Cinemassist: An Intelligent Interactive System for Real-Time Cinematic Composition Design

Rui He
School of Design
The Hong Kong Polytechnic University
Kowloon, Hong Kong
henry.he@connect.polyu.hk

Ying Cao
ShanghaiTech University
Shanghai, China
caoying59@gmail.com

Johan F. Hoorn
School of Design and Department of Computing
The Hong Kong Polytechnic University
Kowloon, Hong Kong
johan.f.hoorn@polyu.edu.hk

Huaxin Wei*
School of Design
The Hong Kong Polytechnic University
Kowloon, Hong Kong
huaxin.wei@polyu.edu.hk

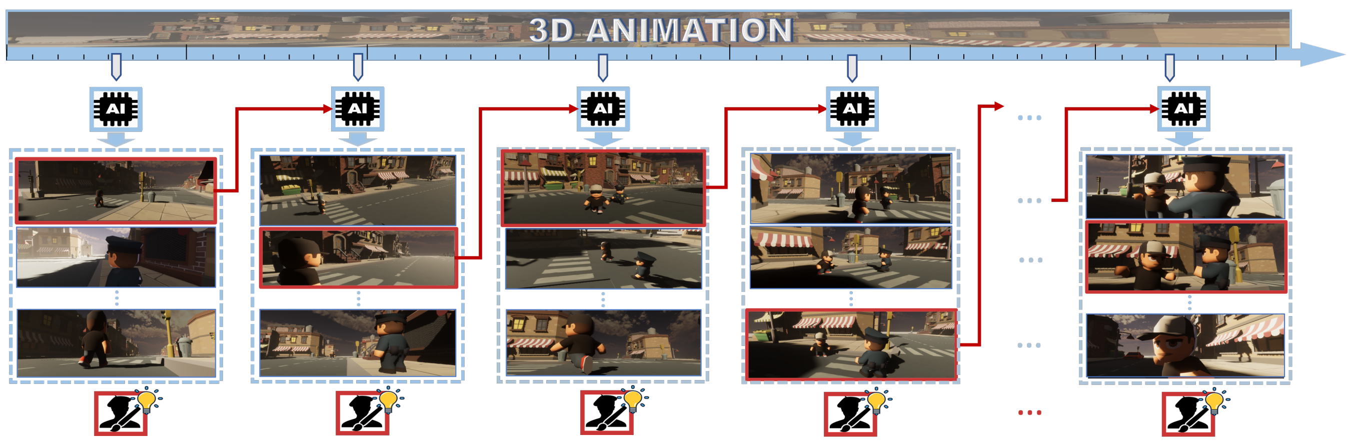

Figure 1: Cinemassist is an interactive system that helps users design cinematic compositions for 3D animations by providing suggestions based on movie genre, intended emotion state, and user-selected key time points. It combines human decision-making with AI-generated suggestions, allowing users to have creative control while benefiting from AI assistance.

ABSTRACT

Designing cinematic compositions, which involves moving cameras through a scene, is essential yet challenging in filmmaking. Machinima filmmaking provides a real-time virtual environment for exploring different compositions, but it still requires significant cinematography skills and creativity. To address this, we introduce Cinemassist, a tool that helps users develop camera trajectories in a 3D animation by generating multiple composition proposals for creative exploration. Preliminary user study indicates that our system can generate useful design suggestions for experts and novices, and facilitate users' exploration and evaluation during the cinematic composition design process.

*Huaxin Wei is the corresponding author.

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3650898

CCS CONCEPTS

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

KEYWORDS

Creativity Support Tool, Digital Filmmaking, Machine Learning, Intelligent Cinematography, Machinima

ACM Reference Format:
Rui He, Ying Cao, Johan F. Hoorn, and Huaxin Wei. 2024. Cinemassist: An Intelligent Interactive System for Real-Time Cinematic Composition Design. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3613905.3650898

1 INTRODUCTION

Designing effective cinematic compositions is a daunting task in filmmaking since it requires extensive cinematographic knowledge, experience and creativity. Recently, machinima filmmaking within virtual environments has become increasingly popular due to its unique advantage of enabling more efficient exploration and evaluation of design alternatives. Unfortunately, cinematic composition design in the machinima environments remains very challenging, as one needs to go through many trials to find an ideal composition for a moment at frame level and ensure the compositions of different moments are connected coherently at the entire scene level [9]. Although many cinematographic rules and referential examples are available, machinima filmmakers still find it difficult to draw inspiration from them in practice. There are a few studies targeted at automating cinematography in virtual environments. However, none of them can serve as effective and ready-to-use solutions to support creativity in the cinematic composition design process.

In this work, we propose Cinemassist, an intelligent creativity support tool that uses machine learning to aid in the design of cinematic compositions in a machinima environment. The tool allows users to design camera poses in a 3D animation by providing conditioning semantics such as movie genre and intended emotion state. Consequently, these design tasks are carried out in an interactive workflow (shown in Figure 1) that alternates between human decision and computer suggestion throughout the design process, so that the user can get inspired by diverse suggested options while enjoying considerable amounts of freedom to customize the design. Particularly, we formed our design requirements for Cinemassist according to our literature survey and the feedback from a field interview with three professional digital filmmakers who use machinima tools in their work. In view of these requirements, Cinemassist was designed and implemented using a deep generative neural network model that is capable of auto-regressively synthesizing a rich variety of plausible camera behaviours for storytelling, conditioned on the input animation and semantics. To enable our model, we construct a dataset of film clips from movies of different genres, with estimated 3D character pose and relative camera annotations. When trained on the dataset, our model essentially acquires the prior knowledge that is necessary to augment users in Cinemassist, without relying on extensive hard-coded rules.

To evaluate the feasibility of our system, we conducted a preliminary user study with both experts and novice machinima filmmakers. The results suggest that Cinemassist can recommend inspiring cinematic composition design alternatives, and improve designers' exploration and self-evaluation in their creative process. Furthermore, based on the study results, we discussed and drew several design implications for future iterations and improvement of the system.

2 BACKGROUND AND RELATED WORK

2.1 Filmmaking in Machinima Environment

In traditional filmmaking, a pivotal stage in the production process is the shot design phase [5]. During this stage, the cinematic composition of every ongoing "beat" [29] of a story scene is meticulously planned using "storyboards" as references for the final imagery implementation [16]. Creative cinematic composition translates a scene from a unique viewing perspective besides the intended emotion expression (referred to as "focalization" in narratology) [1, 8]. To acquire this skill, novice designers often turn to the principles of "cinematography", which encapsulate the fundamental "grammar of the shot" [8]. Conversely, experienced designers tend to expand their design repertoire by studying "film analysis" and drawing inspiration from master exemplars[6, 28, 33].

Machinima is defined as "animated filmmaking within a real-time virtual 3D environment" [27]. Popular machinima environments include commercial digital game engines like Unity and Unreal, as well as 3D filmmaking platforms such as 3ds Max and MAYA. In contrast to traditional filmmaking, which often relies on fixed storyboards, the "real-time" advantage of machinima allows for a more flexible and cost-effective exploration of creative cinematic composition [26]. Given this strength, the film and animation industry has increasingly employed advanced machinima techniques for script rehearsal and the exploration of innovative cinematic composition design ideas before committing to storyboarding [24]. A standout example is the production of the animation film "*EVANGELION:3.0+1.0 Thrice Upon a Time*" [3], which used the machinima tool "Previs" to delve into "unexpected" cinematic effects through the use of 3D character avatars [23]. Nevertheless, despite the advantages offered by machinima, crafting coherent storytelling through composition sequences within the machinima environment remains a challenge [23], which requires machinima filmmakers to explore and evaluate an overwhelming amount of design options. In this work, we focus on designing a system to facilitate this process.

2.2 Creativity Support Tool

HCI-oriented creativity research has been referred to as the third wave of modern creativity studies [15]. This wave saw the emergence of the concept of Creativity Support Tools (CSTs) [14] when Fischer [13] and Shneiderman [32] proposed that computer tools could advance human creativity. However, only a limited number of CST studies within this wave have delved into the specialized context of digital filmmaking.

In the context of machinima filmmaking, Davis et al. [9] featured the according creative process with a "Distributed Exploratory Visualization" model. This model depicts an iterative exploration, evaluation and refinement loop of vague mental design images across object, scene, and narrative layers. Importantly, this model suggests that the design of CSTs should concentrate on facilitating "immediate feedback" and helping boost this creative loop by "lowering the cost of exploration and increasing the fluidity of evaluation" across different layers. In light of these findings, Nicolas et al. [10] proposed a CST system that employs a rule-based approach to assess the "correctness" of each individual shot design outcome. However, this system has only focused on facilitating the evaluation of shot design for novices who lack fundamental cinematography knowledge. The demand from more experienced machinima filmmakers, especially those looking for immediate feedback on the creative exploration and evaluation of shots and shot sequences, has been somewhat overlooked previously. In this paper, we attempt to address this demand with our proposed system that suggests and visualizes high-quality, diverse shot designs for a given story scene in real time.

2.3 Intelligent Cinematography

The automation of cinematography tasks in both physical and virtual settings has been the focus of numerous autonomous or interactive cinematography tools. For physical scenes, numerous studies have proposed methods for suggesting coherent cinematic composition sequences based on provided video clips[4, 31, 34]. Additionally, many other studies [7, 17, 19, 22] utilized drones to capture physical scenes and provide real-time recommendations. However, these works are specially designed for physical scenes, rendering them not amenable to supporting machinima cinematography in real time within virtual environments, which is the focus of this paper.

For virtual scenes, many studies [11, 18, 20, 25] have been built upon classical cinematography rules to enable autonomous cinematography in the machinima environment. However, formalizing all possible rules into computational models can be challenging and there are also great variations on how the guidelines can be applied in practice. Consequently, such rule-based methods are prone to generating results with limited variety. Recent works started to investigate data-driven approaches, training machine learning models from data to predict cinematic compositions directly [11, 21]. Unfortunately, all of these methods do not allow users to specify their preferences and inject their thoughts into the results. Evin et al. [12] introduced Cine-AI, an interactive tool for cinematography that generates machinima shot sequences resembling those created by human directors. However, the tool's generation process relies solely on predefined rules and is agnostic to the input scene content, resulting in low-quality sequences that are often irrelevant and lack coherence.

To the best of our knowledge, how intelligent cinematography techniques can be adopted within the design of interactive systems to substantially support machinima filmmaking has not yet been investigated or assessed. In this work, we aim to formulate the design requirements of this system and adopt a machine learning-based technique to fulfil these requirements.

3 FIELD INTERVIEW AND DESIGN REQUIREMENTS FORMULATION

To better understand machinima filmmakers' design process and design thinking, we conducted field interviews with three professionals in their natural work setting, with access to their currently used software tools. The three interviewees — P1 (female, junior game designer), P2 (female, junior game designer), and P3 (male, lead game artist) — come from a world-leading digital game company. P1 and P2 have 3~5 years of professional narrative design experience using machinima tools for MMO game cutscene prototyping. P3 has more than 10 years of professional experience using machinima tool for not only cutscenes but also 3D film production. Each interview lasted for approximately 60 minutes and was conducted face-to-face. Our interview questions were structured around three topics with no specific limitations on the answers: current machinima filmmaking workflow, current cinematic composition design process, and difficulties encountered in the design process.

Based on the feedback we collected in this interview, we summarize our design requirements for building our CST as follows. First, the tool should significantly enhance the exploration of creative cinematic composition designs within the typical machinima filmmaking workflow (**R1**). Secondly, it should provide real-time recommendations for cinematic composition design, at both keyframe and scene levels [9] (**R2**). Thirdly, it should recommend a diverse range of plausible design alternatives that serve as inspirational examples, expanding upon conventional design paradigms rooted in cinematography rules (**R3**). Fourthly, the tool should recommend coherent cinematic compositions that seamlessly align with the specific scene context, minimizing the need for manual refinement. (**R4**). Finally, the cinematic composition designs recommended by the tool can be displayed in real-time to facilitate users' quick evaluation (**R5**).

4 SYSTEM DESIGN

Based on the design requirements distilled from the field interview, we have developed Cinemassist, a creativity support tool designed to enhance the exploration of creative cinematic composition designs within machinima environments. At the crux of our system is a Transformer-based generative model, which is capable of synthesizing plausible and diverse camera pose sequences in an autoregressive manner, conditioned on a sequence of keyframes and input semantics (including genre and intended emotion state). To train our model, we construct a dataset of 961 samples from 19 movies that are highly rated by IMDb and belong to three different genre categories: romance, action and thriller. We then estimate the intended emotion type of each sample from its accompanying dialogue subtitles using a text-to-emotion detection algorithm [2]. Powered by this model, Cinemassist implements an interactive paradigm that alternates between human decision and intelligent suggestion throughout the design process of cinematic compositions. In this section, we provide an overview of Cinemassist.

The Cinemassist interface is shown in Figure 2, comprising three components: (a) a control panel for configuring the input 3D animation and high-level semantics; (b) a design panel for designing cinematic compositions at frame level; (c) a storyboard panel for visualizing recommended composition sequences at scene level. Furthermore, the interface provides a scene view to display the 3D animation in real time, along with suggested camera poses in 3D space, a cinema view to preview the composition from a particular camera pose, as well as an animation timeline to facilitate keyframe selection. In the remainder of this section, we elaborate on each of the three components.

Input configuration. A user can load a 3D animation into the system. The animation will be displayed in the scene view, under which the timeline is presented. Then, on the control panel (Figure 2(a)), the user can select two scene objects as the characters of interest (Cinemassist assumes a scene includes at least two characters), and a camera object in the scene as the default control camera (**R1**). Additionally, the user can choose one of the movie genre categories (including "action", "romance" and "thriller") that the story of the animation is expected to belong to, and one intended emotion state that the user intends to express through cinematic compositions to the audience. We consider five common intended emotion states including "happy", "angry", "surprise", "sad" and "fear" according to Plutchik's wheel of emotions [30].

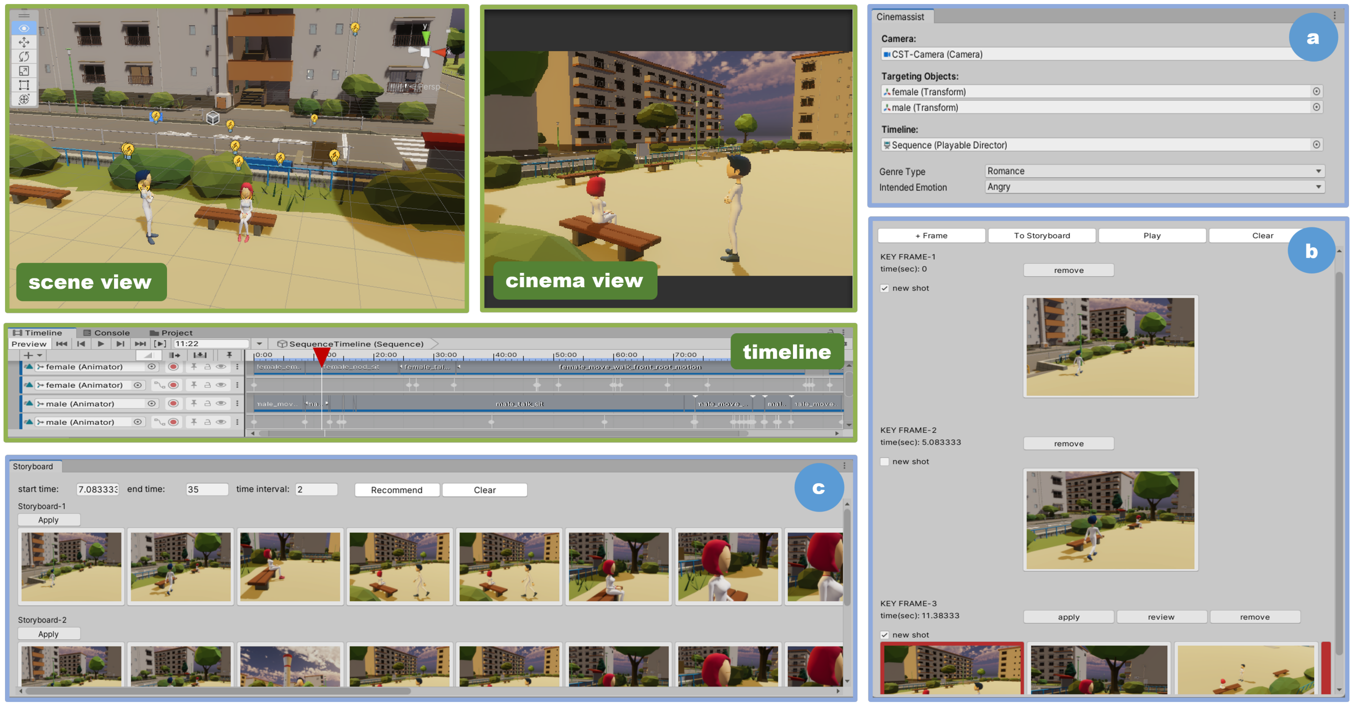

Figure 2: Cinemassist interface.

Frame-level design exploration. To start the design of cinematic compositions for the given animation, the user can drag the red arrow along the timeline to preview the entire animation, and then move the arrow to a time point that corresponds to the first keyframe. To design the cinematic composition for the current keyframe, the user clicks the "+Frame" button on the frame-level suggestion panel (Figure 2(b)) to insert a new keyframe. Then, our system generates a set of camera poses for it, which are presented to the user as suggestions (**R2, R3**). The suggested camera poses will be visualized as light bulbs in the scene view to allow for *in-situ* exploration in 3D space. Whenever the user clicks an option in the scene view, its corresponding 2D composition will be displayed in the cinema view for the user to evaluate in real time (**R5**). Alternatively, the user can also quickly explore and evaluate all the suggestions by clicking the "review" button. This will display the 2D compositions of all the suggested camera poses in a row, ranked by their quality scores predicted by our model (**R3, R5**). Once a desirable option is found, the user can confirm it by clicking the "apply" button, which will add the chosen option to the ongoing sequence of compositions. Then, the user moves on to identify the next keyframe, on which the aforementioned system suggestion and user decision procedures are repeated. For each keyframe, our system can additionally suggest if it is a shot boundary after which a new shot should be started, to facilitate better camera planning in practice. In this way, the user and our system work together to iteratively develop a camera trajectory through the 3D scene, presenting a way to portray the story. Notably, at the end of each iteration, the user-chosen option is fed into our model so that our model makes per-iteration predictions based on both the current keyframe content and the previously designed compositions, resulting in relevant and coherent composition suggestions. At last, the user can export the resulting composition sequence and review the animation rendered under it in the cinema view by clicking the "play" button, to get a quick sense of its overall quality (**R5**).

Scene-level design exploration. In addition to the frame-level design exploration above, our system also provides a functionality to allow the user to explore across sequences of cinematic compositions at scene level. To do this, on the storyboard panel, the user first chooses a range on the timeline by specifying the start and end time points, and setting a time interval. Then, after the "recommend" button is clicked, our system automatically samples a sequence of uniformly spaced keyframes using the interval over the specified time range and suggests multiple storyboards (Figure 2 (c)) (**R2, R3, R4**). Each storyboard is a sequence of cinematic compositions, one for each keyframe. Note that the chosen range can vary to cover the entire animation or only a small part of it. The user can select a satisfying sequence to export into the design panel and continue iterating on it via the frame-level design exploration. Further, the user can also export a cinematic composition sequence in progress from the design panel into the storyboard panel, and our system can automatically extend the partial sequence into multiple complete sequences by generating the remaining compositions (**R1**).

5 PRELIMINARY USER STUDY

We conducted a small-scale preliminary user study to evaluate our current prototype of Cinemassist. The main goal of this evaluation was to assess the feasibility of Cinemassist and confirm our key design decisions by gathering some preliminary feedback from end users. This study invited 4 participants divided into two groups: professional machinima filmmakers, and trained media practitioners with filmmaking or video editing experience. We regard the first group as the expert machinima filmmakers group and the second as the novice group. In the following, we detail the procedure of our study and then report the results. Based on the results, we discuss the design implications we drew in our study.

Participants. For the expert group, we invited back two professional machinima filmmakers (P1 and P2) who participated in our field interviews (section 3). Their previous feedback helped us determine our current design decisions and computational methodology for building up Cinemassist. For the novice group, we invited two media school PhD students, P4 (male, aged between 30 and 40) and P5 (female, aged between 20 and 30). According to a pre-study questionnaire, they all had experience using machinima tools in their studies and work.

Procedures. After a short tutorial to introduce Cinemassist's interface and key functionalities within Unity Timeline, the participants were asked to complete a cinematic composition design task for a representative 3D animation on Unity Timeline and output their design outcomes, using a virtual scene in Unity constructed with open-source assets. The 3D animation consisted of three phases following the standard "Hero's Journey" story pattern [12], which is widely adopted in digital game design and filmmaking. Given the formative nature of the evaluation, no specific constraints were placed on the use of Cinemassist or completion time. Participants were encouraged to use a think-aloud method to provide insights into the usefulness of Cinemassist. With participants' consent, their whole design process was recorded for further analysis. After each participant finished the design task, we conducted a debrief interview to gather the participant's feedback on Cinemassist. The interview was semi-structured and lasted for approximately 30 minutes. Notably, The interview questions were structured based on our observations of the participants' design process and the following four key issues: the influence of the use of Cinemassist on the design process of cinematic composition; major flaws in the current design of Cinemassist and the according suggestions for refinement.

Feedback. Regarding the first key issue, all participants confirmed that the functionalities of Cinemassist facilitated their design process differently. Notably, regarding the frame-level design exploration function, P2 commented: "*The inspirational bulbs popped out in the scene view at each keyframe enabled me to start my design exploration directly from multiple feasible setting-ups*". Regarding the scene-level design exploration function, all participants felt that the composition sequences generated by our generative model were almost feasible and some of them even looked novel. P1 commented: "*I never know the whole story can be told in this way. I already tend to settle on a recommended sequence or only do some minor adjustments for one or several keyframes within the sequence.*". P4 complimented: "*Previously, adjusting camera placement to track dynamic objects involved cumbersome mouse and keyboard controls. The generated sequences now handle this seamlessly.*". However, P4 also pointed out: "*I prefer to implement the cinematic work myself. I only refer to the recommended sequences for overview purpose.*". The above feedback coincides with our observations of the two participants' design process. Particularly, Figure 3 compares the design processes of P1 and P4. As is shown, P1 frequently used our scene-level design exploration function and applied the recommended composition sequences (highlighted by green outlines) to her design at different time ranges along the animation's timeline. In contrast, P4 more often used the frame-level design exploration function to achieve his intended composition design for several keyframes on the timeline (highlighted by the blue outline). On the second issue, both P1 and P4 expressed concerns about the effectiveness and efficiency of the genre and intended emotion type settings on the customization panel. P1 pointed out, "*I am uncertain about which semantic types can produce specific effects. I tend to adopt a design if it appeals to me visually, regardless of its semantic classification.*" P4 added, "*I don't see a strong correlation between the specified semantic types and the style of the recommended results. In practice, the same composition may indeed serve multiple semantic types.*" Moreover, there were shared concerns about the "storyboards panel", where all participants noted that it presented an overwhelming amount of information to review simultaneously. This issue became especially pronounced when recommended sequences grew longer, making it challenging for participants to efficiently assess the recommendations and locate specific keyframes. Regarding this concern, P4 contributed an enhancement idea: "*Could the next version of Cinemassist simplify the reviewing of recommended composition sequences by just animating them in multiple cinema views?*"

6 DISCUSSION AND FUTURE WORK

In this work, we interviewed three professionals and established five key design requirements for a software tool to support creative machinima filmmaking. To fulfil these design requirements, our system employs a deep generative network model trained on a diverse film clip dataset, capable of autonomously generating a wide range of cinematic compositions at both frame and scene levels. This model surpasses the limitations of existing data-driven approaches [11, 21] that restrict camera poses or rely on reference clips. In addition, unlike the rule-based systems [11, 18, 20, 25], our model, informed by various content, genres, directors, and eras, expands upon canonical cinematography principles, enabling the generation of camera poses from scratch. Participants in our preliminary user study noted that Cinemassist can recommend coherent and contextually appropriate sequences with minimal manual adjustments. In contrast, systems like CineAI [12] struggle with diversity, relevance and continuity, often requiring further manual intervention to achieve desirable results, which constrains the creative process and increases the cost and effort in evaluating potential designs at the frame and scene levels.

In our preliminary user study, we also discerned two distinct types of user needs related to the functionality of our system. The first type prefers to maintain control over the design process, preserving the human designer's initiative throughout the design process. Conversely, the second type leans towards delegating a significant portion of the design tasks to our system. As such, we emphasize the importance of considering "levels of control" in our next design iterations. The study also revealed that participants tend to assess the recommended designs primarily based on their

Figure 3: The cinematic composition sequences designed by two participants in our study: P1 (top) and P4 (bottom). Each composition is created manually (red), with Cinemassist's frame-level exploration (green) or with Cinemassist's scene-level exploration (blue).

own professional judgements, rather than considering the recommendations' "genre" and "intended emotion" classifications. This phenomenon could infer several intriguing research questions to be addressed in our future and other relevant CST studies of digital filmmaking. Firstly, could the classifications of "genre type" and "intended emotion type" be sufficient to fully encapsulate the user's desired animation semantics? Secondly, how to evaluate whether the specified semantic prompts can lead to the recommendations of the users' intended cinematic styles? Thirdly, will fixed semantic prompts confine the design space of the CST when providing inspirational recommendations whose appropriateness and novelty are still influenced by other factors in the virtual scene? In a broader sense of future CST designs, how to select appropriate semantic prompts or weigh their influence on the divergence of the inspirational recommendations besides facilitating the users to achieve their intended convergent result?

Notably, we only evaluated Cinemssist by a small-scale formative user study. Although the preliminary feedback we received from the 4 participants in the study was quite positive, the usefulness of Cinemassist in facilitating the creative process of cinematic composition design is still subject to more rigorous and larger-scale summative user studies. Additionally, our current system is designed for a specific machinima filmmaking context. Future iterations will expand the system's capabilities to support a wider range of 3D filmmaking workflows where scene animation is created concurrently with, or after, the development of storyboards.

REFERENCES

[1] H Porter Abbott. 2002. *The Cambridge introduction to narrative.* Cambridge University Press.
[2] Shivam Sharma Karan Bilakhiya Aman Gupta, Amey Band. 2020. *text2emotion.* https://pypi.org/project/text2emotion/
[3] Hideaki Anno. 2021. Evangelion: 3.0+1.01 Thrice Upon a Time [Film]. *IMDb* (2021).
[4] Ido Arev, Hyun Soo Park, Yaser Sheikh, Jessica Hodgins, and Ariel Shamir. 2014. Automatic Editing of Footage from Multiple Social Cameras. *ACM Trans. Graph.* 33, 4, Article 81 (jul 2014), 11 pages. https://doi.org/10.1145/2601097.2601198
[5] Daniel Arijon. 1991. *Grammar of the film language.* Silman-James Press.
[6] J. Aumont, University of Texas Press, A. Bergala, M. Marie, R. Neupert, and M. Vernet. 1992. *Aesthetics of Film.* University of Texas Press. https://books.google.com.hk/books?id=nmpZAAAAMAAJ
[7] Rogerio Bonatti, Arthur Bucker, Sebastian Scherer, Mustafa Mukadam, and Jessica Hodgins. 2021. Batteries, camera, action! Learning a semantic control space for expressive robot cinematography. In *2021 IEEE International Conference on Robotics and Automation (ICRA).* 7302–7308. https://doi.org/10.1109/ICRA48506.2021.9560745
[8] Christopher Bowen. 2013. *Grammar of the Shot.* Routledge.

[9] Nicholas Davis, Boyang Li, Brian O'Neill, Mark Riedl, and Michael Nitsche. 2011. Distributed Creative Cognition in Digital Filmmaking. In *Proceedings of the 8th ACM Conference on Creativity and Cognition* (Atlanta, Georgia, USA) *(C&C '11)*. Association for Computing Machinery, New York, NY, USA, 207–216. https://doi.org/10.1145/2069618.2069654

[10] Nicholas Davis, Alexander Zook, Brian O'Neill, Brandon Headrick, Mark Riedl, Ashton Grosz, and Michael Nitsche. 2013. Creativity support for novice digital filmmaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 651–660.

[11] Edirlei E. S. de Lima, Cesar T. Pozzer, Marcos C. d'Ornellas, Angelo E. M. Ciarlini, Bruno Feijó, and Antonio L. Furtado. 2009. Virtual Cinematography Director for Interactive Storytelling. In *Proceedings of the International Conference on Advances in Computer Entertainment Technology* (Athens, Greece) *(ACE '09)*. Association for Computing Machinery, New York, NY, USA, 263–270. https://doi.org/10.1145/1690388.1690432

[12] Inan Evin, Perttu Hämäläinen, and Christian Guckelsberger. 2022. Cine-AI: Generating Video Game Cutscenes in the Style of Human Directors. *Proceedings of the ACM on Human-Computer Interaction* 6, CHI PLAY (2022), 1–23.

[13] Gerhard Fischer. 2004. Social creativity: turning barriers into opportunities for collaborative design. In *Proceedings of the eighth conference on Participatory design: Artful integration: interweaving media, materials and practices-Volume 1*. 152–161.

[14] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the landscape of creativity support tools in HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–18.

[15] Jonas Frich, Michael Mose Biskjaer, and Peter Dalsgaard. 2018. Twenty years of creativity research in human-computer interaction: Current state and future directions. In *Proceedings of the 2018 Designing Interactive Systems Conference*. 1235–1257.

[16] Francis Glebas. 2012. *Directing the story: professional storytelling and storyboarding techniques for live action and animation*. Routledge.

[17] Mirko Gschwindt, Efe Camci, Rogerio Bonatti, Wenshan Wang, Erdal Kayacan, and Sebastian Scherer. 2019. Can a Robot Become a Movie Director? Learning Artistic Principles for Aerial Cinematography. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 1107–1114. https://doi.org/10.1109/IROS40897.2019.8967592

[18] Li-wei He, Michael F Cohen, and David H Salesin. 1996. The virtual cinematographer: A paradigm for automatic real-time camera control and directing. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. 217–224.

[19] Chong Huang, Fei Gao, Jie Pan, Zhenyu Yang, Weihao Qiu, Peng Chen, Xin Yang, Shaojie Shen, and Kwang-Ting Cheng. 2018. ACT: An Autonomous Drone Cinematography System for Action Scenes. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. 7039–7046. https://doi.org/10.1109/ICRA.2018.8460703

[20] Arnav Jhala and R Michael Young. 2011. Intelligent machinima generation for visual storytelling. In *Artificial Intelligence for Computer Games*. Springer, 151–170.

[21] Hongda Jiang, Bin Wang, Xi Wang, Marc Christie, and Baoquan Chen. 2020. Example-driven virtual cinematography by learning camera behaviors. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 45–1.

[22] Niels Joubert, Jane L. E, Dan B Goldman, Floraine Berthouzoz, Mike Roberts, James A. Landay, and Pat Hanrahan. 2016. Towards a Drone Cinematographer: Guiding Quadrotor Cameras using Visual Composition Principles. arXiv:1610.01691 [cs.GR]

[23] Aki Kubota. 2021. Hideaki Anno: The Final Challenge of Evangelion [Film]. *NHK* (2021).

[24] Nicole E Lemon. 2012. *Previsualization in Computer Animated Filmmaking*. Ph. D. Dissertation. The Ohio State University.

[25] Christophe Lino, Marc Christie, Fabrice Lamarche, Schofield Guy, and Patrick Olivier. 2010. A real-time cinematography system for interactive 3d environments. In *SCA'10 Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. 139–148.

[26] Henry Lowood and Michael Nitsche. 2011. *The machinima reader*. MIT Press.

[27] Paul Marino. 2004. *3D game-based filmmaking: The art of machinima*. Paraglyph Press.

[28] Marcos Mateu-Mestre and Jeffrey Katzenberg. 2010. *Framed ink: Drawing and composition for visual storytellers*. Design Studio Press.

[29] R. McKee. 1997. *Story: Substance, Structure, Style and the Principles of Screenwriting*. HarperCollins Publishers.

[30] Manshad Abbasi Mohsin and Anatoly Beltiukov. 2019. Summarizing emotions from text using Plutchik's wheel of emotions. In *7th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS 2019)*. Atlantis Press, 291–294.

[31] K. L. Bhanu Moorthy, Moneish Kumar, Ramanathan Subramanian, and Vineet Gandhi. 2020. GAZED– Gaze-Guided Cinematic Editing of Wide-Angle Monocular Video Recordings. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3313831.3376544

[32] Ben Shneiderman. 2002. Creativity support tools. *Commun. ACM* 45, 10 (2002), 116–120.

[33] T. Sobchack and V.C. Sobchack. 1987. *An Introduction to Film*. Little, Brown. https://books.google.com.hk/books?id=SwMbAQAAIAAJ

[34] Anh Truong, Floraine Berthouzoz, Wilmot Li, and Maneesh Agrawala. 2016. QuickCut: An Interactive Tool for Editing Narrated Video. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 497–507. https://doi.org/10.1145/2984511.2984569