

A Study on the Controllability of Character Actions in Generative AI Animation: A Case Study of Dong Yong Burying His Father

JIALE YANG*, Design graduate school, Dongseo University
Busan, Korea (south), 13201770275@163.com

KIM Hae Yoon*, Design graduate school, Dongseo University
Busan, Korea (south), haeyoonkim@empal.com

* Corresponding author

Abstract

With the increasing application of generative artificial intelligence technologies in animation production, the efficiency of character action generation has improved, yet problems remain in the accuracy and stability of action execution. Taking the AI-generated animation Dong Yong Burying His Father, created using Jimeng AI, as a practical case, this study analyzes issues including unstable posture transitions, discontinuities in complex sequential actions, incorrect character orientation, and inaccurate understanding of traditional etiquette movements. Based on fixed character three-view reference sheets, background scenes, and character-scene images, two prompt-based action control methods were tested: precise and concise prompts, and time-segmented prompts. The practical results show that for simple actions such as turning the head, raising the head, and waving a hand, precise and concise prompts are more effective in clarifying character orientation and action goals. For continuous and complex actions such as kneeling, falling, and standing up, dividing the action according to specific time points can improve the stability of action sequence and transitions. This study argues that improving the controllability of character actions in generative AI not only depends on the generative model itself, but also requires the integration of preliminary visual settings and well-designed prompts, so that the generated results can better conform to storyboard requirements and the creator's intended actions.

I. Research Background and Problem Statement

With the rapid development of generative artificial intelligence technologies, the application of AI has expanded from static image generation to dynamic video generation. Generative technologies represented by text-to-video and image-to-video models enable creators to rapidly produce continuous dynamic content through text prompts, character reference images, keyframes, and other forms of input. OpenAI noted in its technical report on Sora that large-scale video generation models are already capable of generating continuous videos with relatively high visual quality from text prompts, indicating that generative AI is gradually developing the ability to simulate real-world space, character movement, and temporal changes. For animation creators, the emergence of such technologies means that animation production no longer relies entirely on the complex workflows of traditional frame-by-frame drawing and 3D animation, while also providing new creative approaches for independent animation and small-scale animation projects. Character movement is an important component of character development and narrative expression in animation. Canadian animator Norman McLaren once described animation as “drawn movement,” a view that emphasizes the central role of movement in animation. Therefore, regardless of how animation production technologies evolve, whether character actions can accurately convey a character's intentions and emotions remains an important issue in animation creation. However, although current generative AI technologies can rapidly generate character motion, their ability to control such actions still has certain limitations. Fang et al., in their study MotionCharacter, pointed out that existing text-to-video models still face difficulties in fine-grained character action control, while more specific action descriptions can help improve the controllability of character motion. This issue is particularly evident in practical AI animation production. For example, when creators require a character to perform actions such as “turning around,” “kneeling down,” “standing up,” “running,” “performing a traditional bow,” or “helping another person stand up,” the generated results may exhibit incomplete actions, incorrect movement directions, abnormal limb structures, abrupt transitions between actions, or unintended changes in the character's position. Based on these issues, this study takes the production of the generative AI animation Dong Yong Burying His Father using the generative AI software Jimeng AI as a practical case study. Actions such as kneeling, running, performing a traditional bow, and helping another person stand up are selected to analyze the controllability of character actions generated by generative AI. By comparing the generation results produced under different prompt descriptions and keyframe control methods, this study explores the major factors affecting the accuracy and stability of character action generation and, on the basis of practical production experience, summarizes specific methods for improving the controllability of character actions in AI-generated animation.

II. Major Problems in Character Action Generation in Dong Yong Burying His Father

During the production of the AI-generated animation Dong Yong Burying His Father, character actions directly affect character performance and visual storytelling. In practical use of Jimeng AI for motion generation, simple actions are relatively stable. However, when the action process becomes more complex, involves interactions between multiple characters, or includes traditional etiquette movements, the generated results are more likely to deviate from the creator's intended design.

A. Insufficient Stability in Posture Transitions

For example, in the scene where “Dong Yong kneels down to plead with the physician,” the character needs to gradually transition from a standing position by bending the knees, lowering the body, and finally kneeling on both knees. However, during actual production, problems such as missing intermediate movements, abrupt posture transitions, and abnormal leg structures occurred, indicating that AI still lacks stability in controlling changes in the body's center of gravity and the

B. Complex Continuous Actions Are Prone to Discontinuities

In continuous action sequences such as “running—falling—standing up—continuing to run,” entering multiple complex action instructions at once can easily lead to omitted actions, disordered sequences, or abrupt transitions. The more action steps involved, the greater the uncertainty during the generation process.

C. Character Orientation and Movement Direction Are Prone to Deviations

In scenes where characters turn around, walk, interact with other characters, or engage in conversation, directional errors may still occur even when the prompts clearly specify the character's orientation. For example, when Dong Yong is required to face the physician while speaking, the character may turn toward the camera or face the opposite direction during the action, and vertical positional misalignment may also occur. When the character is instructed to walk toward the left side of the frame, the generated result may also show the character moving to the left. Errors in character orientation directly affect the characters' eyeline relationships and the spatial continuity between shots.

D. Misinterpretation of Traditional Etiquette Actions

Traditional etiquette actions such as kneeling in respect, performing a traditional bow with clasped hands, and making a cupped-hand salute appear in the animation Dong Yong Burying His Father. AI sometimes generates the action of “zuoyi” as an ordinary bow, or represents the “gongshou” salute as pressing the palms together, indicating that its understanding of traditional cultural movements still contains inaccuracies. Therefore, the controllability of character actions involves not only technical accuracy, but also whether traditional cultural content can be represented correctly.

III. Experimental Analysis of Character Action Controllability

Before dynamic generation, the main characters and scenes in Dong Yong Burying His Father were first standardized through a unified design process. For the characters, front, side, and back three-view reference sheets were used to define their facial features, hairstyles, costumes, and body proportions; for the scenes, fixed backgrounds such as the thatched cottage, medical hall, and street were generated in advance. The characters were then placed into the corresponding scenes according to the specific storyboard shots, forming static scene images that met the requirements for shot composition and character positioning. These images served not only as the starting images for animation generation, but also as an important basis for maintaining consistency in both characters and scenes. In this way, the experiment could focus on “how the characters move,” while avoiding the simultaneous regeneration of character designs, background structures, and character positions during dynamic generation. To address the problems encountered during the production of Dong Yong Burying His Father, including unstable posture transitions, discontinuities in continuous actions, incorrect character orientation, and misinterpretation of traditional etiquette movements, this study tested two different prompt-control methods in Jimeng AI. The first method emphasized accurate and concise instructions and was mainly applied to shots involving relatively simple actions and clearly defined directions. The second method divided actions according to temporal sequence, providing more detailed specifications for the character's behavior, orientation, and state during different time intervals.

A. Control Method Using Precise and Concise Prompts

For single and relatively clear actions such as turning around, lowering the head, kneeling down, facing a specific character, and performing a traditional bow, prompts should not include too much irrelevant wording. Instead, they should directly specify the character, the action, the direction, and the final posture. For example, in the shot where Dong Yong talks with the physician, adding too many action descriptions often leads to problems such as character misalignment and discontinuous movement. In such cases, the issue can be addressed by using a starting frame together with a concise prompt, such as: “Use the scene image as a reference for character positioning. Fixed camera. The physician waves his hand and then says: ‘It is not that I refuse to save him, but this illness is beyond the reach of medicine. You... should prepare for the funeral.’ No background sound except the dialogue. Character movements should be continuous.” For actions such as zuoyi that are difficult for the model to recognize, a corresponding action image can be provided as a motion reference. Similarly, in the shot of “Dong Yong kneeling down,” simplifying the prompt and making it explicit as “Dong Yong faces the physician, the camera moves closer, he kneels on both knees, continuously kowtows, and says: ‘Please, I beg you, save him.’ He is anxious, crying, and in tears.” is more effective in highlighting the core action than adding excessive environmental and emotional descriptions. This method is suitable for addressing errors in character orientation as well as some simple posture-transition problems. Its key lies not in increasing the length of the prompt, but in reducing vague expressions so that the AI can quickly identify the main action requirements.



Figure 1. Image of Dong Yong Talking with the Doctor



Figure 2. Shot of Dong Yong Kneeling Down

B. Control Method Using Time-Segmented Prompts

When a shot contains multiple continuous actions, using only brief instructions is often insufficient to ensure the correct action sequence. Therefore, it is necessary to divide the actions according to time segments. For example: “Seconds 1–2: Dong Yong, while kneeling, turns around. Seconds 3–6: Dong Yong quickly crawls toward the rich man, grabs hold of him, and says, ‘Master, I can do anything.’ Seconds 7–10: Continuing from the 6-second mark, Dong Yong then raises his head to look at the rich man and says, ‘I only ask that you advance me some silver so I may bury my father. I am willing to serve you as a slave for the rest of my life.’ Throughout the scene, Dong Yong remains anxious and crying, speaks at a relatively fast pace, and there is no background sound other than the dialogue. The camera follows Dong Yong's movement.” In this way, a complex action can be broken down into several continuous smaller actions, and the character's bodily state in different time segments can be specified separately, allowing the action to develop in the intended order. This method is essentially equivalent to breaking a complete animated performance into multiple action nodes, enabling the AI to understand more clearly the sequence in which the actions occur.



Figure 3. Image of the Camera Movement as Dong Yong Quickly Crawls Forward While Kneeling



Figure 4. Image of the Camera Movement as Dong Yong Quickly Crawls Forward While Kneeling

C. Selection Between the Two Different Methods

Based on the actual generation results, the two prompt methods should be selected according to the action characteristics of different shots. For simple action shots such as walking, turning the head, raising the head, and waving a hand, using a starting frame together with accurate and concise prompts can more directly constrain the direction of movement and the character's final state, while reducing interference from irrelevant information. For long shots containing continuous actions, however, the time-segmented method is more suitable. By dividing a complete action into several consecutive stages, the character can perform the movements according to the intended sequence, thereby reducing omitted actions and unnatural transitions.

IV. Conclusion

Through the practical production of Dong Yong Burying His Father, it can be observed that although generative AI has already demonstrated strong capabilities in dynamic video generation, a certain degree of instability remains in the control of character actions. This is particularly evident in posture transitions, continuous actions, character orientation, and traditional etiquette movements such as zuoyi and kneeling in respect, where the generated results can easily deviate from the creator's original action design. Therefore, accurately conveying the creator's intended actions to AI is one of the key issues that currently needs to be addressed in AI animation production. Based on fixed character three-view reference sheets, background scenes, and character-scene images, this study conducted a practical analysis of two prompt-control methods in Jimeng AI. The results show that for relatively simple actions such as walking, turning the head, raising the head, and waving a hand, accurate and concise prompts can reduce interference from irrelevant information and enable the AI to focus more effectively on executing the required character actions and directions. For traditional actions such as zuoyi, corresponding action images can be selected as motion references. In contrast, for long shots involving multiple action processes, dividing the complete action according to different time points can define the action sequence more clearly and reduce problems such as omitted actions, abrupt transitions, and incorrect movement directions.

V. Acknowledgements

This research was supported by BK21 Four Service Design driven Social Innovation Educational Research Team in Dongseo University.

- [1] T. Brooks, B. Peebles, C. Holmes, et al., “Video generation models as world simulators,” OpenAI, Feb. 2024. [Online].
- [2] H. Fang, D. Qiu, B. Mao, and H. Tang, “MotionCharacter: Fine-Grained Motion Controllable Human Video Generation,” Proc. AAAI Conf. Artif. Intell., vol. 40, no. 5, pp. 3831–3839, 2026.
- [3] J. Ho, T. Salimans, A. Gritsenko, W. Chan, M. Norouzi, and D. J. Fleet, “Video Diffusion Models,” in Advances in Neural Information Processing Systems (NeurIPS), vol. 35, 2022, pp. 8633–8646.
- [4] Y. Guo, C. Yang, A. Rao, et al., “AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Tuning,” in Proc. Int. Conf. Learning Representations (ICLR), 2024.
- [5] F. Thomas and O. Johnston, Disney Animation: The Illusion of Life, 1st ed. New York: Abbeville Press, 1981.
- [6] A. Thain, “In the Blink of an Eye: Norman McLaren Between Dance and Animation,” in The Oxford Handbook of Screendance Studies, D. Rosenberg, Ed. Oxford University Press, 2016, pp. 167–186.