Animation and Human-Computer Interaction: A Systematic Review of Applications, Challenges, and Research Opportunities

Yiyang Liu[1]

Received July 4, 2025
Accepted July 10, 2026
Electronic access August 15, 2026

Animation and human-computer interaction (HCI) are being used in many interactive technologies, such as VR/AR, gesture interaction, etc. However, there is still a lack of systematic synthesis on their mechanisms, impacts and challenges. To fill this gap, this paper conducts a systematic literature review. In terms of HCI, our review shows that animation improves HCI intuitiveness, useful feedback and waiting time perception. While, from animation perspective, HCI technologies change the workflow of animation production by lowering the barrier, improving interaction method, and making animation storytelling more immersive. There are still many challenges existing in this field, including lack of standardization in principles, technical limitations (e.g., latency, accuracy), and high creation barrier. This paper reviews above challenges and calls for future work to standardize evaluation heuristics, solve technical bottlenecks, and develop more available creator-centered tools.

Keywords: Animation · Human-Computer Interaction · User Experience · Systematic Review

Introduction

As a creative field, animation has undergone remarkable changes from the traditional frame-by-frame production. Now it can emphasize interactivity, realism and other elements, and create an immersive world for the audience or users. Nowadays, animation is no longer limited to the concept of passive viewing the screen, but actively attracts users to 3D space. At the same time, the field of human-computer interaction (HCI) has developed out of the category of original mouse and keyboard interface design, including enhanced sensors including computer imaging and artificial intelligence, and has developed into the so-called "new era of hybrid media", covering gestures, limb posture, eye tracking, voice commands and touch interface. Recently, these two fields are also increasingly integrated. Researchers and developers have opened up a new world for the application of human-computer interaction technology by proposing real-time 3D motion capture to generate and control 3D characters, use voice recognition to drive lip movements and emotional expression, or produce animated stories with the help of immersive VR/AR. The combination of animation and human-computer interaction not only stays at the conceptual level, but also seems to enjoy the marketing dividend. This integration is now more like a business phenomenon. Analysts and investors predict that by 2025, the global market value of applications involving VR/AR and animation, digital narrative and other functions will reach about 28 billion to 35 billion US dollars, which fully proves its huge commercial and technological potential.

The combination of human-computer interaction and animation is very important, because it makes up for the inherent shortcomings in their respective fields and realizes the cultural values of both sides. Classic animation has the advantages of visual narrative, strong expressiveness and conveying complex content in an easy-to-understand way. However, its medium form is less flexible in terms of user participation, which may cause users to only act as bystanders. Traditional human-computer interaction provides rich natural input possibilities and system feedback, leaving room for interaction and higher autonomy. Both of these media can use animation in a more dynamic way than ever before, so that the characters and stories in the animation can respond to user interaction. Users do not need the help of controllers, but explore the virtual world through natural actions, and the computer can realize branch narrative by recognizing the user's choices and voice commands. This integration brings a more intelligent and personalized experience. For example, users can view the body type by scanning their own bodies, or use gesture controls to create personalized animation sequences. Examples of the integration of cutting-edge technology include the launch of interactive VR animated films that can influence the direction of the plot, the development of AR mobile phone applications for cartoon characters to "break into" the living room, and the system that can now capture the movements of live performers and directly make virtual characters make corresponding actions – performers have thus become one of the creative links.

[1] *Shanghai Qingpu World Foreign Language School, China*

© The National High School Journal of Science 2026

However, despite its rapidly expanding potential and remarkable breadth, the integration of animation and human-computer interaction at the interdisciplinary level has been proven to be challenging in academic research. As far as we know, this literature review is the first strategic review focusing on the integration of animation and human-computer interaction. At the same time, there are many differences in the existing research background. Relevant research is scattered in the field of personalization, such as animation creation based on gesture design, voice-driven character animation, immersive narrative design based on VR, and even artificial intelligence-driven animation production. The theoretical heterogeneity of this method makes it difficult for newcomers to this field to understand its real development, major achievements and problems that need to be solved in the end. This review integrates the system and deeply analyzes the existing research results, and presents the relevant application scenarios, core advantages and existing limitations of interactive animation in a structured way. At present, although the integration of advanced human-computer interaction tools and animation production has brought innovation to creation, it has also raised the threshold of industry access due to technological complexity. Based on this, this article will sort out the frontier of field research, accurately identify core technology bottlenecks, and focus on emerging solutions such as artificial intelligence driving tools and cost-effective motion capture devices. This provides comprehensive information support and capability empowerment for animation creators, and helps to reduce the creative cost and technical threshold of the next generation of interactive animation experience.

Method

This study adopts a systematic literature review to examine how animation and human-computer interaction (HCI) are integrated in existing research. The aim is to identify the major application scenarios of this interdisciplinary field, summarize its main benefits, clarify current challenges, and highlight future research opportunities. To ensure transparency and reproducibility, the review followed a structured process including literature retrieval, screening, data extraction, and thematic synthesis.

Literature Search and Selection Criteria

The literature search was conducted in major academic databases relevant to computer science, digital media, and design, including Scopus, Web of Science, IEEE Xplore, ACM Digital Library, and Google Scholar. To capture recent progress in animation generation, immersive media, and interactive systems, the search was restricted to English-language studies published within the selected time span. The search terms combined keywords from both animation and HCI, such as animation, interactive animation, animation generation, human-computer interaction, gesture interaction, voice interaction, VR, AR, motion capture, and immersive storytelling. In addition to database retrieval, backward and forward snowballing was used to identify additional relevant studies from the references and citations of key papers.

To maintain consistency during screening, explicit inclusion and exclusion criteria were defined. Studies were included if they focused on animation or animated content as a central topic and incorporated a clear HCI component, such as gesture, voice, touch, motion capture, VR/AR, embodied interaction, or creator-centered interactive tools. Only studies with substantial methodological, technical, or analytical content were retained. In contrast, studies were excluded if they treated animation only as passive visual media without meaningful interaction, discussed HCI without direct relevance to animation, lacked sufficient methodological detail, duplicated another study, or did not provide accessible full text.

Screening and Data Extraction

After the initial search, duplicate records were removed. The remaining studies were first screened by title and abstract to eliminate clearly irrelevant papers. Full-text reading was then conducted on the candidate studies to determine final eligibility. Through this multi-stage filtering process, the final review corpus was established.

For each included study, a structured extraction form was used to record key information, including publication year, research objective, application scenario, animation type, HCI technology, methodology, sample or system setting, major findings, and reported limitations. This procedure ensured that the subsequent analysis was based on comparable dimensions across studies rather than unsystematic description.

Thematic Analysis

To synthesize the selected literature, this study employed thematic analysis. Since the reviewed work spans multiple areas, including animation generation, interactive systems, immersive experiences, and creator-support tools, thematic analysis provides an appropriate way to identify shared patterns across heterogeneous studies.

The analysis began with open coding, in which each study was examined to identify its main concerns, such as line-art colorization, customized character generation, image-to-video synthesis, 4D animation, gesture-based interaction, voice-driven animation, immersive storytelling, workflow transformation, latency, consistency, and evaluation issues. These initial codes were then grouped into broader conceptual categories and further refined into several higher-level themes.

© The National High School Journal of Science 2026

The final themes were developed around four major aspects: animation generation paradigms, HCI-enabled interaction and creation, the benefits brought to user experience and production workflow, and the major challenges and future research opportunities. These themes were derived not only from the objectives of this review but also from repeated patterns emerging across the selected studies. In this way, the review moves beyond a simple narrative summary and provides a systematic synthesis of how animation and HCI are currently integrated, what progress has been achieved, and which issues still require further investigation.

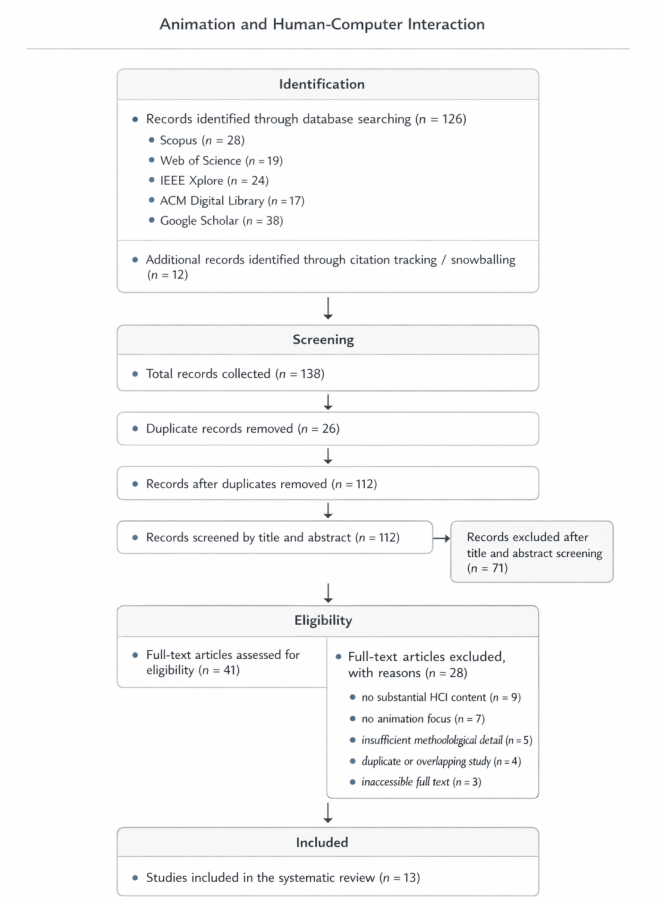

Figure. 1 Flowchart of the systematic literature search, screening, and study selection process.

Literature Review

Anime Diffusion Colorization

The Anime Diffusion Colorization task aims to use computer vision and deep learning techniques to automatically convert an uncolored animated character line drawing into a color image with professional aesthetic quality. In most cases, the input may be a face close-up or a full-body line draft. The key challenge of this task lies in reference-based color transfer: the algorithm must not only fill colors into the target line art, but also accurately learn and reproduce the color style, shading relationship, and tonal atmosphere provided by a reference image, rather than performing arbitrary colorization.

As shown in Figure 2, the standard process of this task is as follows: a reference color image and a target line draft are taken as input, and the model generates a colorized result that is consistent with the target line art in structure and details, while remaining visually aligned with the reference image in terms of color style.

In AnimeDiffusion: Anime Diffusion Colorization, Cao et al. introduced the diffusion model into reference-based anime line-art colorization and addressed several limitations of earlier GAN-based methods, especially in controllability and generation quality[1]. Their method adopts a hybrid end-to-end training strategy. The first stage focuses on denoising and structural capture, while the second stage refines color alignment and correspondence with the reference image. As a result, the model produces higher-quality colorization results than earlier approaches and improves both automation and visual fidelity. The architecture of this method is illustrated in Figure 3.

(a) reference image (b) line drawing (c) original image

Figure. 2 Schematic diagram of the line art colorization task.

Compared with a simple description of line-art colorization as a conditional image generation task, a more systematic view shows that current animation colorization research mainly differs in three aspects: control signal, consistency objective, and generation backbone[2–4]. Early methods were mostly based on GANs and relied on user guidance such as scribbles, tags, or sparse interaction signals[2,3]. These methods improved interactivity and reduced manual effort, but often suffered from unstable shading, weak generalization, and limited fidelity to reference styles[2]. Later work gradually introduced stronger structural priors, attention mechanisms, and similarity-based matching to improve color propagation and local detail consistency[1,4]. More recent studies have shifted toward diffusion-based frameworks, which are

better at modeling complex color distributions and reference-guided appearance transfer, and therefore produce more natural and controllable results [1,4,5]. For example, MangaNinja emphasizes precise reference following, while AnimeDiffusion focuses on high-quality reference-based anime colorization with a diffusion pipeline [1,4]. At the video level, reference-based line-art video colorization further extends the problem from single-image consistency to temporal consistency across frames, which is essential for animation production [6,7].

Despite this progress, animation colorization still faces several persistent challenges. First, semantic ambiguity remains fundamental: line art contains sparse structural cues but little texture or material information, making it difficult for a model to infer plausible colors without external guidance [1,2]. Second, reference alignment is still difficult, especially when the target line art and the reference image differ in pose, viewpoint, or local shape details [1,4]. Third, many methods struggle with fine-grained controllability, meaning that users may want not only global style transfer but also precise local color placement [1,4,8]. Fourth, in production settings, temporal consistency becomes crucial, since frame-wise colorization can easily introduce flicker or color drift in animation sequences [6,7]. Finally, there is a continuing trade-off between automation and user control: highly automatic systems may reduce labor but can be less predictable, while strongly guided systems may require more manual interaction [1,2,4]. These issues suggest that future research should not only pursue better visual quality, but also improve controllability, robustness, and workflow compatibility for real animation production [1,4,6].

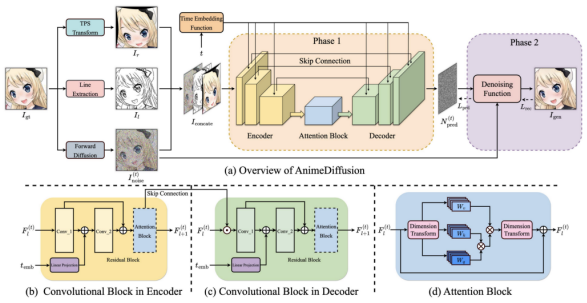

Figure. 3 Schematic diagram of the AnimeDiffusion method architecture.

Customized Image Generation of Anime Characters

Text-driven customized image generation of anime characters has become an important branch of generative animation research. Different from general text-to-image generation, this task not only requires semantic alignment between text prompts and generated images, but also demands stable preservation of a specific anime character's identity, including hairstyle, eye color, clothing elements, and drawing style. As a result, the core problem is no longer simple text-image correspondence, but how to reliably bind a fictional character concept to a generative model and reproduce that identity under diverse prompts and scenes.

Table 1 Representative methods for animation line-art colorization.

Method	Year	Backbone	Input / control
UDALC [2]	2018	GAN	Line art + user scribbles
Tag2Pix [3]	2019	GAN	Line art + text/color tags
DLAVC [6]	2020	Ref-based	Line-art video + few reference images
CSA [9]	2021	GAN+Attn	Line art + hints
UGFF [10]	2021	FillNet	Line art + user scribbles
Diffusart [5]	2023	Diffusion	Line art + user hints
Anime [1]	2024	Diffusion	Reference image + target line art
MangaNinja [4]	2025	Diffusion	Line art + reference image + point control

As shown in Figure 4, the standard pipeline usually takes a text prompt and one or more character reference images as input. The model is expected to generate a new image that follows the target prompt while maintaining the key visual traits of the designated anime character. For example, a prompt such as "Haibara is in chef's outfit" should not only produce a chef-themed anime image, but also preserve Haibara's recognizable identity in the generated result.

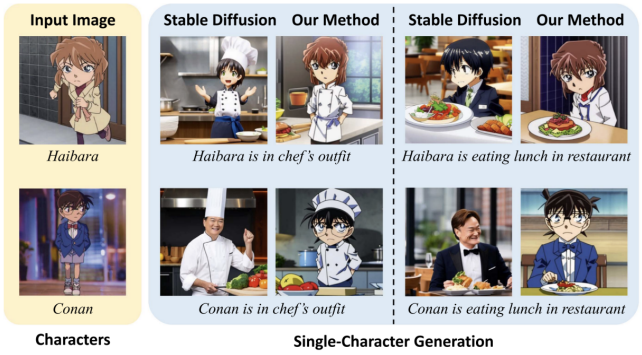

Figure. 4 Comparison of character-specific generation results.

A representative method for this task is AnimeDiff, proposed by Jiang et al. [8]. This method was designed to address two common problems in customized anime character generation: degradation of image quality and identity confusion in multi-character scenarios. Its key idea is to directly bind the target anime character's name to the corresponding visual concept, rather than introducing newly optimized identifier tokens that may distort the original text embedding space. In this way, the model can learn the unique appearance of a character more stably. In addition, AnimeDiff introduces a copy-paste data augmentation strategy to synthesize multi-character

© The National High School Journal of Science 2026

scenes during training, which improves the model's ability to distinguish different character attributes and reduces feature mixing across characters. The framework is illustrated in Figure 5.

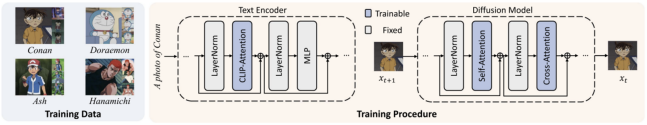

Figure. 5 Framework design of AnimeDiff for customized character generation.

From a broader perspective, current customized anime character generation methods mainly differ in three aspects: personalization mechanism, identity consistency objective, and parameter adaptation strategy. Early personalized generation methods such as Textual Inversion [11] represent a concept through a learned token embedding, while DreamBooth [12] binds a subject to a unique identifier by fine-tuning the diffusion model. Later methods such as Custom Diffusion [13] reduce training cost by updating only a subset of cross-attention parameters, and Mix-of-Show [14] further addresses multi-concept composition and identity conflicts. More recent methods, including AnimeDiff [8] and SerialGen [15], place stronger emphasis on identity preservation, whole-body consistency, and prompt controllability in character-centered generation.

Despite substantial progress, several challenges remain unresolved. First, identity drift is still common: a generated character may gradually deviate from the reference appearance when the prompt becomes more complex or the pose changes significantly. Second, attribute binding remains difficult, especially in multi-character scenes where visual attributes such as hair color or clothing style may be assigned to the wrong character. Third, many methods face a trade-off between text controllability and identity fidelity: stronger personalization may preserve the character better but reduce the model's ability to respond flexibly to new prompts. Fourth, style consistency is especially important in anime scenarios, since anime characters are not only defined by identity traits but also by a highly stylized visual language. Finally, efficiency and scalability remain practical concerns, because repeated fine-tuning for each new character can be costly in both storage and computation.

These observations suggest that future research should move beyond single-character reconstruction quality and focus more on robust identity preservation, compositional controllability, and scalable adaptation mechanisms for production-oriented anime content generation.

Table 2 Representative methods for customized anime character generation.

Method	Year	Backbone	Focus
Textual Inversion [11]	2022	Token	Embedding
DreamBooth [12]	2023	Full FT	Subject binding
Custom Diffusion [13]	2023	Partial FT	Efficient tuning
Mix-of-Show [14]	2023	LoRA	Multi-concept
AnimeDiff [8]	2024	Diffusion	Name binding
SerialGen [15]	2025	Two-stage	Body consistency

3D Animation: Image-to-Video Synthesis

Image-to-video character animation aims to generate a dynamic video sequence from a static reference image while preserving the character's appearance and following a specified motion signal. Compared with ordinary video generation, this task imposes two core constraints: appearance consistency and motion controllability. The first requires the generated frames to remain highly faithful to the reference image in terms of clothing texture, hairstyle, color, and identity details, so as to avoid flickering or visual drift. The second requires the model to accurately animate the character according to external control signals, such as pose sequences, skeletal keypoints, or driving videos.

As shown in Figure 6, the standard pipeline takes a static reference image together with a motion control sequence as input, and outputs a temporally coherent animation video. Under this setting, the character should complete the target motion while maintaining the original visual identity.

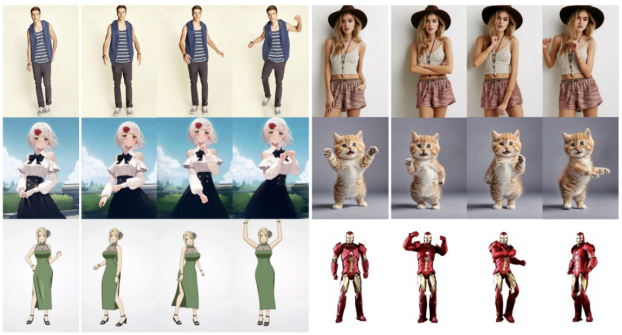

Figure. 6 Qualitative results of reference-based character animation generation.

A representative work in this direction is Animate Anyone, proposed by Hu et al. [16]. This method introduces a diffusion-based two-stage framework tailored for character animation. Its first key component is a ReferenceNet, which injects fine-grained appearance details from the reference image into the generation process through spatial attention, thereby improv-

ing appearance preservation. Its second key component is a pose guider together with temporal modeling, which allows the generated video to follow the target motion sequence while maintaining smooth transitions across frames. As illustrated in Figure 7, this framework provides a strong solution to the joint problem of consistency and controllability.

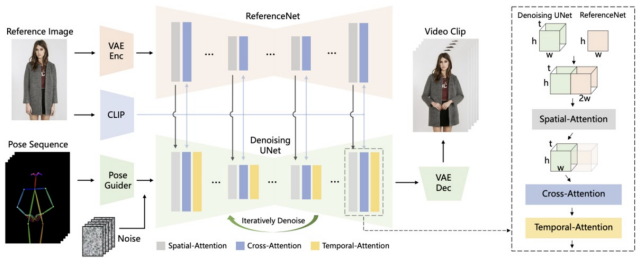

Figure. 7 Architecture overview of the diffusion-based character animation system.

From a broader perspective, current image-to-video character animation methods mainly differ in three aspects: motion guidance, appearance preservation strategy, and temporal modeling mechanism. Recent studies show a clear transition from frame-based or warping-based animation to diffusion-based video synthesis with stronger temporal reasoning [7,16–20]. For example, MagicAnimate enhances temporal consistency through a video diffusion model and an appearance encoder [7], while Champ introduces 3D parametric guidance based on SMPL to improve pose-shape alignment [17]. VividPose further improves identity retention and pose alignment by combining an identity-aware appearance controller with a geometry-aware pose controller [18]. More recent methods such as Cinemo and DisPose place stronger emphasis on controllable motion magnitude, smoother temporal transitions, and more generalizable pose control [19,20].

Despite the rapid progress, several challenges remain central to this task. First, appearance drift is still common when large motion changes or long sequences are involved, causing the generated character to deviate from the reference identity [7,16]. Second, motion control remains imperfect, especially when sparse pose signals fail to fully specify body shape, hand motion, or occluded regions [17,18,20]. Third, temporal inconsistency remains a major issue in video generation, often appearing as flicker, abrupt motion transitions, or unstable textures [7,19]. Fourth, many methods face a trade-off between controllability and visual realism: stronger constraints may improve motion following but can also reduce naturalness or diversity. Finally, generalization is still limited when the target character differs significantly from the driving source in body shape, style, or viewpoint [16,20]. These observations suggest that future research should focus not only on frame quality, but also on identity robustness, richer motion conditioning, and scalable video generation for real animation production.

Table 3 Representative methods for image-to-video character animation.

Method	Year	Backbone	Focus
Animate Anyone [16]	2024	Diffusion	Ref + pose
MagicAnimate [7]	2024	VideoDiff	Temp. consistency
Champ [17]	2024	Diffusion	3D guidance
VividPose [18]	2024	SVD	Identity + geometry
Cinemo [19]	2025	MotionDiff	Smooth control
DisPose [20]	2025	ControlNet	Pose disentangle

4D Animation

The core objective of 4D content generation is to inject dynamic changes into static 3D assets and produce animation sequences that remain consistent in both spatial and temporal dimensions. Compared with 2D image generation and static 3D modeling, 4D generation is substantially more challenging because it must preserve not only frame-level visual quality but also temporal continuity and multi-view geometric consistency. In other words, a successful 4D model should avoid distortions, flickering artifacts, and structural contradictions when an object is observed across different viewpoints and different time steps.

As shown in Figure 8, the standard paradigm of this task takes a static 3D object or scene representation as input and aims to output a dynamic 4D result. A common strategy is to first render the static 3D asset from multiple views, and then use these geometrically consistent renderings as conditions for generating temporally coherent multi-view videos. Based on these videos, the final 4D representation can be reconstructed or optimized.

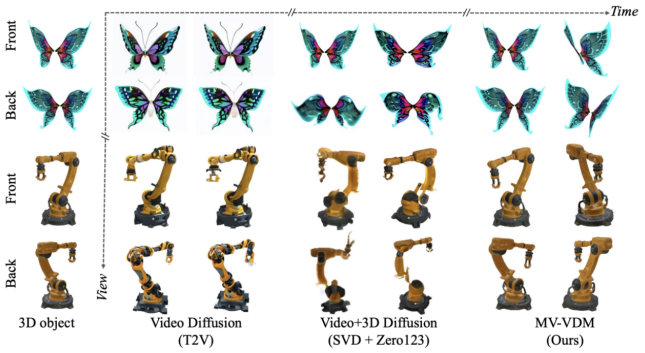

Figure. 8 Qualitative comparison of 4D generation results.

A representative work in this direction is Animate3D, proposed by Jiang et al. [21]. This framework addresses the com-

© The National High School Journal of Science 2026

mon problem of spatial-temporal inconsistency when animating existing static 3D models. Its main contribution is a customized multi-view video diffusion model, referred to as MV-VDM, which explicitly introduces consistency constraints across both time and viewpoint dimensions. Conditioned on multi-view renderings of a static 3D model, MV-VDM generates dynamic multi-view videos that preserve object identity while remaining temporally smooth. On top of this, Animate3D combines video-guided reconstruction with 4D score distillation sampling to refine the final dynamic geometry. As illustrated in Figure 9, the framework provides a practical route from static 3D assets to temporally and spatially coherent 4D animation.

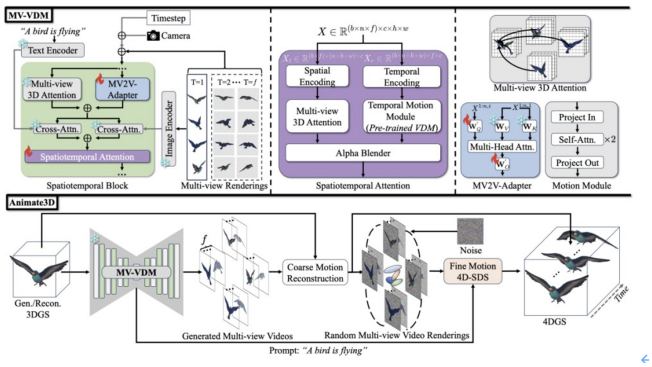

Figure. 9 Illustration of the MV-VDM and Animate3D framework for 4D generation.

From a broader perspective, current 4D generation methods mainly differ in three aspects: input condition, dynamic representation, and consistency mechanism. Some methods, such as 4D-fy, focus on text-to-4D generation through hybrid score distillation sampling[22]. Others, such as Consistent4D, 4Diffusion, and CAT4D, start from monocular video and use diffusion priors to recover multi-view dynamic content[23–25]. Another line of work emphasizes efficient explicit representations. For example, DreamGaussian4D and 4D Gaussian Splatting exploit Gaussian-based dynamic representations to improve rendering speed and geometric efficiency[26,27]. More recent methods such as Diffusion4D further emphasize fast spatial-temporal consistency by transferring temporal priors from video diffusion models into 4D generation pipelines[28]. Overall, the field is moving from slow optimization-heavy pipelines toward more scalable and controllable diffusion-driven 4D generation.

Despite this rapid progress, several important challenges remain. First, spatial-temporal inconsistency is still the most fundamental issue: a method may generate visually plausible frames while failing to preserve object geometry across views or time[21,28]. Second, motion controllability remains limited in many frameworks, especially when the dynamic behavior of a generated object must follow user-specified motion or semantic intent[22,26]. Third, optimization cost is still high for many 4D pipelines, especially those relying on score distillation sampling or reconstruction-heavy stages[22,24]. Fourth, representation quality versus efficiency is still a major trade-off: implicit fields may offer better fidelity, while explicit Gaussian-based methods offer better speed but may struggle with complex deformations[26,27]. Finally, generalization remains difficult, because methods trained on limited object or scene distributions may not transfer well to arbitrary assets, large deformations, or open-domain prompts[21,25]. These observations suggest that future research should pay greater attention to controllability, efficiency, and robust spatial-temporal coherence in production-oriented 4D animation systems.

Table 4 Representative methods for 4D animation generation.

Method	Year	Backbone	Focus
Consistent4D[23]	2024	DyNeRF	Mono video
4D-fy[22]	2024	SDS	Text-to-4D
DreamGaussian4D[26]	2024	4D-GS	Efficient 4D
4D-GS[27]	2024	Gaussians	Real-time render
4Diffusion[24]	2024	MV-VDM	Multi-view consistency
Animate3D[21]	2024	MV-VDM	3D-to-4D
Diffusion4D[28]	2024	VideoDiff	Fast consistency
CAT4D[25]	2025	MV-VDM	Video-to-4D

Discussion

The reviewed studies show that animation and human-computer interaction are becoming increasingly integrated. Animation is no longer only a passive visual product, but is gradually developing into a controllable and interactive medium. At the same time, HCI is no longer limited to traditional interface design, but now directly influences how animation is generated, edited, and experienced.

A clear trend in the literature is the shift from manual or weakly assisted production to AI-driven controllable generation. In 2D colorization, research has evolved from hint-based and GAN-based methods to diffusion-based frameworks with stronger controllability and better visual quality. In customized character generation, the focus has moved from general text-image alignment to identity-preserving generation. In image-to-video synthesis, static characters can now be animated under explicit motion control, while 4D generation further extends this process to spatiotemporally consistent dynamic content. These developments indicate that current animation research increasingly values not only realism and style, but also controllability and interaction.

However, several common challenges remain across these

tasks. The most important one is consistency. In 2D animation, this appears as stable color transfer and style preservation; in customized generation, it becomes character identity consistency; in image-to-video synthesis, it requires temporal smoothness and appearance fidelity; and in 4D generation, it further expands to spatial-temporal coherence across viewpoints and frames. Although recent diffusion-based approaches have improved performance significantly, consistency is still difficult to guarantee under complex conditions such as large motion, multiple characters, or viewpoint change.

Another important issue is the trade-off between automation and user control. Highly automatic systems reduce labor cost, but may also weaken predictability and fine-grained editing ability. By contrast, stronger control often requires more detailed input or more complex interaction. From an HCI perspective, this means that future animation systems should not only generate better results, but also provide more transparent and creator-friendly interfaces.

In addition, the current evaluation of this field is still fragmented. Most studies focus on technical metrics such as image quality, motion accuracy, or temporal consistency, while less attention is given to usability, creator experience, and interaction effectiveness. This suggests that future work should combine algorithmic evaluation with human-centered assessment.

Conclusion

This literature review has thoroughly discussed the intersection of animation and human-computer interaction (HCI), showing a critical shift in passive, frame-by-frame creation to exist as an immersive and interactive environment based on AI. The overview of the literature reveals a paradigm shift: the early models of GAN-based methods like Style2Paints have been replaced with the more advanced Latent Diffusion Models with better aesthetic quality and automatic control. At the heart of this development has been the effective reduction of consistency drift, which is attained by character identity binding in AnimeDiff, the Reference-Image Integration module in Animate Anyone, and multi-view video diffusion (MV-VDM) in Animate3D, all of which ensure structural and visual consistency across 2D, 3D and 4D time series. Moreover, the introduction of HCI technologies has completely changed the animation production process, reduced the creation barriers and providing more natural ways of interaction, such as gesture and voice recognition. Even with these advancements, the discipline continues to struggle with serious technical bottlenecks in inference latency, accuracy concerns, and the lack of a standardized evaluation methodology. The future should therefore emphasize more effective, creator-friendly tools and a universal set of principles that will ensure the maximum commercial and cultural utilization of this very promising age of hybrid media.

References

1 Y. Cao, X. Meng, P. Mok, T. Lee, X. Liu and P. Li, *AnimeDiffusion: Anime Diffusion Colorization*, 2024, 10.1109/TVCG.2024.3357568.

2 Y. Ci, X. Ma, Z. Wang, H. Li and Z. Luo, *User-Guided Deep Anime Line Art Colorization with Conditional Adversarial Networks*, 2018.

3 H. Kim, H. Jhoo, E. Park and S. Yoo, *Tag2Pix: Line Art Colorization Using Text Tag with SECat and Changing Loss*, 2019.

4 Z. Liu, K. Cheng, X. Chen, J. Xiao, H. Ouyang, K. Zhu, Y. Liu, Y. Shen, Q. Chen and P. Luo, *MangaNinja: Line Art Colorization with Precise Reference Following*, 2025.

5 H. Carrillo, M. Clement, A. Bugeau and E. Simo-Serra, *Diffusart: Enhancing Line Art Colorization with Conditional Diffusion Models*, 2023.

6 M. Shi, J. Zhang, S. Chen, L. Gao, Y. Lai and F. Zhang, *Deep Line Art Video Colorization with a Few References*, 2023.

7 Z. Xu *et al.*, *MagicAnimate: Temporally Consistent Human Image Animation Using Diffusion Model*, 2024.

8 Y. Jiang *et al.*, *AnimeDiff: Customized Image Generation of Anime Characters Using Diffusion Model*, 2024.

9 Z. Yuan *et al.*, *Line Art Colorization with Concatenated Spatial Attention*, 2021.

10 L. Zhang, C. Li, T. Wong, Y. Ji and C. Liu, *User-Guided Line Art Flat Filling with Split Filling Mechanism*, 2021.

11 R. Gal, Y. Alaluf, Y. Atzmon, O. Patashnik, A. Bermano, G. Chechik and D. Cohen-Or, *An Image is Worth One Word: Personalizing Text-to-Image Generation Using Textual Inversion*, 2022.

12 N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein and K. Aberman, *DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation*, 2023.

13 N. Kumari, B. Zhang, S. Wang, E. Shechtman, R. Zhang and J. Zhu, *Multi-Concept Customization of Text-to-Image Diffusion*, 2023.

14 Y. Gu, X. Wang, J. Wu, Y. Shi, Y. Chen, Z. Fan, W. Xiao, R. Zhao, S. Chang, W. Wu, Y. Ge, Y. Shan and M. Shou, *Mix-of-Show: Decentralized Low-Rank Adaptation for Multi-Concept Customization of Diffusion Models*, 2023.

15 C. Xie, H. Zou, R. Yu, Y. Zhang and Z. Zhan, *SerialGen: Personalized Image Generation by First Standardization Then Personalization*, 2025.

16 L. Hu, X. Gao, Z. Zhang, K. Yang, L. Zhao *et al.*, *Animate Anyone: Consistent and Controllable Image-to-Video Synthesis for Character Animation*, 2024.

17 S. Zhu, J. Chen, Z. Dai, Q. Su, Y. Xu, X. Cao, Y. Yao, H. Zhu and S. Zhu, *Champ: Controllable and Consistent Human Image Animation with 3D Parametric Guidance*, 2024.

18 Q. Wang, Z. Jiang, C. Xu, J. Zhang, Y. Wang, X. Zhang, Y. Cao, W. Cao, C. Wang and Y. Fu, *VividPose: Advancing Stable Video Diffusion for Realistic Human Image Animation*, 2024.

19 X. Ma, Y. Wang, G. Jia, X. Chen, T. Wong, Y. Li and C. Chen, *Consistent and Controllable Image Animation with Motion Diffusion Models*, 2025.

20 H. Li, Y. Li, Y. Yang, J. Cao, Z. Zhu, X. Cheng and L. Chen, *DisPose: Disentangling Pose Guidance for Controllable Human Image Animation*, 2025.

21 Y. Jiang, L. Zhang, J. Gao, W. Hu and Y. Yao, *Animate3D: Animating Any 3D Model with Multi-View Video Diffusion*, 2024.

22 S. Bahmani, W. Huang, J. Bai *et al.*, *4D-fy: Text-to-4D Generation Using Hybrid Score Distillation Sampling*, 2024.

23 Y. Jiang, L. Zhang, J. Gao, W. Hu and Y. Yao, *Consistent4D: Consistent 360 Degree Dynamic Object Generation from Monocular Video*, 2024.

24 H. Zhang *et al.*, *4Diffusion: Multi-View Video Diffusion Model for 4D Generation*, 2024.

© The National High School Journal of Science 2026

25 R. Wu, Y. Gao *et al.*, *CAT4D: Create Anything in 4D with Multi-View Video Diffusion Models*, 2025.
26 J. Ren, C. Pan, Z. Wang *et al.*, *DreamGaussian4D: Generative 4D Gaussian Splatting*, 2023.
27 G. Wu, T. Yi, J. Fang, L. Xie, X. Zhang, W. Wei, W. Liu, Q. Tian and X. Wang, *4D Gaussian Splatting for Real-Time Dynamic Scene Rendering*, 2024.
28 H. Liang *et al.*, *Diffusion4D: Fast Spatial-Temporal Consistent 4D Generation via Video Diffusion Models*, 2024.

© The National High School Journal of Science 2026