. 2024. iPose: Interactive Human Pose Reconstruction from Video. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 945, 14 pages. doi:10.1145/3613904.3641944

[20] Allan MacLean, Richard Young, Victoria Bellotti, and Thomas Moran. 1991. Questions, Options, and Criteria: Elements of Design Space Analysis. *Human-Computer Interaction* 6 (09 1991), 201–250. doi:10.1080/07370024.1991.9667168

[21] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. 2019. Expressive Body Capture: 3D Hands, Face, and Body from a Single Image. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.

[22] Georgios Pavlakos, Dandan Shan, Ilija Radosavovic, Angjoo Kanazawa, David Fouhey, and Jitendra Malik. 2024. Reconstructing Hands in 3D with Transformers. In *CVPR*.

[23] Photographer. 2025. *Photon [Stable Diffusion model].* https://civitai.com/models/84728/photon CivitAI.

[24] Xuanchi Ren, Tianchang Shen, Jiahui Huang, Huan Ling, Yifan Lu, Merlin Nimier-David, Thomas Müller, Alexander Keller, Sanja Fidler, and Jun Gao. 2025. GEN3C: 3D-Informed World-Consistent Video Generation with Precise Camera Control. arXiv:2503.03751 [cs.CV] https://arxiv.org/abs/2503.03751

[25] Nazmus Saquib, Rubaiat Habib Kazi, Li-Yi Wei, and Wilmot Li. 2019. Interactive Body-Driven Graphics for Augmented Video Performance. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300852

[26] Zehong Shen, Huaijin Pi, Yan Xia, Zhi Cen, Sida Peng, Zechen Hu, Hujun Bao, Ruizhen Hu, and Xiaowei Zhou. 2024. World-Grounded Human Motion Recovery via Gravity-View Coordinates. In *SIGGRAPH Asia 2024 Conference Papers*. 1–11.

[27] Ben Shneiderman. 2000. Creating creativity: user interfaces for supporting innovation. *ACM Trans. Comput.-Hum. Interact.* 7, 1 (March 2000), 114–138. doi:10.1145/344949.345077

[28] Ben Shneiderman. 2007. Creativity support tools: accelerating discovery and innovation. *Commun. ACM* 50, 12 (Dec. 2007), 20–32. doi:10.1145/1323688.1323689

[29] Hariharan Subramonyam, Wilmot Li, Eytan Adar, and Mira Dontcheva. 2018. TakeToons: Script-driven Performance Animation. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (Berlin, Germany) *(UIST '18)*. Association for Computing Machinery, New York, NY, USA, 663–674. doi:10.1145/3242587.3242618

[30] Huanchen Wang, Tianrun Qiu, Jiaping Li, Zhicong Lu, and Yuxin Ma. 2025. HarmonyCut: Supporting Creative Chinese Paper-cutting Design with Form and Connotation Harmony. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 661, 22 pages. doi:10.1145/3706598.3714159

[31] Nora S. Willett, Wilmot Li, Jovan Popovic, and Adam Finkelstein. 2017. Triggering Artwork Swaps for Live Animation. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 85–95. doi:10.1145/3126594.3126596

[32] Vasco Xu, Chenfeng Gao, Henry Hoffmann, and Karan Ahuja. 2024. MobilePoser: Real-Time Full-Body Pose Estimation and 3D Human Translation from IMUs in Mobile Consumer Devices. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 70, 11 pages. doi:10.1145/3654777.3676461

[33] Zhuoyi Yang, Jiayan Teng, Wendi Zheng, Ming Ding, Shiyu Huang, Jiazheng Xu, Yuanming Yang, Wenyi Hong, Xiaohan Zhang, Guanyu Feng, Da Yin, Yuxuan Zhang, Weihan Wang, Yean Cheng, Bin Xu, Xiaotao Gu, Yuxiao Dong, and Jie Tang. 2025. CogVideoX: Text-to-Video Diffusion Models with An Expert Transformer. arXiv:2408.06072 [cs.CV] https://arxiv.org/abs/2408.06072

[34] Zhendong Yang, Ailing Zeng, Chun Yuan, and Yu Li. 2023. Effective Whole-body Pose Estimation with Two-stages Distillation. arXiv:2307.15880 [cs.CV] https://arxiv.org/abs/2307.15880

[35] Mark YU, Wenbo Hu, Jinbo Xing, and Ying Shan. 2025. Trajectorycrafter: Redirecting camera trajectory for monocular videos via diffusion models.

[36] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. arXiv:2302.05543 [cs.CV] https://arxiv.org/abs/2302.05543

[37] Mingxu Zhou, Dengming Zhang, Weitao You, Ziqi Yu, Yifei Wu, Chenghao Pan, Huiting Liu, Tianyu Lao, and Pei Chen. 2024. StyleFactory: Towards Better Style Alignment in Image Creation through Style-Strength-Based Control and Evaluation. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 107, 15 pages. doi:10.1145/3654777.3676370