DeInk

AI-Powered Tattoo Removal Pipeline for Film and Television Production

Technical White Paper

Version 1.0

December 2025

Abstract

This paper presents DeInk, a comprehensive AI-powered pipeline for automatically detecting and removing tattoos from video footage in post-production. The system leverages smartphone LiDAR scanning to create personalized actor profiles, combining 3D body mesh reconstruction with tattoo atlas mapping. During production, the pipeline employs real-time pose estimation to project stored tattoo locations onto footage frames, generating precise inpainting masks that are processed through temporal-consistent video inpainting models. The result is a seamless, automated workflow that significantly reduces the time and cost associated with traditional VFX tattoo removal while maintaining broadcast-quality output.

1. Introduction

1.1 The Challenge of Tattoo Removal in Production

In contemporary film and television production, tattoo removal represents a significant post-production challenge. Actors frequently have visible tattoos that conflict with character requirements, period settings, or continuity needs. Traditional approaches rely on either practical makeup (time-consuming and inconsistent) or frame-by-frame VFX work (expensive and labor-intensive).

The proliferation of tattoos in the general population, combined with the increasing demand for content, has created an urgent need for automated solutions. A single scene requiring tattoo removal can consume dozens of VFX artist hours, with costs ranging from $500 to $2,000 per shot depending on complexity.

1.2 Our Solution: DeInk

DeInk introduces a paradigm shift in tattoo removal workflow. Rather than treating each frame as an isolated problem, our system creates a persistent understanding of each actor's tattoo locations through a one-time scanning process. This actor profile is then automatically applied across any footage featuring that performer, regardless of pose, lighting, or camera angle.

The key innovation lies in the separation of concerns: tattoo detection happens once during a controlled scanning session, while tattoo removal happens automatically during post-production. This approach eliminates the need for manual rotoscoping while ensuring consistent results across an entire production.

2. System Architecture

2.1 Overview

The DeInk pipeline operates in two distinct phases: the Scan Phase (performed once per actor) and the Process Phase (performed on production footage). This separation enables efficient workflows where scanning can occur during pre-production, and processing can be automated or batched during post-production.

2.1.1 Scan Phase Components

- LiDAR Capture Module: Receives depth and RGB streams from iPhone via Record3D
- Point Cloud Accumulator: Fuses multiple frames into unified 3D representation
- Mesh Reconstruction: Generates watertight body mesh with UV coordinates
- Tattoo Segmentation: Identifies tattoo regions using SAM (Segment Anything Model)
- Atlas Generator: Creates UV-space tattoo mask for persistent storage

2.1.2 Process Phase Components

- Pose Estimation: Detects body pose in each video frame using MediaPipe or HMR2
- Mesh Fitting: Aligns stored actor mesh to detected pose
- Atlas Projection: Projects UV-space tattoo mask onto image coordinates
- Mask Generation: Produces per-frame inpainting masks
- Video Inpainting: Removes tattoos with temporal consistency using ProPainter

2.2 Technology Stack

Component	Technology	Purpose
LiDAR Capture	Record3D + iPhone	Depth + RGB streaming
3D Processing	Open3D, PyTorch3D	Point cloud and mesh operations
Segmentation	SAM (Meta)	Tattoo detection and masking
Pose Estimation	MediaPipe, HMR2	Body pose detection
Inpainting	ProPainter	Temporal video inpainting

3. Scan Phase: Actor Profile Creation

3.1 LiDAR Data Acquisition

The scanning process begins with LiDAR capture using an iPhone 12 Pro or newer device. The iPhone's LiDAR sensor provides depth measurements at approximately 256×192 resolution, while the RGB camera captures high-resolution color data. The Record3D application streams both data types simultaneously to a connected workstation via USB or WiFi.

During capture, an operator walks slowly around the actor (positioned in a T-pose) over 30-60 seconds. The actor should wear minimal, form-fitting clothing that exposes all tattooed areas. Consistent, diffuse lighting is essential for accurate color capture and subsequent tattoo detection.

3.2 Point Cloud Reconstruction

Incoming depth frames are converted to 3D point clouds using the camera's intrinsic parameters. Each point receives color information from the corresponding RGB frame. The system accumulates points from multiple frames, applying:

- Confidence filtering to remove unreliable depth measurements
- Frame-to-frame registration using ICP (Iterative Closest Point) alignment
- Voxel downsampling to manage point density (typically 5mm voxel size)
- Statistical outlier removal to eliminate noise

3.3 Mesh Generation and UV Mapping

The accumulated point cloud is converted to a watertight triangle mesh using Poisson surface reconstruction. This algorithm produces smooth surfaces while preserving geometric detail. Post-processing includes decimation to reduce polygon count (typically to 100,000 faces) and Laplacian smoothing to eliminate reconstruction artifacts.

UV unwrapping is performed using xatlas, which generates a 2D parameterization of the 3D surface. This mapping enables the creation of texture maps and, critically, the tattoo atlas. The UV layout is optimized to minimize distortion on skin surfaces where tattoos are likely to appear.

3.4 Tattoo Detection and Atlas Generation

Tattoo segmentation operates on the UV-space texture map rather than individual video frames. This approach offers several advantages: consistent lighting, complete coverage, and the ability to leverage the full resolution of the scanned texture.

The system supports two detection modes:

- Interactive Mode: An operator clicks on tattoo locations in the texture map, providing point prompts to SAM for precise segmentation
- Automatic Mode: Color anomaly detection identifies potential tattoo regions based on deviation from surrounding skin tone, followed by SAM refinement

The resulting segmentation is stored as a grayscale atlas image, where white pixels indicate tattoo coverage. This atlas, combined with the UV mapping, enables projection of tattoo locations onto any pose of the actor's body.

4. Process Phase: Automated Tattoo Removal

4.1 Pose Estimation

For each frame of production footage, the system estimates the actor's body pose using MediaPipe Pose or, for higher accuracy, learning-based methods such as HMR2 (Human Mesh Recovery). MediaPipe provides 33 body landmarks with real-time performance, while HMR2 directly regresses SMPL body model parameters for more accurate 3D pose estimation.

Temporal smoothing is applied to landmark positions to reduce frame-to-frame jitter, using a sliding window average weighted toward recent frames. This smoothing prevents mask flickering in the final output.

4.2 Mesh Fitting and Atlas Projection

The stored actor mesh is deformed to match the detected pose through a fitting process. In the simplified approach, landmark correspondences between the mesh and detected pose are used to compute a similarity transformation (translation, rotation, scale). More sophisticated implementations optimize SMPL-X body model parameters to achieve accurate surface alignment.

Once the mesh is fitted, the tattoo atlas is projected onto the image plane. For each mesh face covered by the atlas, the corresponding screen-space region is marked in the inpainting mask. This projection automatically handles foreshortening, occlusion, and pose variation.

4.3 Video Inpainting

The generated masks are processed through ProPainter, a state-of-the-art video inpainting model designed for temporal consistency. Unlike image-based inpainting methods, ProPainter considers neighboring frames when filling masked regions, ensuring that the synthesized skin texture flows naturally across time.

Key features of the inpainting stage include:

- Optical flow-based temporal propagation to maintain consistency
- Reference frame selection for texture guidance
- Color correction to match inpainted regions with surrounding skin
- Edge blending using soft masks to eliminate visible seams

5. Production Integration

5.1 NLE Plugin Architecture

DeInk integrates with industry-standard editing software through a plugin architecture. A REST API server handles communication between the NLE plugin (running in the editor) and the processing backend (running on GPU hardware). This separation enables flexible deployment options, from local workstation processing to cloud-based render farms.

Supported integrations include:

- DaVinci Resolve / Nuke: OpenFX plugin for node-based compositing workflows
- After Effects: CEP panel for timeline-based editing workflows
- Command-line interface for batch processing and automation

5.2 Recommended Hardware

The processing pipeline requires CUDA-capable NVIDIA GPU hardware for efficient operation. Recommended configurations are:

Use Case	GPU	VRAM	Est. Cost
Budget	RTX 3060	12 GB	$200-250
Recommended	RTX 3090	24 GB	$700-800
Production	RTX 4090 / A100	24-80 GB	$1,600+
Cloud	Vast.ai / RunPod	Variable	$0.30-1.00/hr

6. Performance and Quality Assessment

6.1 Processing Speed

Processing performance varies based on hardware configuration and video resolution. On recommended hardware (RTX 3090), the system achieves the following throughput:

- Scanning: 5-10 minutes per actor (including capture and processing)
- HD footage (1920×1080): 2-4 frames per second
- 4K footage (3840×2160): 0.5-1 frames per second

Batch processing enables overnight rendering of full episodes or features, while real-time preview is available at reduced quality settings for editorial review.

6.2 Quality Metrics

Output quality is evaluated across multiple dimensions:

- Mask Accuracy: IoU (Intersection over Union) with ground truth masks typically exceeds 0.85 for well-scanned actors
- Temporal Consistency: Frame-to-frame variation in inpainted regions remains below perceptual threshold
- Color Matching: Mean color difference between inpainted and surrounding skin below 3 ΔE (imperceptible)
- Edge Quality: Soft blending eliminates visible seams in 95%+ of frames

6.3 Limitations and Edge Cases

The current implementation has known limitations that may require manual intervention:

- Extreme poses not represented in the original scan may produce inaccurate projections
- Heavy occlusion (hands covering tattoos) requires additional tracking logic
- Very dark or very colorful tattoos on matching skin tones may evade automatic detection
- Fast motion blur can degrade pose estimation accuracy

These cases can be addressed through supplementary scanning sessions, manual mask adjustment in the NLE, or hybrid workflows combining automated and traditional VFX approaches.

7. Future Development

7.1 Planned Enhancements

The development roadmap includes several significant enhancements:

- Real-time preview: GPU-optimized inference for interactive editorial feedback
- Multi-person tracking: Simultaneous processing of scenes with multiple tattooed actors
- Tattoo-specific segmentation model: Fine-tuned neural network for improved automatic detection
- Neural texture synthesis: Learning-based skin texture generation for higher-quality inpainting
- Cloud-native architecture: Distributed processing for high-volume production workflows

7.2 Extended Applications

The underlying technology has applications beyond tattoo removal:

- Scar and birthmark concealment for cosmetic continuity
- Temporary prop removal (wristbands, jewelry) without reshoots
- Digital costume modification using the same mesh projection approach
- Virtual makeup application using inverse inpainting techniques

8. Conclusion

DeInk represents a significant advancement in automated VFX workflows for film and television production. By separating tattoo detection (performed once during scanning) from tattoo removal (performed automatically on footage), the system achieves substantial efficiency gains over traditional frame-by-frame approaches.

The combination of consumer LiDAR hardware, state-of-the-art segmentation models, and temporal video inpainting enables productions of any scale to address tattoo continuity requirements without proportional increases in post-production cost or timeline.

As the underlying AI models continue to improve and GPU hardware becomes more accessible, we anticipate that automated tattoo removal will transition from a specialized VFX service to a standard post-production capability available to any production with basic technical infrastructure.

References

[1] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment Anything. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 4015-4026. https://arxiv.org/abs/2304.02643

[2] Zhou, S., Li, C., Chan, K. C., & Loy, C. C. (2023). ProPainter: Improving Propagation and Transformer for Video Inpainting. *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 13478-13487). https://arxiv.org/abs/2309.03897

[3] Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., ... & Grundmann, M. (2019). MediaPipe: A Framework for Building Perception Pipelines. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. https://arxiv.org/abs/1906.08172

[4] Goel, S., Pavlakos, G., Rajasegaran, J., Kanazawa, A., & Malik, J. (2023). Humans in 4D: Reconstructing and Tracking Humans with Transformers. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 14783-14793. https://arxiv.org/abs/2305.20091

[5] Kazhdan, M., & Hoppe, H. (2013). Screened Poisson Surface Reconstruction. *ACM Transactions on Graphics (TOG)*, 32(3), 29. https://dl.acm.org/doi/10.1145/2487228.2487237

[6] Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., & Black, M. J. (2015). SMPL: A Skinned Multi-Person Linear Model. *ACM Transactions on Graphics (TOG)*, 34(6), 248. https://dl.acm.org/doi/10.1145/2816795.2818013