



FilmAgent: Automating Virtual Film Production Through a Multi-Agent Collaborative Framework

Zhenran Xu
Harbin Institute of Technology
China
xuzhenran.hitsz@gmail.com

Jifang Wang
Harbin Institute of Technology
China
23S151116@stu.hit.edu.cn

Longyue Wang
Dublin City University
Ireland
vincentwang0229@gmail.com

Zhouyi Li
Tsinghua University
China
lizhouyi23@mails.tsinghua.edu.cn

Senbao Shi
Harbin Institute of Technology
China
shisenbaohit@gmail.com

Baotian Hu*
Harbin Institute of Technology
China
hubaotian@hit.edu.cn

Min Zhang
Harbin Institute of Technology
China
zhangmin2021@hit.edu.cn

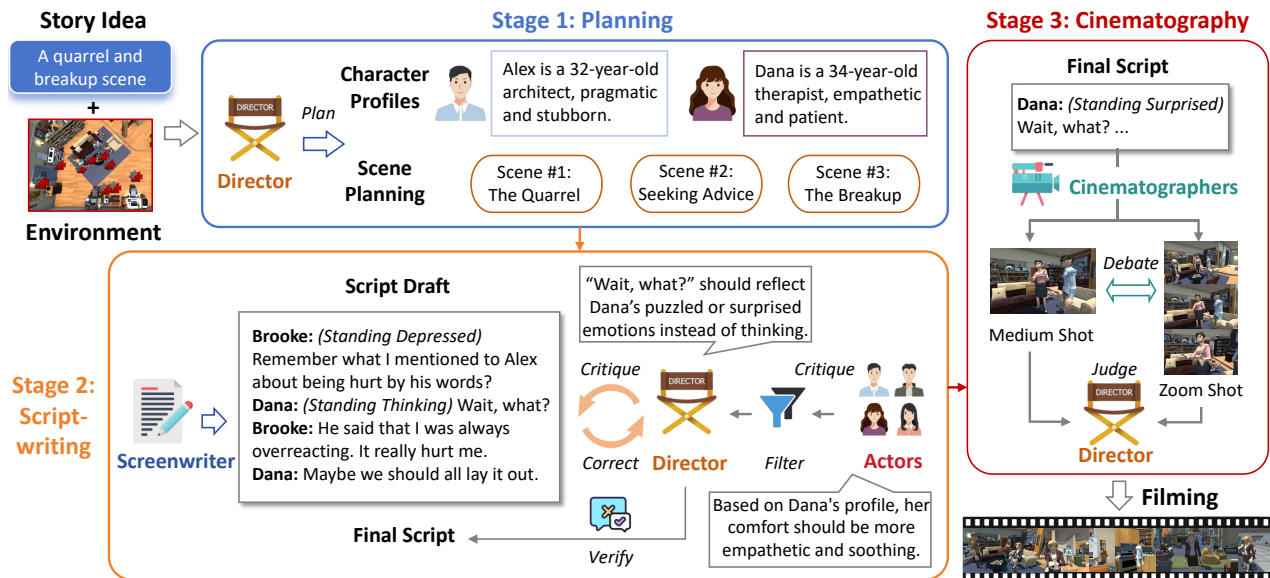


Figure 1: Workflow of our film production system FilmAgent. Given a 3D environment and a story idea, the director first creates potential character profiles, and converts the idea into a scene outline. Next, actors, the screenwriter and the director collaborate to develop the dialogue and choreograph movements. Then multiple cinematographers design and discuss the camera setups for each line, with director making final decisions. Finally, the film is shot within our constructed 3D environment.

*Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
SA Technical Communications '24, December 03–06, 2024, Tokyo, Japan
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-1140-4/24/12
<https://doi.org/10.1145/3681758.3698014>

Abstract

Virtual film production requires intricate decision-making processes, including scriptwriting, virtual cinematography, and precise actor positioning and actions. Remarkable progress in automated decision-making have utilized agent societies powered by large language models (LLMs). This paper introduces FilmAgent, a novel LLM-based multi-agent collaborative framework designed to automate and streamline the film production process. FilmAgent simulates key crew roles—directors, screenwriters, actors, and cinematographers—within a sandbox environment, integrating efficient human workflows. The process is divided into three stages: planning, scriptwriting, and cinematography. Each stage engages

a team of film crews providing iterative feedback, thus verifying intermediate results and reducing errors. Our evaluation of generated videos reveals that collaborative FilmAgent significantly outperforms individual efforts in line consistency, script coherence, character actions, and camera settings. Further analysis highlights the importance of feedback and verification in reducing hallucinations, enhancing script quality, and improving camera choices. We hope that this project lays the groundwork and shows the potential of integrating LLMs into creative multimedia tasks¹.

CCS Concepts

• **Information systems** → **Multimedia content creation.**

Keywords

multi-agent system, large language model, virtual cinematography

ACM Reference Format:

Zhenran Xu, Jifang Wang, Longyue Wang, Zhouyi Li, Senbao Shi, Baotian Hu, and Min Zhang. 2024. FilmAgent: Automating Virtual Film Production Through a Multi-Agent Collaborative Framework. In *SIGGRAPH Asia 2024 Technical Communications (SA Technical Communications '24)*, December 03–06, 2024, Tokyo, Japan. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3681758.3698014>

1 Introduction

Virtual film production requires a disciplined approach to directing, camera placement and actor positioning [He et al. 2023]. Films are produced through the dialogues spoken by the characters, the screenplays that outline the story, and the guidance given by directors [Jiang et al. 2020, 2024]. Therefore, filmmaking is fundamentally a communication-driven collaborative task, motivating our design of a multi-agent system based on large language models (LLMs). Multiple agents can work together and accomplish more complex tasks than a single agent can, showing the emergence of collective intelligence [Du et al. 2023; Guo et al. 2024; Xu et al. 2023].

Inspired by these developments, we propose FilmAgent, the first LLM-based multi-agent collaborative framework to automate virtual film production. In this framework, LLM-based agents fulfill various film crew roles such as directors, screenwriters, actors, and cinematographers. As shown in Figure 1, the collaborative process emulates the human workflow and divides the process into planning, scriptwriting, and cinematography. In the planning stage, the director starts with a story idea and develops character profiles, expanding it into a detailed scene outline that specifies the where, what, and who of each segment. During scriptwriting, the director, screenwriter, and actors collaborate on dialogue development and choreograph movements. In the cinematography stage, the cinematographers and director design camera setups, choosing between static and dynamic shots to effectively convey the narrative visually. To facilitate virtual production, we have meticulously built a 3D environment, including 15 locations, 21 actions, 65 designated actor positions, 272 static and dynamic shots, and speech audio generation. In addition, we propose two multi-agent collaboration algorithms, *Critique-Correct-Verify* and *Debate-Judge*, used in scriptwriting and cinematography stages respectively.

¹For more information, including open-source Unity environment, codes and videos, please visit our project page at <https://filmagent.github.io/>.

Human evaluations of the generated videos validate the effectiveness of our framework. The results show that the collaborative FilmAgent significantly outperforms single-agent efforts across four aspects: plot coherence, alignment between dialogue and actor profiles, appropriateness of camera setting, and accuracy of actor actions. Further preference analysis underscores the importance of feedback and verification in correcting inaccuracies, enhancing plot coherence and improving camera choices. This project lays the groundwork for automated virtual film production, showing the potential of collaborative AI agents in this creative domain.

In summary, our main contributions are as follows:

- We introduce FilmAgent, the first LLM-based multi-agent collaborative framework for automating virtual film production, with a well-crafted 3D environment.
- We incorporate two collaboration strategies within the workflow, which substantially reduces hallucinations and enhances the quality of scripts and camera settings.
- Extensive human evaluations validate FilmAgent, indicating LLM-based multi-agent collaboration as a promising avenue for automating virtual film production.

2 FilmAgent

2.1 Overview

FilmAgent is an LLM-based multi-agent framework for automated virtual film production in a sandbox environment². The whole process is illustrated in Figure 1. Clear role specialization allows for the breakdown of complex work into smaller and more specific tasks [Hong et al. 2024; Li et al. 2023]. In FilmAgent, we define four main characters: **Director**, **Screenwriter**, **Actor** and **Cinematographer**. Each of these roles carries its own set of responsibilities.

The **Director** initiates and oversees the entire filmmaking project. This role includes setting character profiles, planning video outlines, providing feedback on the script, engaging in discussions with crew members, and making final decisions when conflicts arise. The **Screenwriter's** responsibilities go beyond writing dialogue; they also specify the positioning and actions for each line, and continuously update the script to ensure it is coherent, captivating, and well-structured, based on the Director's critiques. **Actors** are responsible for making minor adjustments to their lines based on their profiles, ensuring the dialogue aligns with the characters, and communicating necessary changes to the Director. **Cinematographers** select the camera settings for each line according to shot usage guidelines, collaborate with peers to compare and discuss these choices, and ensure the appropriateness of camera settings.

2.2 Agent Collaboration Strategies

In this section, we introduce two collaboration strategies employed in this work, including *Critique-Correct-Verify* and *Debate-Judge*. The pseudo codes are in supplementary materials.

Critique-Correct-Verify Collaboration. This strategy involves two agents working collaboratively. First, the *Action agent* P generates a response R based on the given context C and instruction I. Next, the *Critique agent* Q reviews the response R and writes critiques F highlighting potential areas for improvement. The Action agent P

²An introduction of our constructed 3D environment is in supplementary materials.

then integrates the critiques and corrects the response. Finally, the *Critique agent* Q evaluates the updated response R to determine whether the critiques F have been adequately addressed or if further iterations are necessary.

Debate-Judge Collaboration. The *Debate-Judge* strategy involves multiple agents who propose their responses and then engage in a debate to persuade each other. A third-party agent ultimately concludes the discussion and delivers the final judgment. During each iteration, two *peer agents* P and Q independently generate their responses and then critique each other’s work. Based on the critiques received, each agent may revise their response or maintain the original. After several rounds of debate, the *Judgment agent* J synthesizes the discussion and formulates the final judgment R.

2.3 Workflow

In Figure 1, we divide the virtual film production process into three sequential stages: **planning**, **scriptwriting** and **cinematography**.

In the **planning** stage, from a brief story idea, the director generates various character profiles that could be relevant to the story. The profiles include key attributes such as gender, occupation, and personality traits. Using these profiles and a set of 15 locations in our 3D environment, the director expands the story idea into a detailed scene outline, specifying the where, what, and who of each segment.

Scriptwriting involves three key roles: the screenwriter, the director and the actors. The scriptwriting stage can be divided into three parts: (1) *Initial Draft*: The screenwriter drafts the initial script, including character positioning, dialogue, and actions. (2) *Director-Screenwriter Discussion*: The director and screenwriter engage in a *Critique-Correct-Verify* process. The director (the Critique agent Q) thoroughly reviews the script and provides critiques on the plot coherence and the appropriateness of character actions. The screenwriter (the Action agent P) then revises the script based on the director’s critiques. The director verifies the updated script to determine if further adjustments are needed. (3) *Actor-Director-Screenwriter Discussion*: Actors provide feedback based on their understanding of characters to ensure consistency between the script and character profiles. The director filters and aggregates this feedback, then, in collaboration with the screenwriter, employs the same *Critique-Correct-Verify* cycle to refine the script.

Cinematography involves a collaborative process among two peer cinematographers and the director in the *Debate-Judge* manner to ensure diverse and appropriate camera choices. The two cinematographers (agents P and Q) independently assign their camera choices to each line of the script. They then engage in a debate to address any discrepancies in their choices, refining their decisions as the discussion progresses. After several rounds of debate and revision, the director (the Judgment agent J) summarizes the debate process, synthesizes the final choices from both cinematographers, and ultimately determines the camera decisions.

After these stages, we can simulate the entire script within the constructed 3D environment and begin filming. Each line in the script is specified with the positions of the actors, their actions, and the chosen camera shots. The duration of each line in the video is determined by the length of the corresponding speech audio.

Table 1: Comparison of baselines using human annotations for actor actions, overall plot coherence, script alignment with actor profiles, and appropriateness of camera settings. The evaluation metric for Action is accuracy (0-1), while the others use a 5-point Likert scale (1-5).

	Action	Plot	Profile	Camera
CoT	0.62	1.33	3.26	1.67
FilmAgent (Solo)	0.80	1.87	4.20	2.07
FilmAgent (Group)	0.88	3.53	4.44	3.53

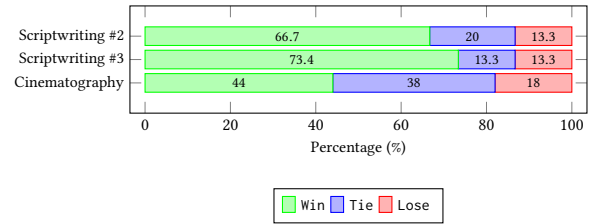


Figure 2: Compared with the original version, the win, tie, and lose rates of the updated script and camera choices after multi-agent collaboration.

3 Experiments

3.1 Experimental Setup

Data. We manually brainstorm 15 story ideas that can be implemented in the locations and action spaces within our constructed 3D environment, such as “a quarrel and breakup scene”, “late night brainstorming for a startup” and “casual meet-up with an old friend”.



Evaluation Scheme. We evaluate the videos across four key aspects: the appropriateness of camera settings, the alignment of the script with actor profiles, the accuracy of actor actions, and the overall plot coherence. For the action aspect, we randomly select 50 actions from the generated scripts and annotate their accuracy. We use a 5-point Likert scale to assess the remaining three aspects.

Baselines. Following the experimental setup of AgentVerse [Chen et al. 2024], to validate the superiority of FilmAgent in facilitating agent collaboration over standalone agents, we compare it against these baselines: (1) **Chain-of-Thought (CoT)**: A single agent generates the chain-of-thought rationale and the complete script. (2) **Solo**: A single agent is responsible for planning, scriptwriting, and cinematography, representing our FilmAgent framework without multi-agent collaboration algorithms. (3) **Group, i.e. the full FilmAgent framework**, utilizing multi-agent collaboration.

3.2 Results

From the results in Table 1, agents configured using FilmAgent (both Solo and Group) consistently outperform the standalone CoT agent. This shows the efficacy of decomposing complex tasks into manageable sub-tasks. We find that the CoT agent struggles with generating accurate camera selections, and often suggests actor movements outside of the action space, leading to low camera and action scores. Comparative analysis between the Solo and Group configurations highlights the benefits of the multi-agent framework. FilmAgent facilitates iterative feedback and revisions

Table 2: Comparisons of the scripts and camera settings before (left) and after (right) multi-agent collaboration, with excerpts from their discussion process. Case #1 is from the Critique-Correct-Verify method in Scriptwriting #2. Case #2 is from the Debate-Judge method in Cinematography.

Case #1	<p>Scene #1 (Roadside) Emma: I'd love that. Where should we meet? Alex: (Standing suggest) There's a cafe just around the corner from here. How about tomorrow at 3? Emma: (Standing happy) Perfect! See you tomorrow. Scene #2 (Alex's living room) Alex: (Standing greeting) Welcome to my humble abode! Make yourself comfortable.</p>	<p>Scene #1 (Roadside) Emma: I'd love that. Where should we meet? Alex: (Standing thinking) How about at my place? Tomorrow at 3? Emma: (Standing happy) Perfect! See you tomorrow. Scene #2 (Alex's living room) Alex: (Standing greeting) Welcome to my humble abode! Make yourself comfortable.</p>
<p>Critiques from the Director: For the reasonableness of actions, {"dialogue": "There's a cafe ...?", "current_action": "Standing suggest", "suggested_revision": "Standing thinking"}. For the fluency of the script, the dialogue in Scene 1 mentions meeting up in cafe, but Scene 2 shows them at Alex's house instead. Consider changing Alex's dialogue to mention catching up at his place to make Scene 2 more natural.</p>		
Case #2	 <p>Tracking Shot</p>	 <p>Medium Shot</p>
<p>The selected shots for the last line in Case #1. Debate from one Cinematographer: Tracking Shot is not applicable as Alex is not moving, violating the guideline of Tracking Shot usage. Instead, the Medium Shot correctly shows Alex's body language.</p>		

through collaboration, leading to significant improvements in all aspects, especially in plot coherence and camera settings.

3.3 Preference Analysis

To further analyze the effectiveness of multi-agent collaboration, we compare 15 scripts before and after *Critique-Correct-Verify*, i.e. *Director-Screenwriter Discussion* (denoted as Scriptwriting #2) and *Actor-Director-Screenwriter Discussion* (denoted as Scriptwriting #3), and 50 randomly-selected modifications on the camera choices before and after *Debate-Judge* in the Cinematography Stage.

The results, shown in Figure 2 as the winning rates of revised scripts, indicate a clear preference by human evaluators for the revised scripts over the original versions. This demonstrates the effectiveness of iterative feedback and verification. For Scriptwriting #2, as illustrated by Case #1 in Table 2, the *Director-Screenwriter discussion* reduces hallucinations in non-existent actions (e.g., standing suggest), enhances plot coherence, and ensures consistency across scenes. For Cinematography, Case #2 shows the correction of an inappropriate dynamic shot, which is replaced with a medium shot to better convey body language. Additionally, multi-agent collaboration improves line consistency with character profiles and increases the diversity of camera choices³.

4 Conclusion

We present FilmAgent, an LLM-based multi-agent framework that automates virtual film production. This framework features a meticulously crafted 3D environment, simulates efficient human workflows, and employs multi-agent collaboration strategies. Human evaluations show the effectiveness of FilmAgent, showing that it

significantly enhances script quality and improves camera selection. These results highlight the potential of FilmAgent to advance virtual film production through multi-agent collaboration.

Acknowledgments

This work is supported by Natural Science Foundation of China (No. 62376067).

References

- Weize Chen, Yusheng Su, Jingwei Zuo, Cheng Yang, Chenfei Yuan, Chi-Min Chan, Heyang Yu, Yaxi Lu, Yi-Hsin Hung, Chen Qian, Yujia Qin, Xin Cong, Ruobing Xie, Zhiyuan Liu, Maosong Sun, and Jie Zhou. 2024. AgentVerse: Facilitating Multi-Agent Collaboration and Exploring Emergent Behaviors. In *The Twelfth International Conference on Learning Representations*. <https://openreview.net/forum?id=EHg5GDnyq1>
- Yilun Du, Shuang Li, Antonio Torralba, Joshua B. Tenenbaum, and Igor Mordatch. 2023. Improving Factuality and Reasoning in Language Models through Multiagent Debate. arXiv:2305.14325 [cs.CL]
- Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V. Chawla, Olaf Wiest, and Xiangliang Zhang. 2024. Large Language Model based Multi-Agents: A Survey of Progress and Challenges. arXiv:2402.01680 [cs.CL]
- Li-wei He, Michael F. Cohen, and David H. Salesin. 2023. *The Virtual Cinematographer: A Paradigm for Automatic Real-Time Camera Control and Directing* (1 ed.). Association for Computing Machinery, New York, NY, USA.
- Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiawu Zheng, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, Chenyu Ran, Lingfeng Xiao, Chenglin Wu, and Jürgen Schmidhuber. 2024. MetaGPT: Meta Programming for A Multi-Agent Collaborative Framework. In *The Twelfth International Conference on Learning Representations*.
- Hongda Jiang, Bin Wang, Xi Wang, Marc Christie, and Baoquan Chen. 2020. Example-driven virtual cinematography by learning camera behaviors. *ACM Trans. Graph.* 39, 4, Article 45 (aug 2020), 14 pages. <https://doi.org/10.1145/3386569.3392427>
- Hongda Jiang, Xi Wang, Marc Christie, Libin Liu, and Baoquan Chen. 2024. Cinematographic Camera Diffusion Model. arXiv:2402.16143 [cs.GR]
- Guohao Li, Hasan Abed Al Kader Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. 2023. CAMEL: Communicative Agents for "Mind" Exploration of Large Language Model Society. In *Thirty-seventh Conference on Neural Information Processing Systems*. <https://openreview.net/forum?id=3lyL2XWDkG>
- Zhenran Xu, Senbao Shi, Baotian Hu, Jindi Yu, Dongfang Li, Min Zhang, and Yuxiang Wu. 2023. Towards Reasoning in Large Language Models via Multi-Agent Peer Review Collaboration. arXiv:2311.08152 [cs.CL]

³Examples of Scriptwriting #3 and Cinematography are in supplementary materials.