International Journal of Computational Research in Science and Technology
E-ISSN: 3108-0871, Volume 2 Issue 5
Sep-Oct-2026, PP 01-06

Artificial Intelligence In Visual Effects: A Review Of Automated Matting, Markerless Motion Capture, Neural Rendering And Ethical Production

Wavhal Vedant Hemant[1]; Prof. S. S. Bhor[2]; Prof. M. C. Kolse[3]; Prof. S. K. Said[4]
Department Of Artificial Intelligence And Data Science Jaihind College Of Engineering Pune, Maharashtra, India

Abstract—Artificial intelligence and machine learning are reshaping visual effects pipelines in modern cinema. Traditional VFX requires extensive manual labour for repetitive tasks such as rotoscoping, chroma keying, rigging and matchmoving. This paper reviews AI-driven techniques that automate these processes. It examines video object segmentation and high-resolution video matting that generate mattes without a green screen; structure-from-motion and visual SLAM that recover camera paths for matchmoving; markerless motion capture based on parametric body models and large motion-capture archives; neural rendering with neural radiance fields and real-time Gaussian splatting for novel-view synthesis and environment reconstruction; learned fluid simulation that approximates expensive physical solvers; and generative video synthesis and facial reenactment used for digital de-aging and performance transfer. The paper discusses how these methods improve efficiency and reduce post-production time and cost, and then examines the accompanying concerns of dataset copyright, performers' likeness and consent, data privacy and workforce displacement, together with content provenance standards as part of an ethical framework. A five-stage AI-assisted VFX pipeline covering acquisition, automated preparation, asset creation, rendering and review is proposed, with research gaps, objectives, a comparison of tasks and challenges. The review concludes that AI should function as a collaborative assistant that extends rather than replaces human creativity.

Index Terms—Visual effects, neural rendering, NeRF, Gaussian splatting, video matting, markerless motion capture, generative video, content provenance

I. Introduction

A single feature film shot may require an artist to trace the outline of a moving actor frame by frame, a process known as rotoscoping, or to adjust hundreds of tracking points so that a computer-generated object stays locked to the camera. Such tasks consume much of a visual effects budget without adding creative value.

AI now automates many of them. Segmentation networks produce mattes automatically, pose estimation captures performance without markers, and neural rendering reconstructs environments from ordinary footage. At the same time, generative tools raise questions about copyright, consent and jobs, which the industry is actively debating.

This paper reviews the main AI techniques used in VFX production and proposes a pipeline in which ethical safeguards are explicit.

II. Background and Fundamental Concepts

2.1 VFX Tasks and AI Techniques

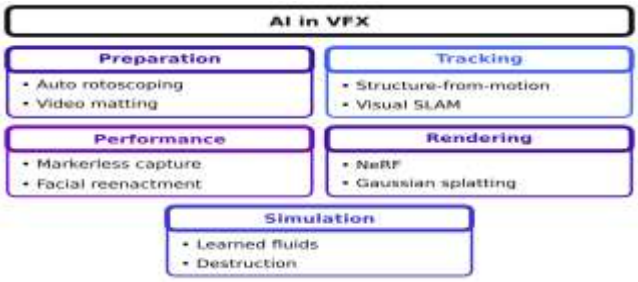


Fig. 1 maps the major VFX tasks to the AI methods that address them.
Fig. 1. VFX tasks and the AI techniques that address them.

2.2 Matting

Matting decomposes an image into foreground, background and an alpha channel. The compositing equation is

$$I = \alpha F + (1 - \alpha) B \qquad (1)$$

Traditional chroma keying solves this with a green screen; learned matting estimates α directly from arbitrary backgrounds.

III. Literature Survey

Segmentation and matting. Pont-Tuset *et al.* [1] established the DAVIS benchmark for video object segmentation, and Lin *et al.* [2] surveyed video segmentation and tracking, the basis of automated rotoscoping. Sengupta *et al.* [3] showed that a clean background plate allows high-quality matting without a green screen, and Lin *et al.* [4] achieved robust real-time high-resolution video matting with temporal guidance. Segment Anything [5] further generalized interactive segmentation.

Camera tracking. Schönberger and Frahm [6] improved structure-from-motion reconstruction, and Mur-Artal *et al.* [7] developed ORB-SLAM for accurate monocular camera tracking, both used for matchmoving.

Markerless motion capture. SMPL [8] provides a parametric body model that pose estimation methods fit to images, and AMASS [9] unified many motion-capture datasets into a common surface representation, enabling learning-based capture. OpenPose [10] supplies real-time 2D keypoints.

Neural rendering and generation. Mildenhall *et al.* [11] introduced neural radiance fields for photorealistic novel-view synthesis, and Kerbl *et al.* [12] achieved real-time rendering with 3D Gaussian splatting. Kim *et al.* [13] demonstrated deep video portraits that transfer head pose and expression, the technical basis of de-aging and performance transfer, and Wang *et al.* [14] synthesized temporally coherent video from semantic inputs. Kim *et al.* [15] trained a generative network to produce parameterized fluid simulations far faster than numerical solvers.

Provenance. The C2PA specification [16] defines signed provenance metadata for media, relevant as synthetic content becomes common. Table I summarizes these works.

TABLE I Summary of Literature Survey

Ref.	Author(s) / Year	Technique / Approach	Key Contribution	Limitation / Gap
[8]	Loper et al., 2015	SMPL body model	Parametric human model	Body only
[7]	Mur-Artal et al., 2015	ORB-SLAM	Accurate camera tracking	Textured scenes
[6]	Schönberger and Frahm, 2016	Structure-from-motion	Robust reconstruction	Offline
[13]	Kim et al., 2018	Deep video portraits	Pose / expression transfer	Ethical concerns
[11]	Mildenhall et al., 2020	NeRF	Photorealistic view synthesis	Slow, hard to edit
[3]	Sengupta et al., 2020	Background matting	No green screen needed	Needs clean plate
[4]	Lin et al., 2022	Robust video matting	Real-time high resolution	Hair / blur artefacts
[12]	Kerbl et al., 2023	3D Gaussian splatting	Real-time radiance fields	Large memory

IV. Research Gap

G1: Automated mattes still need artist correction on hair, motion blur and transparency.

G2: Neural rendering assets are difficult to edit with conventional tools.

G3: Training data often include copyrighted footage without clear licensing.

G4: Digital likeness use requires consent frameworks that are still evolving.

G5: Pipeline integration and version control for learned assets are immature.

V. Objectives

1. To review AI methods for matting, tracking, capture and rendering in VFX.

2. To examine generative techniques for de-aging and performance transfer.

3. To assess efficiency gains and remaining manual effort.

4. To propose an AI-assisted VFX pipeline with ethical safeguards.

5. To discuss copyright, consent, privacy and workforce concerns.

VI. Proposed System Architecture

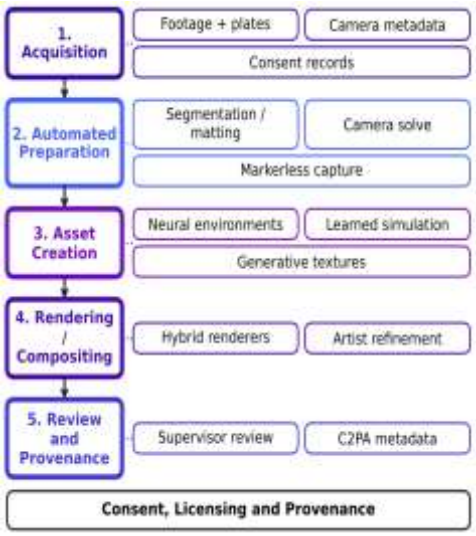


Fig. 2 presents the proposed pipeline.

Fig. 2. Proposed five-stage AI-assisted VFX pipeline.

6.1 Acquisition Stage

Footage, clean background plates, lens and camera metadata, and reference scans are captured on set, with written consent for likeness use.

6.2 Automated Preparation Stage

Shots are segmented and matted automatically, camera paths are solved by structure-from-motion or SLAM and performance is captured markerlessly.

6.3 Asset Creation Stage

Environments are reconstructed with neural radiance fields or Gaussian splatting; learned simulators generate fluid and destruction effects; generative tools assist texture and crowd creation.

6.4 Rendering and Compositing Stage

Learned and traditional renderers are combined; artists adjust mattes, lighting and integration in standard compositing software.

6.5 Review and Provenance Stage

Supervisors review shots; provenance metadata records which assets were AI-generated and under which licences and consents.

VII. Methodology and Working

Fig. 3 describes the workflow. Efficiency can be quantified by the reduction in manual hours,

$$Saving = (H_manual - H_AI) / H_manual \quad (2)$$

while quality is judged perceptually and by metrics such as alpha error for mattes and reprojection error for camera solves,

$$E_reproj = (1/N) \Sigma \|x_i - \pi(X_i)\| \quad (3)$$

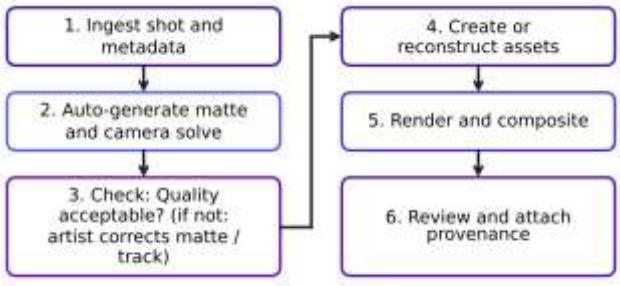


Fig. 3. AI-assisted shot production workflow.

VIII. Comparative Analysis

Table II compares VFX tasks by AI maturity, remaining manual effort and risk.

Task	AI Maturity	Manual Effort Left	Key Risk
Rotoscoping / matting	High	Edge cleanup	Quality on hair
Matchmoving	High	Verification	Low-texture shots
Motion capture	Medium–High	Cleanup, retarget	Occlusion
Environment rendering	Medium	Art direction	Editability
De-aging / likeness	Medium	Supervision	Consent, copyright

TABLE II Comparison of AI Adoption Across VFX Tasks

IX. Challenges and Future Scope

Table III summarizes challenges. Future directions include editable neural assets, real-time virtual production on LED stages, standardized consent and licensing for training data, and retraining programmes for artists moving into AI-assisted roles.

TABLE III Challenges and Mitigation

Challenge	Mitigation
Dataset copyright	Licensed and in-house training data
Likeness and consent	Written consent; usage limits
Editability of neural assets	Hybrid pipelines; export to standard formats
Workforce displacement	Reskilling; AI as assistant
Authenticity of media	C2PA provenance metadata

X. Conclusion

AI is transforming visual effects by automating repetitive preparation work and enabling new forms of rendering and capture, shortening schedules and reducing costs. The reviewed literature shows mature methods for matting, tracking, capture and neural rendering, with generative video advancing quickly. Realizing these benefits responsibly requires clear licensing, consent for digital likeness, provenance records and a pipeline in which AI assists rather than replaces the artists who shape the final image.

10.1 Acknowledgment

The authors express their sincere gratitude to the Department of Artificial Intelligence and Data Science, Jaihind College of Engineering, Pune, for providing the facilities, guidance and encouragement needed to carry out this work. The authors also thank the faculty members and colleagues whose valuable suggestions helped improve the quality of this review.

References

1. J. Pont-Tuset, F. Perazzi, S. Caelles, P. Arbeláez, A. Sorkine-Hornung, and L. Van Gool, “The 2017 DAVIS challenge on video object segmentation,” arXiv:1704.00675, 2017.
2. Z. Lin, Z. Zhang, Z. Chen, H. Lu, Y. Liu, and L. Zhang, “Video object segmentation and tracking: A survey,” ACM Trans. Intell. Syst. Technol., vol. 13, no. 4, Art. no. 59, 2022.
3. S. Sengupta, V. Jayaram, B. Curless, S. Seitz, and I. Kemelmacher-Shlizerman, “Background matting: The world is your green screen,” in Proc. IEEE/CVF CVPR, 2020, pp. 2288–2297.
4. S. Lin, L. Yang, I. Saleemi, and S. Sengupta, “Robust high-resolution video matting with temporal guidance,” in Proc. IEEE/CVF WACV, 2022, pp. 3132–3141.
5. A. Kirillov et al., “Segment anything,” in Proc. IEEE/CVF ICCV, 2023, pp. 4015–4026.
6. J. L. Schönberger and J.-M. Frahm, “Structure-from-motion revisited,” in Proc. IEEE CVPR, 2016, pp. 4104–4113.
7. R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, “ORB-SLAM: A versatile and accurate monocular SLAM system,” IEEE Trans. Robot., vol. 31, no. 5, pp. 1147–1163, 2015.
8. M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, “SMPL: A skinned multi-person linear model,” ACM Trans. Graph., vol. 34, no. 6, Art. no. 248, 2015.
9. N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. J. Black, “AMASS: Archive of motion capture as surface shapes,” in Proc. IEEE/CVF ICCV, 2019, pp. 5442–5451.
10. Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, “OpenPose: Realtime multi-person 2D pose estimation using part affinity fields,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 1, pp. 172–186, 2021.

11. B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “NeRF: Representing scenes as neural radiance fields for view synthesis,” in Proc. ECCV, 2020, pp. 405–421.
12. B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, “3D Gaussian splatting for real-time radiance field rendering,” ACM Trans. Graph., vol. 42, no. 4, Art. no. 139, 2023.
13. H. Kim et al., “Deep video portraits,” ACM Trans. Graph., vol. 37, no. 4, Art. no. 163, 2018.
14. T.-C. Wang et al., “Video-to-video synthesis,” in Proc. NeurIPS, 2018.
15. B. Kim, V. C. Azevedo, N. Thuerey, T. Kim, M. Gross, and B. Solenthaler, “Deep fluids: A generative network for parameterized fluid simulations,” Comput. Graph. Forum, vol. 38, no. 2, pp. 59–70, 2019.
16. Coalition for Content Provenance and Authenticity, “C2PA technical specification, version 1.3,” 2023. [Online]. Available: https://c2pa.org/specifications/