CineCraft: Unified Shot Planning, Capture, and Post-Processing for Mobile Cinematography

Nhan (Nathan) Tran
Computer Science
Cornell University
Ithaca, New York, USA
nhan@cs.cornell.edu

Sam Belliveau
Electrical and Computer Engineering
Cornell University
Ithaca, New York, USA
srb343@cornell.edu

Zixin Xu
Computer Science
Cornell University
Ithaca, New York, USA
zx35@cornell.edu

Abe Davis
Computer Science
Cornell University
Ithaca, New York, USA
abedavis@cornell.edu

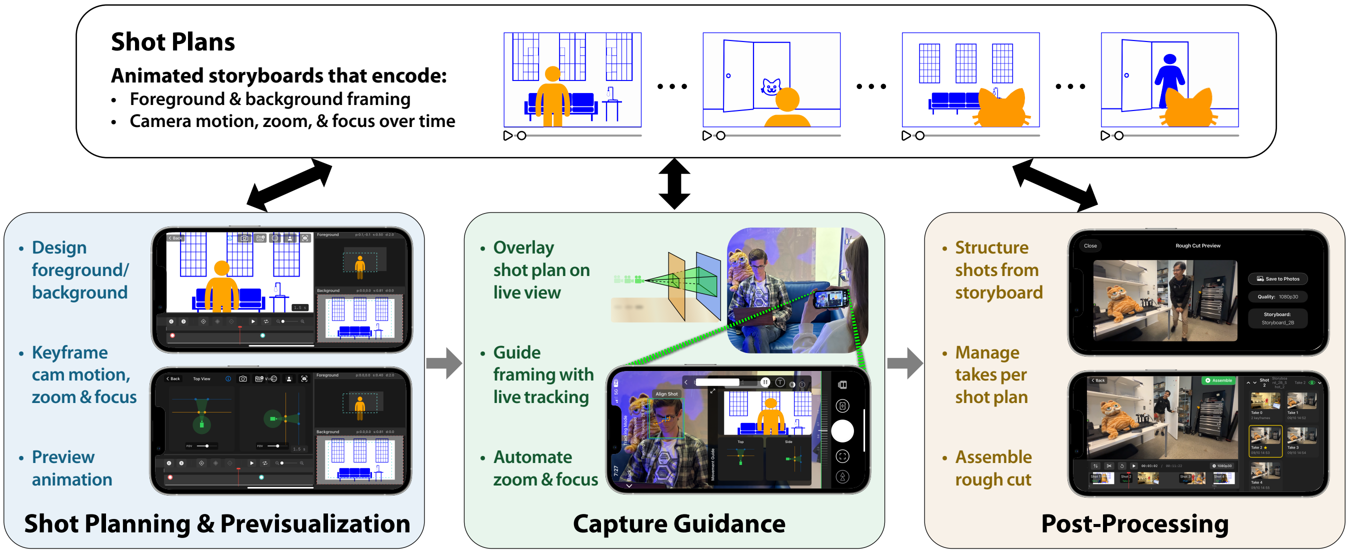

Shot Plans
Animated storyboards that encode:
• Foreground & background framing
• Camera motion, zoom, & focus over time
• Design foreground/ background
• Keyframe cam motion, zoom & focus
• Preview animation
Shot Planning & Previsualization
• Overlay shot plan on live view
• Guide framing with live tracking
• Automate zoom & focus
Capture Guidance
• Structure shots from storyboard
• Manage takes per shot plan
• Assemble rough cut
Post-Processing

Figure 1: CineCraft System Overview. Our tool uses animated storyboards—*shot plans* **(Top)**—to unify three stages of mobile filmmaking. Our shot planning interface **(Left)** lets users design and previsualize shots by keyframing foreground/background layers and camera motion. During capture **(Center)**, we overlay the shot plan on the live view, track subjects to guide framing, and automate zoom and focus. During post-processing **(Right)**, we use the planned storyboard to group captured footage under each shot panel, let users select the best take for each shot, and string them into a rough cut—all within a single app. Project page: https://megatran.github.io/cinecraft.

Abstract

We present CineCraft, an interactive mobile application that unifies planning, capture, and post-processing for cinematography on a single device. Our key design insight is to use a storyboard-like shot plan as a persistent representation that connects different stages of the filmmaking process, emulating coordination strategies used by professional film crews. Our shot plans extend common storyboarding conventions to encode time-varying parameters (e.g., camera movement, focus, and zoom) on a shared timeline, enabling previsualization during planning and precise synchronization during capture. CineCraft uses shot plans to generate camera movement instructions, provide augmented-reality (AR) framing guidance during filming, automate focus and zoom, and organize takes for review and rough-cut assembly. By consolidating stages that are often fragmented across separate mobile apps and ad hoc workflows, our system enables rapid on-location iteration with immediate playback. We demonstrate our system through a range of examples and two user studies.

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791532

CCS Concepts

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

Keywords

Mobile Cinematography, Augmented Reality Guidance, Camera Automation, Creativity Support Tools, Filmmaking Tools

ACM Reference Format:
Nhan (Nathan) Tran, Sam Belliveau, Zixin Xu, and Abe Davis. 2026. CineCraft: Unified Shot Planning, Capture, and Post-Processing for Mobile Cinematography. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3772318.3791532

1 Introduction

Mobile videography has become immensely popular with the rise of video-sharing platforms. However, the kind of dynamic shot design and complex camerawork that often characterizes professional film is rarely seen in mobile videos. This is largely due to the level of planning, coordination, and iteration that goes into executing advanced cinematography. A single scene may involve extensive planning and previsualization before a shoot takes place, precise synchronization of camera and scene properties during capture, and careful management of footage for rapid review and re-shooting once takes have been recorded.

On a professional film set, this is accomplished by a coordinated team of experts, each with distinct roles and responsibilities. While a single-user mobile app cannot replace all the functions of a film crew, it does offer a compact and convenient computational platform that we can use to design new creative tools. In this work, we draw inspiration from workflows used in professional film production to design an interactive application that helps users plan and execute complex cinematography on mobile devices.

We start by examining typical filmmaking workflows and how they divide tasks among crew members. We then propose a design for consolidating key roles into a single application that we call *CineCraft*[1]. We show that a single, storyboard-derived representation can serve as a *multi-functional backbone* for mobile cinematography—supporting ideation, guiding real-world execution, and organizing results for rapid iteration. Central to our design is the use of animated, storyboard-like *shot plan*s to connect different stages of film production:

- **Shot Planning & Previsualization**: The first stage in pulling off complex cinematography is planning. The coordination required to shoot scenes with precise timing and camera control does not happen spontaneously. Directors use storyboards to iterate on ideas and work out logistics, often long before anyone steps foot on set. Our shot plan structure is derived from storyboard conventions and organized into scenes, supporting both shot-level iteration and animatic previsualization.
- **Choreographed Camera Control**: Difficult shots often require precisely synchronized control over several camera parameters at once during capture, including camera position, rotation, focus, and zoom. On a professional camera crew, these parameters are often assigned to different human operators. Our innovation here is to use the shot plans that a user creates during planning to help automate control of some camera parameters during filming, making complex shots significantly easier to execute, even with limited crew and equipment.
- **Shot Management, Review, & Reshooting**: On a professional set, slates and shot logs help crews organize footage so editors can efficiently locate and assemble takes during post-production. Scenes and shot plans provide analogous structure for organizing captured data. Each shot plan in a scene can be captured and re-captured as many times as a user wants, and all resulting takes are organized so users can compare alternatives and select takes for automatically assembling rough cuts.

Connecting these stages in a single mobile app enables rapid end-to-end iteration, letting creators plan shots, execute complex camerawork, and review rough cuts from a pocket-sized mobile device.

Contributions

Our contributions include:

- CineCraft, an integrated mobile cinematography application connecting planning, AR-guided capture, and post-processing.
- A storyboard-derived shot plan representation that encodes time-varying camera intent and serves as the connective thread across all three stages.
- Evidence from two user studies: a technical validation study comparing CineCraft to the standard mobile camera app (n=8), and a complete workflow evaluation in which participants planned, captured multiple takes per shot, and assembled a short film (n=9).

2 Background: Cinematography and Production

Film production involves coordinated workflows across specialists and stages. Understanding these practices motivates our design.

2.1 Camera Crews and Division of Labor

Much of the inspiration for this work comes from existing practices in film production. In particular, our tool's automation of focus and zoom mirrors the standard division of responsibilities between the camera operator (sometimes performed by the cinematographer or the director of photography, or "DP") and the assistant camera (sometimes "focus puller" or "AC"). Typically, the camera operator handles the physical movement of the camera, while the assistant camera (1st AC) controls focus and occasionally zoom [4, 51].

2.2 Storyboarding Camera Maneuvers

Storyboards are a common tool for previsualization and shot planning in film production. Our interfaces are inspired by storyboarding techniques used to convey camera maneuvers [1, 21]. In this notation (Figure 2), numbered viewport rectangles overlaid on a scene sketch indicate the camera's starting and ending framing, giving the director a way to communicate intended camera movements to the crew during pre-production and on set.

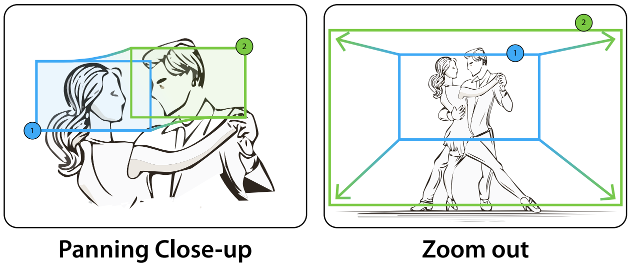


Figure 2: Storyboarding Camera Maneuvers. Two examples of storyboard notation for camera movement. **(Left)** A panning close-up that starts on one subject (1) and pans to another (2). **(Right)** A zoom out that starts tight on the subjects (1) and widens to reveal the full scene (2). CineCraft adopts this convention, letting users keyframe layered viewports on a timeline to define camera moves.

[1] Project page: https://megatran.github.io/cinecraft

2.3 Layer-Based Cinematography

Our interaction design draws inspiration from established animation workflows, specifically the Multiplane Camera [46] developed by Walt Disney Studios and modern industry-standard tools like Storyboard Pro [47]. These systems use layers moving at varied speeds to produce a convincing illusion of depth and parallax.

These tools offer an abstraction that lets animators reason about shots as layered compositions rather than explicit camera movements. CineCraft builds on this same abstraction: users compose shots using foreground and background planes, and the system derives the camera movement needed to achieve that composition.

2.4 Workflow Integration Challenges

Cinematography has traditionally involved disconnected stages. In professional productions, storyboarding, camera operation, focus pulling, stabilization, and editing each require separate tools and personnel, with inevitable information loss at each handoff [2, 27]. High-end solutions mitigate this fragmentation in different ways: hardware ecosystems like the DJI RS 4 Pro [9] couple LiDAR tracking with stabilization and focus motors, software frameworks like USD [41] and Virtual Production [13] let departments work within a shared digital environment. Generative AI approaches [19, 39] sidestep physical capture altogether by synthesizing video from text. Each addresses fragmentation at a different scale and cost.

CineCraft targets workflow integration for *live-action mobile capture*, where mobile deployment enables scenarios impractical with traditional equipment: spontaneous capture without setup time, solo on-location production, and rapid iteration with immediate playback, all on a single device. The shot plan serves as a persistent representation connecting planning, filming, and post-processing, ensuring creative intent flows end-to-end.

3 Related Work

3.1 Guidance for Mobile Photography & Video

Research on capture guidance has largely focused on static photography. Prior work has explored ways to derive dynamic AR-based guidance from photography principles related to framing and shot structure [11] as well as composition [10], while simultaneously using this guidance to teach users about those principles [12]. Other systems have helped users specify and reproduce photographic framing [28, 31], or combine on-device estimation of aesthetic criteria to offer suggestions during capture [34].

Extending to time-varying capture, systems like ReCapture [53], MeCapture [48], and ARticulate [49] utilize overlay guidance to help users reposition their camera to match a reference configuration, primarily to facilitate time-lapse creation. While effective for guiding users toward a target camera pose, these systems address an orthogonal goal to CineCraft, which coordinates focus, zoom, and movement across a planned shot sequence.

More closely related to our goals is ARCAM [30], which extracts camera movements from sample videos and visualizes them as AR trajectory overlays that novice users can follow during filming. Where ARCAM focuses on reproducing existing camera movements, our system is an authoring tool that enables users to design original shot plans. It also automates optical parameters such as focus and zoom during capture, while adapting to the inevitable variations of handheld execution and functioning as a digital assistant camera operator to help footage match the planned composition.

3.2 Virtual Camera Control & Previsualization

Previsualization describes techniques used by filmmakers to iterate on creative ideas before committing resources to full production. Different approaches strike different tradeoffs between cost and fidelity, from blocking out shots using cardboard cutouts [38] to sophisticated 3D computer graphics [8, 17, 22]. Some have explored previsualization in head-mounted VR [15] or AR settings [23, 44].

Recent tools have leveraged mobile devices to bridge the physical and virtual worlds. CamARa [29] focuses on virtual camera layout, allowing users to record camera motion paths in AR that are exported to 3D software like Blender. Similarly, CollageVis [24] allows filmmakers to create video collages by using a mobile device as a controller to navigate a 2.5D virtual stage.

These tools excel at previsualization and virtual layout, operating in proxy environments to plan camera work before physical production. CineCraft extends the pipeline further by functioning as an on-device production tool: it uses the shot plan to control mobile camera parameters during filming, carrying the user's vision from planning directly into final capture without requiring external tracking hardware or 3D scene reconstruction. More recently, generative AI systems like CineVision [50] have streamlined storyboard creation from scripts, reducing planning effort but not producing executable camera parameters. Our shot plan interface lets users compose through familiar layered abstractions, encoding camera control parameters inferred from the user's compositions that connect planning to physical capture and post-processing.

3.3 Computational Camera Control

Combining previsualization with camera automation is common in drone cinematography. Joubert et al. [26] did early work integrating drone path-planning with interactive previsualization based on Google Earth data. Subsequent work incorporated cinematography principles [25], path optimization based on high-level user goals [18], and chains of user-specified path segments [52]. Galvane et al. [16] explored directing drones in closer-range cinematic settings.

However, these systems rely on robotic actuation and absolute positioning (GPS/motion capture) unavailable to handheld mobile users. There are also approaches to camera control in 3D virtual environments, such as the toric space representation [32] for interpolating camera poses around identified targets. While relevant, these assume access to 3D scene information that does not hold in mobile capture settings.

On mobile devices, professional camera applications like Filmic Pro and Blackmagic Camera [3, 14] offer manual control over parameters but treat them in isolation, requiring users to adjust sliders while simultaneously moving the camera, a difficult task mirroring the division of labor on professional sets, where the camera operator handles movement while the 1st AC controls focus and zoom [4]. Our approach sits between manual apps and automated camera control. By utilizing a relative, image-space shot plan, we enable programmatic cinematography on handheld devices, automatically synchronizing focus and zoom to the user's physical capture task.

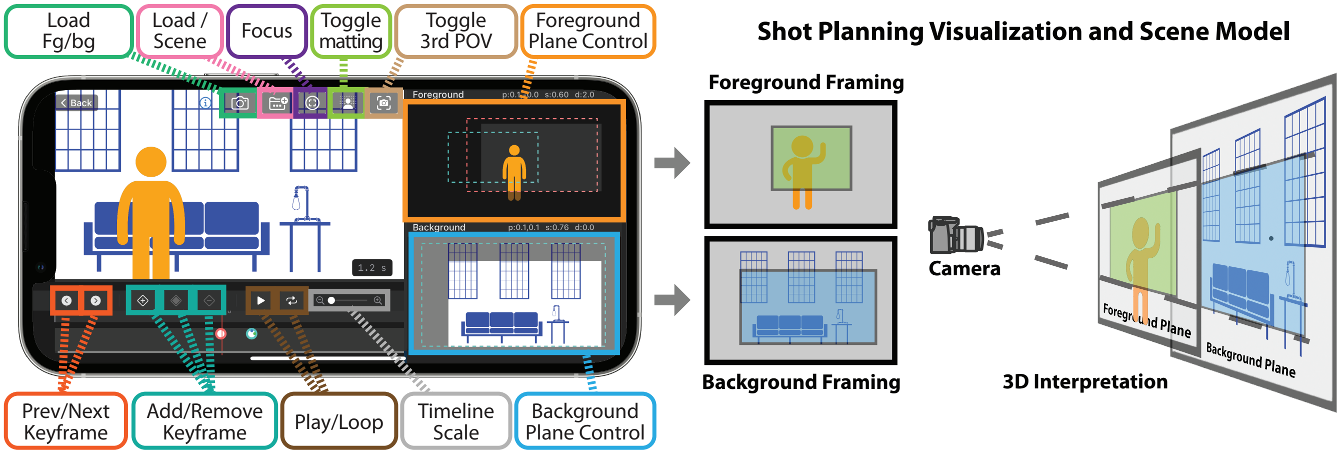


Figure 3: Shot Planning Interface & Scene Representation. **(Left)** Our shot planning interface shows a composite shot preview with a timeline and playback controls. The foreground and background panels display 2D maps of each layer with the camera's visible FOV highlighted by a rectangle. Users keyframe and adjust these rectangles to frame each layer, and the application automatically infers the corresponding 3D parameters. Keyframes are color-coded on both the timeline and the 2D maps, showing how the framing evolves across the shot. **(Right)** We represent the scene with parallel foreground and background planes suspended in 3D space.

3.4 Video Stabilization

Video stabilization removes unwanted camera shake to create smooth results. Hardware solutions such as gimbals and Steadicams offer real-time motorized stabilization [40]. Beyond passive stabilization, systems like LookOut [43] augment gimbal hardware with prescripted camera behaviors and real-time tracking to automate framing during long takes. On the software side, common approaches analyze footage and apply optimization techniques to smooth camera trajectories [20, 33].

While effective, these methods estimate the camera trajectory from captured footage and compute a smoothed version, without reference to what the camera was intended to do. CineCraft's stabilization differs by using the shot plan itself as a reference. Because the system knows what the camera was meant to do, it can preserve deliberate movements while correcting only unintended deviations, creating a direct connection between creative intent and final result.

4 Application Design

Our application supports three distinct stages of filmmaking: planning, capture, and post-processing. These stages are unified by our underlying shot plan representation. We first describe the design rationale behind this representation, then detail the interaction techniques employed at each stage.

4.1 Representing Shot Plans

To design a representation that is intuitive for filmmakers, we build upon established storyboarding conventions [1, 4, 21]. Typically, a storyboard captures narrative and composition via sketches, while camera framing is conveyed through viewport rectangles indicating the field of view (FOV). As shown in Figure 2, these static representations provide a blueprint for the shot. We extend this concept to achieve three key design goals:

4.1.1 Design Goals.

- **DG1: Meaningful at each stage of production**: Shot plans should provide effective previsualization during planning, clear guidance during capture, and preserve metadata for post-processing.
- **DG2: Robustness to scene scale**: The same shot plan should work across scenes with different depth ranges without requiring re-authoring.
- **DG3: Robustness to minor variations during capture**: Shot plans should tolerate inevitable deviations in handheld execution while preserving shot intent.

4.1.2 Two-Plane Representation. To address **DG1**, we extend the traditional single-plane storyboard to separate foreground and background planes (Figure 3). Users keyframe viewport rectangles on both planes independently. Separating the scene into two depth layers allows the system to distinguish between camera operations like dolly, zoom, pan, and dolly zoom, since each produces a distinct combination of changes across the two layers (Figure 3, right). This enables the previsualization of complex parallax effects that are difficult to convey with 2D sketches alone. This representation translates directly to capture guidance by computing the camera parameters to align the viewing frustum with these two viewports.

4.1.3 Scale Invariance. To achieve **DG2**, we decouple shot composition from absolute scene dimensions (Figure 4). By defining motion and zoom operations based on the relationship between the foreground and background planes, the same shot plan can adapt to different depth ranges. Users can adjust depth estimates for a more accurate preview, but the core compositional intent—the visual relationship between layers—remains transferable across scenes.

4.1.4 Relative Keyframing. Finally, **DG3** addresses the timing variations inherent in handheld capture. Cinematographic shots are typically characterized by the *type* of movement (e.g., pan, dolly)

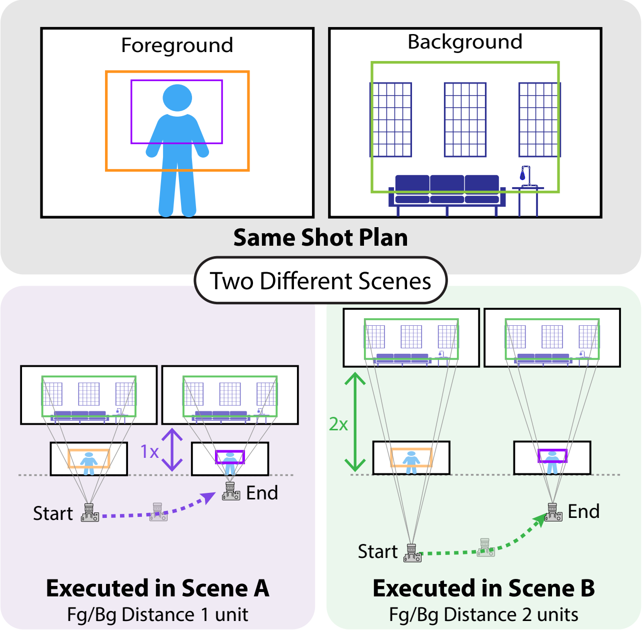


Figure 4: Scaling a shot plan to different scenes. (Top) A single shot plan defines a camera move from a start viewport (orange) to an end viewport (purple). **(Left)** When applied to a scene with a shallow depth (1 unit), the system calculates a specific camera trajectory. **(Right)** When applied to a deeper scene (2 units), the system automatically scales the magnitude of the camera's motion to maintain the exact same visual composition and relative timing, without requiring the user to re-author the shot.

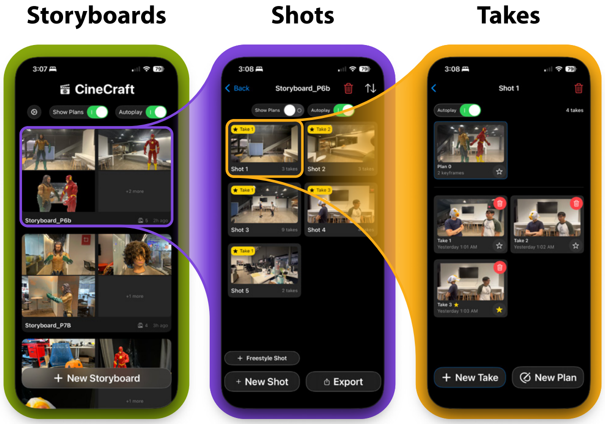


Figure 5: Storyboard View. The app organizes production assets hierarchically. **(Left)** The main view displays a list of storyboards. **(Center)** Selecting a storyboard reveals its shots arranged in narrative order. **(Right)** Tapping a shot opens the take management view, where users can record, review, and star their preferred take.

rather than precise velocity. Defining shot plans with absolute timestamps would cause slight user deviations to break the shot's logic. Instead, we use *Relative Keyframing*, where each keyframe is defined as a transformation relative to the preceding one. This preserves the geometric intent of the shot even if execution speed varies, ensuring the guidance remains valid despite user performance variations.

4.2 Shot Planning

Our shot planning interface (Figure 3, left) implements the two-plane representation to enable interactive previsualization. The interface consists of a shot preview, a timeline, and a camera control panel supporting two distinct modes.

4.2.1 Two-Plane Control. The default mode directly exposes our two-plane shot plan representation. We display the foreground and background planes, highlighting the current viewport against a dimmed background. Users can adjust these viewports and insert keyframes on the timeline; the system visualizes these keyframes with corresponding colors on each plane. As the timeline plays, the viewport interpolates between keyframes, providing the immediate previsualization capability central to **DG1**.

4.2.2 Third-Person Camera Control. We also offer a bird's-eye view mode (Figure 6, center) that renders the scene from third-person perspectives (top-down and side views). This allows users to specify camera motions, such as trucking (lateral movement) or booming (vertical movement), using familiar spatial metaphors. Users can drag the camera icon directly in 3D space to define the path. This visualization is also utilized later to convey shot instructions (Figure 7), helping users understand the physical movement required before they begin filming.

4.2.3 Focus Keyframes. Users can also keyframe focus pulls within the planning interface. To maintain scale independence (**DG2**), focus pulls are not defined by absolute distance, but by associating a keyframe with a specific point in the viewport. This ensures that the camera focuses on the intended subject regardless of the physical scene scale.

4.2.4 Storyboarding and Take Management. The Storyboarding interface allows users to organize plans into a narrative flow. Users can group, reorder, and rename shot plans within a grid layout that displays animated previews of each movement. The system allows users to record and manage multiple *takes* for each shot plan (Figure 5). This directly supports **DG1**: selected takes preserve their link to the original shot plan, ensuring that the creative intent defined during planning is carried through to post-processing. Additionally, shot plans can serve as reusable templates, supporting **DG2** by allowing users to repurpose previous compositions in new contexts.

4.2.5 Matting. To facilitate rapid prototyping, our Matting interface allows users to compose shot plans using simple 2D graphics or segmented photos (Figure 6). Inspired by directors who use action figures or scale models for blocking [42, 45], this interface supports **DG2** by enabling users to use placeholder elements that map to any real-world scale, for instance, using a toy figure to represent an actor. Users can capture, segment, and manipulate these elements (translate, rotate, scale) to compose the foreground and background

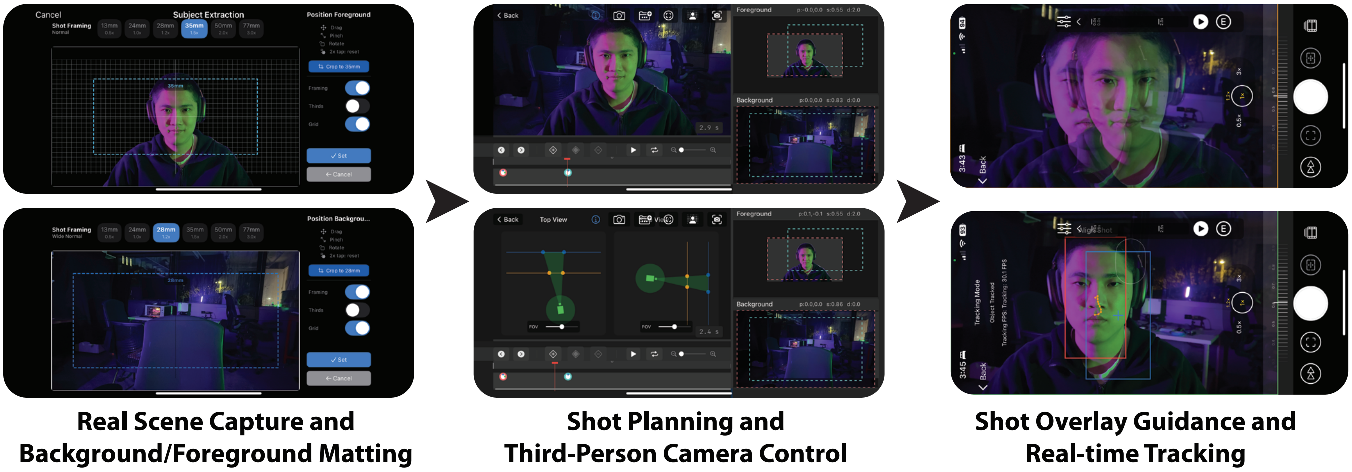

Real Scene Capture and Background/Foreground Matting
Shot Planning and Third-Person Camera Control
Shot Overlay Guidance and Real-time Tracking

Figure 6: Creating Shot Plans with Real-World Images. **(Left)** Users scout locations by capturing backgrounds and subjects using the device's camera. The system uses image matting to segment the foreground subject and allows users to apply cinematic lens presets. **(Middle)** These captured assets populate the two-plane planning interface, where users keyframe camera movement using either two-plane control (top) or third-person camera control (bottom). **(Right)** The resulting shot plan drives live overlay guidance during final capture. Users initiate real-time tracking by drawing a bounding box around the subject. Color-coded feedback indicates alignment to the shot plan (red when misaligned, green when aligned).

planes, enabling meaningful planning and visualization anywhere, even before arriving at the filming location.

4.3 Shot Instructions

While our two-plane interface allows users to declare *what* the shot should look like, users still need to understand *how* to move the camera to achieve it. To connect declarative planning with physical execution (**DG1**), we display a looping animation of the planned camera movement that users can preview before capture.

As shown in Figure 7, this visualization depicts the required camera motion in the XZ (top-down) and XY (screen-parallel) planes. This translation ensures the user is prepared for the physical movements required. Furthermore, by visualizing the expected path, users can better gauge the tolerance built into the system (**DG3**), understanding which deviations are acceptable during a take.

4.4 Shot Guidance

During filming, we leverage the pre-planned composition to provide real-time framing guidance. We overlay an animation of the shot plan directly onto the live camera view, offering continuous feedback on alignment.

To support **DG3** (tolerance to variation), the interface employs an adaptive tracking visualization (Figure 6). Users draw a bounding box to select a foreground subject. The system then displays two boxes: one tracking the subject and one following the planned trajectory. The subject box changes color (e.g., from red to green) based on the intersection with the planned trajectory. This feedback loop guides the user toward the intended framing without demanding robotic precision. The color coding indicates acceptable variance rather than perfect alignment.

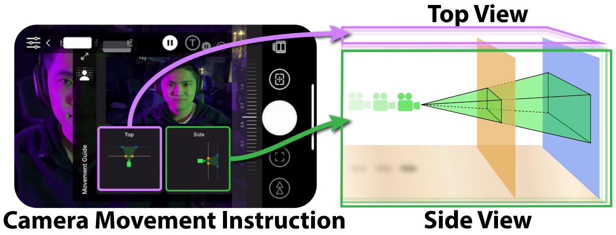


Figure 7: Shot Instruction View. **(Left)** During capture, the app overlays real-time instruction widgets on the camera feed, showing the user how to move the camera. **(Right)** Top and side views illustrate how these widgets map to the 3D scene, with the camera frustum (green) moving toward the target position.

Users can also reference a timeline at the top of the screen, which displays planned keyframes as visual cues for upcoming camera automation events. This allows users to "rehearse" the shot with the playback controls, verifying the automation behavior before committing to a recording.

4.5 Focus & Zoom Automation

Synchronizing zoom and focus with camera motion is one of the most challenging aspects of mobile cinematography. Our system automates these parameters to act as a virtual camera assistant. We track the foreground in real-time and adjust the focal length to match the subject's relative scale as defined in the shot plan.

This automation supports **DG2** by adapting zoom ranges to different scene scales, and **DG3** by compensating for imperfect manual execution. If the user's physical distance deviates from the plan, the system dynamically adjusts the zoom to maintain the intended framing, preserving visual composition despite execution variations.

4.6 Rough Cut Assembly

Our rough-cut interface facilitates the initial edit while maintaining the connection to the original plans (**DG1**). Users can sort through takes, trim clips, and select the best performances using the starring system carried over from the capture stage (Figure 8). Users can assemble and export a rough cut where every clip retains its semantic link to the shot plan. CineCraft exports captures in a hierarchical folder structure compatible with standard Non-Linear Editors (NLEs), allowing users to continue editing in their preferred tools without any conversion process (see Appendix C).

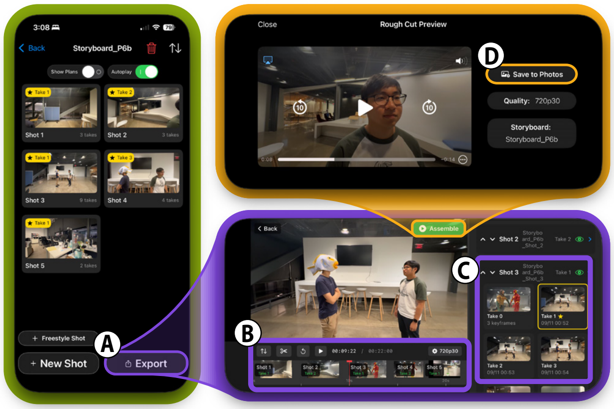


Figure 8: Rough Cut Assembly View. **(A)** The storyboard screen shows all shots with their takes; tapping *Export* opens the assembly workspace. **(B)** A multi-shot timeline allows trimming, reordering, and playback. **(C)** Selecting a shot reveals all its takes, letting users swap in their preferred take. **(D)** The preview screen provides export settings and lets users save the rough cut to the device.

5 Implementation

The system is built using native iOS frameworks with Swift and Metal. Because our two-plane representation is depth-invariant (DG2), the system requires no knowledge of scene geometry, avoiding resource-intensive ARKit scene understanding and 3D reconstruction. This frees on-device compute for real-time object tracking and low-level camera control across a wide range of devices.

5.1 Control System & Constraints

The application's core loop runs at 60 Hz. At each frame, the system processes input from the camera and the object tracker to update the application state. We implement two distinct control logic paths:

- **Time-based:** The system interpolates camera parameters based on the current playback timestamp.
- **Relative Keyframe-based:** The system advances the shot progress based on the geometric similarity between the current tracked composition and the target keyframe.

To ensure geometrically valid shot plans during this process, our constraint solver enforces three rules: a *size constraint* to maintain visual hierarchy between foreground and background, a *viewing angle constraint* to ensure realistic camera perspectives, and a *perspective constraint* to guarantee background visibility. Full constraint details appear in Appendix A.1.

5.2 Relative Keyframing

To support the relative guidance design, the system constructs a transformation tree that stores both absolute and relative transforms between keyframes (Algorithm 1 in Appendix A).

During capture, if the user's tracked position deviates from the plan, the system reconstructs the remaining trajectory from the current position using cached relative transforms (Algorithm 2 in Appendix A). This prioritizes expressiveness over rigidity: for a plan with three operations (e.g., truck left, dolly zoom, focus pull), if the first movement falls short of the planned endpoint, the system advances to the next camera maneuver from that position without requiring correction.

5.3 Post-Processing & Stabilization

Our on-device stabilization computes frame-by-frame similarity transforms that align captured footage with the planned shot trajectory, compensating for handheld jitter while preserving the intended camera path. Users can apply configurable smoothing filters (e.g., the 1€ filter [6]) to the tracking signal before stabilization.

6 User Study 1: Planning & Capture Validation

We conducted an initial study to evaluate CineCraft's planning and capture capabilities. This iteration included the shot planning interface (two-plane representation), capture guidance (AR overlays), and zoom/focus automation, but not yet takes management or rough-cut assembly. Scoping to these features let us isolate how shot plans support the planning-to-capture transition and validate robustness to execution variations. Study 1 thus addresses DG1 (meaningful across stages) primarily for planning and capture, and DG3 (robustness to minor variations). Study 2 evaluates the full workflow, including DG2 (scene-scale robustness).

6.1 Participants

We recruited eight participants (7 males, 1 female) through message boards, including filmmaking communities (IRB-approved). Five participants had formal film/media education and hands-on experience with specialized camera equipment through multiple productions. Three of them had served as Director of Photography on student films, and all five had worked as 1st AC across multiple projects, roles that demand intimate familiarity with complex camera movements and focus control. The remaining three participants represented more casual users: two were social media content creators, and one had no prior filmmaking experience. This mix let us observe how newcomers respond to the interface. The experienced participants also had extensive familiarity with video editing software, giving them a basis for evaluating how CineCraft's keyframing and timeline interactions align with the editing tools they already use in practice.

6.2 Study Design

We conducted 30-minute sessions with each participant, consisting of two tasks. Before each task, participants received a brief tutorial on the relevant features and practiced until they felt ready.

The first task asked participants to perform a dolly zoom using both our system and the standard camera app. Because no existing mobile app integrates shot planning with camera automation, the standard app serves as the natural baseline. The dolly zoom isolates a core challenge, synchronizing zoom with physical movement, where automation has a clear impact. After completing each condition, participants rated their experience on three metrics using 5-point Likert scales: execution success, shot reproducibility, and creative intent matching. They also provided detailed feedback about their strategies and challenges with each approach.

The second task involved a more complex cinematographic sequence: a chase scene requiring multiple camera techniques. Participants watched a sample video that combined lateral tracking, focus transitions, and a dolly zoom effect. Specifically, the sequence began with lateral camera movement following a foreground subject, transitioned focus to reveal background elements, and concluded with a dolly zoom that maintained the foreground subject size while revealing the background. Using our shot planning interface, participants planned and executed their interpretation of this sequence (see supplemental video).

After completing both tasks, participants provided additional feedback through questionnaires and open-ended responses about their experience, strategies, and areas for improvement.

6.3 Quantitative Results

Paired-samples t-tests and Wilcoxon signed-rank tests reveal statistically significant improvements across all three evaluation metrics when comparing our system to the standard camera app. As shown in Figure 9, for execution success, our system ($M = 4.50$, $SD = 0.54$) significantly outperforms the standard app ($M = 3.25$, $SD = 1.04$) ($t(7) = 3.42$, $p = .011$; Wilcoxon $W = 0$, $p = .031$, $r = .76$). Confidence in shot reproducibility shows similar improvements, with our system ($M = 4.25$, $SD = 0.71$) scoring higher than the standard app ($M = 2.50$, $SD = 1.31$) ($t(7) = 3.86$, $p = .006$; Wilcoxon $W = 0$, $p = .031$, $r = .76$). Creative intent matching shows the largest difference, with our system ($M = 4.38$, $SD = 0.52$) significantly exceeding the standard app ($M = 2.00$, $SD = 1.31$) ($t(7) = 4.77$, $p = .002$; Wilcoxon $W = 0$, $p = .016$, $r = .86$).

Individual feature effectiveness ratings are high across all components. Both camera automation and shot planning features receive the highest ratings ($M = 4.75$, $SD = 0.46$), followed by visual guidance ($M = 4.25$, $SD = 0.71$) and previsualization tools ($M = 4.25$, $SD = 0.71$). The low variance across these ratings suggests uniform reception across expertise levels.

6.4 Qualitative Insights

Participant feedback revealed three recurring themes:

6.4.1 Movement Control and Automation. With the standard camera app, participants consistently highlighted challenges with manual control, noting that *"the zoom is very jittery"* and *"difficult to change at a constant rate."* In contrast, participants responded positively to the automation: *"the auto zoom provides a smoother action compared to human hand zoom."* The tracking-based guidance was also well received, with participants appreciating having *"a solid thing to track when doing a complicated camera move."*

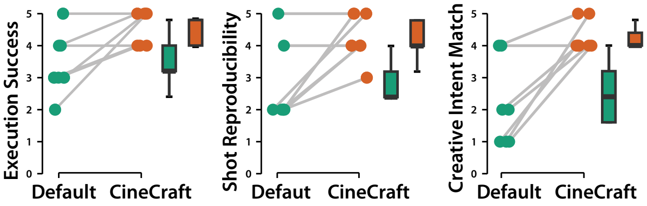


Figure 9: Quantitative Results (Study 1, Task 1). Participants rated CineCraft (orange) higher than the standard camera app (green) across all three metrics when performing a dolly zoom. Box plots show score distributions on a 5-point Likert scale. Gray lines connect ratings from the same participant, showing consistent within-subjects improvement.

6.4.2 Shot Planning and Previsualization. Participants appreciated how the system helps manage multiple parameters, noting it is *"very useful when working solo so you don't have to concern yourself with three things at once."* The previsualization features were also valued for multi-keyframe sequences, helping users *"understand how to transition from frame to frame in a complex shot."*

6.4.3 Alignment with Cinematographic Practice. Participants with cinematographic experience valued how our interface aligns with professional workflows. The observation that *"breaking down shot into subject and background makes good sense from a DP perspective"* reflected how our approach to shot composition mirrors established practice, where focus pullers rely on visual composition rather than metric values.

Overall, these findings reinforced our core design choices while also surfacing areas for refinement. Some noted potential visual interference (*"the subject tracker made the frame a bit busy"*), suggesting the need for customizable guidance visualization. One participant suggested drawing inspiration from the iPhone 16's flashlight controls for more intuitive parameter adjustment.

7 User Study 2: End-to-End Workflow

Building on Study 1, we extended CineCraft to include takes management and rough-cut assembly, completing the full workflow. We recruited nine participants (5 male, 4 female; ages 19–29, $M = 22.4$) via university message boards and filmmaking communities.

Five participants had filmmaking experience: all had completed film or media coursework, three had worked professionally on short films or commercial projects, four had experience with specialized equipment (gimbals, Steadicams), and four had prior storyboarding experience (hand-drawn sketches for class projects). None had used digital storyboarding tools. The remaining four participants were novices with no formal film training, no storyboarding experience, and primarily recorded casual videos (personal events, social media, or research documentation).

Table 1: CineCraft System Metrics (Study 2). Participants rated agreement on a 7-point scale (1=Strongly Disagree, 7=Strongly Agree) with statements adapted from the Creativity Support Index and feature-specific measures. p_{all}: Wilcoxon signed-rank test against scale midpoint (4.0); p_{group}: Mann-Whitney U comparing experienced vs. novice participants. Bold indicates $p < .05$.

Metric	**All (n=9)**			**Expert (n=5)**		**Novice (n=4)**		p_{group}
	M	SD	p_{all}	M	SD	M	SD	
CSI - Shot Plan								
UI: Supports planning needs	5.67	1.00	**.004**	5.20	0.84	6.25	0.96	.161
Visualization: Easy to visualize shots	6.11	1.27	**.008**	5.40	1.34	7.00	0.00	**.042**
Exploration: Easy to explore options	6.00	1.00	**.004**	5.40	0.89	6.75	0.50	**.037**
Expressiveness: Enabled creativity	5.78	1.48	**.012**	5.20	1.64	6.50	1.00	.199
Immersion: Engaged in planning	6.33	0.87	**.002**	5.80	0.84	7.00	0.00	**.042**
CSI - Capture								
UI: Supports capture needs	5.56	1.33	**.012**	5.20	1.30	6.00	1.41	.304
Instruction: Clear movement guidance	5.44	1.51	**.023**	5.40	1.82	5.50	1.29	1.00
Overlay: AR helped maintain framing	5.78	1.30	**.008**	5.00	1.22	6.75	0.50	**.030**
Automation: Helped achieve planned shots	6.00	1.22	**.008**	6.00	1.22	6.00	1.41	1.00
Exploration: No repetitive attempts	6.44	0.53	**.002**	6.20	0.45	6.75	0.50	.157
Expressiveness: Creative during capture	6.33	0.87	**.002**	6.20	0.84	6.50	1.00	.588
Engagement: Engaged in capture	6.22	0.83	**.002**	5.80	0.84	6.75	0.50	.116
CSI - Overall								
Exploration: Encouraged new approaches	6.56	0.73	**.002**	6.40	0.89	6.75	0.50	.661
Expressiveness: Could express ideas	6.33	0.71	**.002**	6.20	0.84	6.50	0.58	.687
Immersion: Immersed in process	6.11	1.05	**.004**	6.20	0.84	6.00	1.41	1.00
Results Worth Effort	6.22	1.30	**.008**	5.80	1.64	6.75	0.50	.558
Takes Management								
Record multiple takes	6.44	0.73	**.002**	6.40	0.55	6.50	1.00	.681
Organize and track takes	7.00	0.00	**.002**	7.00	0.00	7.00	0.00	1.00
Starring/favoriting best takes	6.78	0.44	**.002**	6.60	0.55	7.00	0.00	.237
Compare different takes	6.11	1.05	**.004**	5.60	1.14	6.75	0.50	.118
Multiple attempts helped explore	6.67	0.50	**.002**	6.40	0.55	7.00	0.00	.101
Assembly								
UI: Supports assembly needs	6.11	1.05	**.004**	5.40	0.89	7.00	0.00	**.013**
Preview how takes flow together	6.00	1.00	**.004**	5.60	1.14	6.50	0.58	.241
Creative intent preserved	6.22	1.09	**.004**	5.80	1.30	6.75	0.50	.281
Other Features								
Matting with real photos: Useful	6.56	0.53	**.002**	6.40	0.55	6.75	0.50	.396
Integration: Seamless workflow	5.89	0.93	**.002**	5.40	0.55	6.50	1.00	.116

7.1 Study Design

We conducted 45–60 minute sessions with each participant. During the first 15–20 minutes, participants familiarized themselves with the system by planning shots such as dolly zooms, rack focuses, and combinations of multiple operations, then capturing several takes and selecting their best during rough-cut assembly.

Once comfortable with the workflow, participants used the remaining time to plan, capture, and assemble their own short film entirely within CineCraft, consisting of 4–5 shots. For each shot, participants planned compositions and camera movements using our storyboarding interface, captured at least 2 takes using the AR-guided capture system, reviewed and starred their preferred takes, and assembled the final rough cut. Common shots included establishing shots, tracking movements, focus pulls, dolly zooms, and freestyle shots where participants used a static shot plan with no camera automation, typically for shot-reverse-shot dialogue. After completing their films, participants provided feedback through questionnaires and open-ended responses.

7.2 Quantitative Results

Participants filled out evaluation questionnaires adapted from the Creativity Support Index [7], along with feature-specific measures for takes management, assembly, and workflow integration (Table 1). We measured exploration, expressiveness, and immersion dimensions on 7-point Likert scales for each interface component.

Our adapted exploration items specifically assessed whether participants could achieve their intended results without tedious, repetitive attempts—directly measuring system tolerance to execution variations. Expressiveness items measured creative freedom, while immersion items assessed engagement and enjoyment.

All participants rated the system above the scale midpoint (4.0), indicating overall positive reception. Mann-Whitney U tests comparing experienced (n=5) and novice (n=4) participants revealed that novices rated several planning features significantly higher (Table 1). This pattern suggests CineCraft may be especially valuable for users without prior filmmaking or storyboarding background. The visual guidance helps novices understand shot composition in ways that experienced filmmakers may already have internalized through training and practice.

7.2.1 DG1: Meaningful at Each Stage. The planning interface received high visualization scores (M = 6.11, SD = 1.27). During capture, automation (M = 6.00, SD = 1.22) and AR overlay (M = 5.78, SD = 1.30) were also rated positively. The integrated workflow (M = 5.89, SD = 0.93) and overall results-worth-effort (M = 6.22, SD = 1.30) scores suggest value across stages. Novices rated planning visualization (Novice M = 7.00 vs. Experienced M = 5.40, p = .042) and AR overlay (Novice M = 6.75 vs. Experienced M = 5.00, p = .030) significantly higher, suggesting CineCraft's guidance particularly benefits users without internalized mental models of shot composition.

7.2.2 DG2: Robustness to Scene Scale. The compositional abstraction underlying our two-plane representation was well received. The matting feature (using real-world photo reference for foreground and background, Section 4.2.5) received high scores (M = 6.56, SD = 0.53), indicating users valued replacing abstract placeholders with captured images of their actual scene elements. Novice users rated planning exploration significantly higher than experienced users (Novice M = 6.75 vs. Experienced M = 5.40, p = .037), suggesting this compositional abstraction especially benefited users unfamiliar with traditional camera planning.

7.2.3 DG3: Robustness to Minor Variations. Capture exploration, which assessed whether participants could achieve results without repetitive attempts, showed the lowest variance of all metrics (M = 6.44, SD = 0.53), indicating consistent results despite handheld execution. All participants achieved usable takes regardless of experience level. The two groups did not differ significantly on capture exploration (Experienced M = 6.20 vs. Novice M = 6.75, p = .157), suggesting the system's tolerance for execution variations benefited both groups equally.

7.2.4 Takes Management and Assembly. Takes management received uniformly high scores: organizing and tracking takes (M = 7.00, SD = 0.00), starring/favoriting (M = 6.78, SD = 0.44), and recording multiple takes without disrupting flow (M = 6.44, SD = 0.73). The assembly interface showed a significant expert/novice difference (p = .013): novices gave perfect scores (M = 7.00) while experts rated it lower (M = 5.40). For assembly, participants could preview how takes flow together (M = 6.00, SD = 1.00) and felt the rough cut preserved their creative intent (M = 6.22, SD = 1.09).

7.3 Qualitative Validation

7.3.1 DG1: Meaningful at Each Stage. Participants across experience levels valued each stage of the workflow. For planning, P5 (experienced) noted it was *"intuitive and really helpful for visualizing the film with placeholders,"* while P8 (novice) found *"creating the storyboards with shots help me to overview the overall film plannings."* For capture, P7 (experienced) found *"the AR overlay was very helpful for framing,"* and P8 (novice) noted *"the AR overlay when doing dolly zoom was surprisingly efficient."* For post-processing, P4 (experienced) called the shot management *"the hidden gem in this app...metadata of what shot they belong to,"* and P3 (experienced) appreciated that *"it kept everything organized...really convenient."*

7.3.2 DG2: Compositional Abstraction and Robustness to Scene Scale. Participants appreciated planning camera movement through composition rather than metric specification. P6 (novice) captured this: *"just specifying how foreground and background would look like and let the app plan out the camera trajectory...I'm terrible at reasoning about how to move the camera given how I want the shot to look like."* P5 (experienced) found the matting feature *"helpful to be able to take photos for the background and foreground,"* noting this *"would make location scouting a lot more helpful."*

7.3.3 DG3: Robustness to Minor Variations. The system accommodated typical execution variations. Despite P5 (experienced) noting *"it's difficult to pull off movements of that specificity without a gimbal,"* participants achieved their creative intent. Consistent exploration scores across skill levels confirm robustness to minor deviations. One notable exception was P6's reverse dolly zoom (dolly-in while zooming out), which did not achieve the intended effect. Post-experiment analysis revealed a camera initialization bug (see supplemental materials).

7.3.4 Capture Guidance. Participants found the visual guidance and camera movement preview especially useful during capture. P5 (experienced) noted *"having the 'ghost' protagonist [shot plan overlay] on the screen during capture is super helpful—makes the shots a lot easier to capture."* P8 (novice) added that *"the camera instructions before taking the shot were the most helpful,"* while P6 (novice) appreciated being able to *"practice camera movements before actually capturing."*

7.3.5 Camera Parameters Automation. For capture automation, P4 (experienced) found *"the auto-zoom and timing markers were very helpful, as they helped me pace the shot and make consistently timed shots,"* and P9 (experienced) noted *"the automatic rack focus and the tracking of the face was great."* However, P7 (experienced) reported *"the automated dolly zoom didn't work when the camera lost track of the subject,"* highlighting its dependence on reliable object tracking.

7.3.6 Takes Management and Assembly. Assembly feedback differed by experience level. P8 (novice) found *"trimming each shot was easy to apply,"* reflecting the simplified interface's accessibility. In contrast, P4 (experienced) noted the interface felt *"underpowered for professional non-linear recording workflows,"* preferring to export to dedicated NLEs, a workflow CineCraft supports (Section 4.6) but was not evaluated in our study protocol. P7 (experienced) similarly remarked that *"trimming the shots in assembly was a bit difficult in the small timeline on the iPhone."* This pattern suggests CineCraft's assembly interface serves as an effective entry point for novices, while experienced users may prefer exporting to established NLEs where they already have preferred layouts and shortcuts.

8 Discussion

CineCraft makes professional cinematography more accessible on mobile devices. By unifying shot planning, AR-guided execution, and post-processing into a single workflow, the system helps users achieve complex multi-parameter shots without specialized equipment or crew coordination. Our studies show this integrated approach benefits users without prior cinematographic training while providing workflow efficiencies valued by experienced filmmakers.

8.1 Design Considerations for Mobile Cinematography

Our studies surfaced several design considerations for mobile cinematography tools. We discuss the three most prominent.

Previsualization Fidelity versus Execution Clarity. The classic cinematography manual by Mascelli notes that storyboards range from "simple outlines" to "detailed color renderings," serving different production needs [36]. We observed a parallel tension regarding asset fidelity in CineCraft. Most participants preferred the matting interface when the shot plan's foreground matched the real-world scene. However, when reusing a shot plan across different scenes (DG2), such as overlaying a photographed face onto a different actor, the high fidelity created visual interference that made guidance cues difficult to parse. In these cases, some participants preferred the abstract orange figure or tracking bounding box, noting it provided clearer spatial reference without obstructing their view.

Our tool supports both approaches (2D primitive assets and captured photo mattes) with an opacity slider for adjustment. These findings suggest a design principle of *adaptive abstraction*: creativity support tools should decouple planning fidelity from execution fidelity, providing rich textures for context during planning but simplifying to structural primitives during real-time execution.

Visual Information Density on Mobile Screens. Delivering professional cinematographic control on a mobile device requires presenting substantial information within limited screen real estate. Some users highlighted this tension: one noted the shot planner felt *"slightly dense,"* while another observed that the subject tracker made the frame *"busy"* given the difficulty of cramming features onto a small screen.

CineCraft currently addresses this by making visual guidance optional, but this places the burden of configuration on users. Since the shot plan already encodes the user's task structure, future work could adopt a *generative and malleable interface* approach [5], using the shot plan as a task-driven data model to regenerate the camera interface for each workflow phase—surfacing only the controls and guidance relevant to planning, capturing, or reviewing—rather than relying on users to manually configure visibility.

Pre-Planned Automation versus In-the-Moment Adaptation. Pre-planning camera parameters trades spontaneous adaptation for cognitive offloading. One participant articulated this directly: *"when using a regular camera app you have control in the moment which can be easier in some cases where...your rate of zooming changes with the rate at which you move forward in unpredictable ways."*

Our relative keyframing approach partially addresses this tension. Because each keyframe is defined as a transformation relative to the preceding keyframe rather than an absolute position, the system preserves shot intent even when execution timing varies. This provides a middle ground: users commit to the sequence of operations during planning but retain flexibility in execution pacing. For complex multi-parameter shots where cognitive load is high, this tradeoff favors pre-planning; for simpler improvisational shots, users can forgo automation entirely while still organizing takes within CineCraft's storyboard structure.

8.2 Enabling Solo and Educational Filmmaking

Multiple participants highlighted CineCraft's value for solo creators. By offloading simultaneous demands of zoom, focus, and movement, the system can help individuals achieve shots that typically benefit from coordinated crew roles. Participants also pointed to educational potential: P1 found the system *"surprisingly useful for training me on how to film properly."* The combination of previsualization, real-time capture guidance, and easy retakes may help users internalize cinematographic principles over repeated use. Since shot plans define relative compositions rather than absolute scene parameters, they are portable and shareable: a user could download a template for a specific camera maneuver or repurpose one from other projects and begin camera rehearsal immediately.

8.3 Broader Implications for Creative Tools

CineCraft's design offers insights that generalize to creativity support tools beyond mobile cinematography:

Bridging the Gulf of Execution. Users often think in goal space ("how the shot looks") rather than parameter space ("move camera 10cm"). This aligns with Norman's *Gulf of Execution*: the gap between a user's goals and the actions/controls available to carry them out [37]. This gap appears in many creative domains: Mascelli warns cinematographers against prioritizing mechanics over visual continuity [36], while Ma et al. demonstrate that novice artists struggle to decompose aesthetic goals into executable drawing steps, building computational tools to scaffold this translation [35].

Our approach helps bridge this gulf by letting users specify shots in goal space (composing frames with foreground and background assets) and automating the translation to camera parameters. Its *relative keyframing* mechanism further supports this by encoding each parameter change relative to the prior keyframe, preserving shot intent even when execution timing varies. More broadly, specifying relative behaviors rather than absolute coordinates may help creative tools degrade gracefully when execution conditions vary, supporting outcomes that remain aligned with intent.

Multi-Functional Representations. Traditionally, filmmakers must manually coordinate across distinct tools for storyboarding, shooting, and editing. CineCraft unifies these stages through a single *task-driven data model* [5] that evolves with the user's activity. The shot plan shifts from a compositional sketch (planning) to a guidance overlay (capture) to a take-management bin (assembly). By maintaining this shared representation, CineCraft preserves context and reduces loss of creative intent, providing structure for novices while minimizing information loss at the handoffs professionals rely on. However, spanning multiple stages requires the representation to adapt from detailed during planning to simplified during execution. As a broader insight, creativity support tools should, when possible, anchor ideation and production in a shared representation that adapts to each workflow phase.

8.4 Future Directions

These findings suggest several avenues for future work. Adaptive visual guidance that surfaces relevant controls based on task and expertise could reduce the configuration burden our participants identified, while maintaining flexibility. Generative AI could further streamline planning by connecting generic templates with location-specific imagery for rapid asset creation. Extending the current two-plane model to support additional layers could enable richer compositions with multiple depth planes, and as on-device spatial intelligence matures, it could augment such representations with real-time scene understanding for more contextual guidance during capture. Finally, specialized apps like Filmic Pro [14] and Blackmagic Camera [3] offer advanced capture controls beyond what our current implementation provides. Future work could integrate these or enable interoperability through shared shot plan representations; whether a unified app or an open protocol better serves cinematographers remains an open question.

9 Conclusion

We have presented CineCraft, an interactive tool that makes complex, multi-parameter cinematography more practical on mobile devices by connecting planning, capture, and post-processing through a unified shot plan representation.

Our key contribution is a storyboard-derived shot plan format that captures time-varying camera intent and uses it end-to-end: as an interactive planning and previsualization interface, as capture-time guidance and automation for focus/zoom, and as structure for intent-based stabilization, take management, and rough-cut assembly. Across two user studies, participants were better able to execute complex shots with CineCraft than with the standard mobile camera app, and could complete a full workflow that included multi-take iteration and assembly.

More broadly, CineCraft illustrates a general strategy for mobile creative tools: make coordination tractable by (1) representing intent in the terms creators naturally reason about and (2) automating the translation from that intent to low-level control at the moment of execution. These strategies suggest opportunities to expand the expressive range of mobile filmmaking and support faster, more iterative exploration of cinematic ideas.

Acknowledgments

This work was partially supported by a National Science Foundation Faculty Early Career Development Grant under award #2340448. We thank our study participants and testers, especially Shamus Li and Xinrui Liu, for their help and feedback in developing our app.

References

[1] Daniel Arijon. 1991. *Grammar of the film language* ([1st silman-james press ed.] ed.). Silman-James Press. Citation Key: GrammarOfTheFilmLanguage.

[2] Adrian Azzarelli, Nantheera Anantrasirichai, and David R. Bull. 2025. Intelligent Cinematography: a review of AI research for cinematographic production. *Artificial Intelligence Review* 58, 4 (25 Jan 2025), 108. doi:10.1007/s10462-024-11089-3

[3] BlackmagicDesign. 2023. Blackmagic Camera. https://apps.apple.com/us/app/blackmagic-camera/id6449580241. Accessed: 2025-04-08.

[4] Blain Brown. 2016. *Cinematography: Theory and Practice: Image Making for Cinematographers and Directors* (3rd ed.). Routledge.

[5] Yining Cao, Peiling Jiang, and Haijun Xia. 2025. Generative and Malleable User Interfaces with Generative and Evolving Task-Driven Data Model. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 686, 20 pages. doi:10.1145/3706598.3713285

[6] Géry Casiez, Nicolas Roussel, and Daniel Vogel. 2012. 1 € filter: a simple speed-based low-pass filter for noisy input in interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2527–2530. doi:10.1145/2207676.2208639

[7] Erin Cherry and Celine Latulipe. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. *ACM Trans. Comput.-Hum. Interact.* 21, 4, Article 21 (June 2014), 25 pages. doi:10.1145/2617588

[8] Marc Christie and Patrick Olivier. 2009. Camera control in computer graphics: models, techniques and applications. In *ACM SIGGRAPH ASIA 2009 Courses* (Yokohama, Japan) *(SIGGRAPH ASIA '09)*. Association for Computing Machinery, New York, NY, USA, Article 3, 197 pages. doi:10.1145/1665817.1665820

[9] DJI. 2025. DJI RS 4 Pro – Transcend Potential. https://www.dji.com/rs-4-pro. Accessed: 2025-12-05.

[10] Jane E., Kevin Y. Zhai, Jose Echevarria, Ohad Fried, Pat Hanrahan, and James A. Landay. 2021. Dynamic Guidance for Decluttering Photographic Compositions. In *The 34th Annual ACM Symposium on User Interface Software and Technology (UIST '21)*. Association for Computing Machinery, New York, NY, USA, 359–371. doi:10.1145/3472749.3474755

[11] Jane L. E, Ohad Fried, Jingwan Lu, Jianming Zhang, Radomír Mech, Jose Echevarria, Pat Hanrahan, and James A. Landay. 2020. Adaptive Photographic Composition Guidance. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3313831.3376635

[12] Jane L. E. and James A. Landay. 2021. *Designing Photography Guidance for Rapid In-Camera Iteration.* Springer International Publishing, Cham, 151–165. doi:10.1007/978-3-030-76324-4_8

[13] Epic Games. 2021. *The Virtual Production Field Guide, Volume 2.* Technical Report. Epic Games. https://www.unrealengine.com/en-US/blog/volume-2-of-the-virtual-production-field-guide-now-available

[14] FilmicApps. 2023. Filmic Pro - Video Camera. https://apps.apple.com/us/app/filmic-pro-video-camera/id436577167. Accessed: 2025-04-08.

[15] Quentin Galvane, I-Sheng Lin, Fernando Argelaguet, Tsai-Yen Li, and Marc Christie. 2019. VR as a Content Creation Tool for Movie Previsualisation. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 303–311. doi:10.1109/VR.2019.8798181

[16] Quentin Galvane, Christophe Lino, Marc Christie, Julien Fleureau, Fabien Servant, François-Louis Tariolle, and Philippe Guillotel. 2018. Directing Cinematographic Drones. *ACM Trans. Graph.* 37, 3, Article 34 (jul 2018), 18 pages. doi:10.1145/3181975

[17] Jean-Marc Gauthier. 2005. *Building interactive worlds in 3D: Virtual sets and pre-visualization for games, film & the web.* Focal Press. Citation Key: book:383706.

[18] Christoph Gebhardt, Benjamin Hepp, Tobias Nägeli, Stefan Stevšić, and Otmar Hilliges. 2016. Airways: Optimization-Based Planning of Quadrotor Trajectories according to High-Level User Goals. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2508–2519. doi:10.1145/2858036.2858353

[19] Google DeepMind. 2025. *Veo 3 Technical Report.* Technical Report. Google DeepMind. https://storage.googleapis.com/deepmind-media/veo/Veo-3-Tech-Report.pdf

[20] M. Grundmann, V. Kwatra, and I. Essa. 2011. Auto-directed video stabilization with robust L1 optimal camera paths. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*. IEEE Computer Society, USA, 225–232. doi:10.1109/CVPR.2011.5995525

[21] John Hart. 2007. *The Art of the Storyboard: A filmmaker's introduction* (2 ed.). Focal Press. Citation Key: ArtOfTheStoryboardBook.

[22] Li-wei He, Michael F. Cohen, and David H. Salesin. 2023. *The Virtual Cinematographer: A Paradigm for Automatic Real-Time Camera Control and Directing* (1 ed.). Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3596711.3596786

[23] Ryosuke Ichikari, Keisuke Kawano, Asako Kimura, Fumihisa Shibata, and Hideyuki Tamura. 2006. Mixed reality pre-visualization and camera-work authoring in filmmaking. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*. 239–240. doi:10.1109/ISMAR.2006.297823

[24] Hye-Young Jo, Ryo Suzuki, and Yoonji Kim. 2024. CollageVis: Rapid Previsualization Tool for Indie Filmmaking using Video Collages. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 164, 16 pages. doi:10.1145/3613904.3642575

[25] 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]

[26] Niels Joubert, Mike Roberts, Anh Truong, Floraine Berthouzoz, and Pat Hanrahan. 2015. An interactive tool for designing quadrotor camera shots. *ACM Trans. Graph.* 34, 6, Article 238 (nov 2015), 11 pages. doi:10.1145/2816795.2818106

[27] Steven Douglas Katz. 1991. *Film directing shot by shot: visualizing from concept to screen.* Gulf Professional Publishing.

[28] Minju Kim and Jungjin Lee. 2019. PicMe: Interactive Visual Guidance for Taking Requested Photo Composition. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300625

[29] Hyewon Lee, Christopher Bannon, and Andrea Bianchi. 2025. CamARa: Exploring and Creating Camera Movements with Spatial Reference in Augmented Reality. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 703, 5 pages. doi:10.1145/3706599.3721180

[30] Jiefeng Li, Yingying She, Fang Liu, Chun Yu, Xiaoli Wang, and Yuxin Xu. 2022. Augmented Reality Based Video Shooting Guidance for Novice Users. *Proceedings of the ACM on Human-Computer Interaction* 6, MHCI (Sep 2022), 215:1–215:20. doi:10.1145/3546750

[31] Jie Lin and Chuan-Kai Yang. 2020. A Multi-Person Selfie System via Augmented Reality. *Computer Graphics Forum* 39, 7 (2020), 553–564. doi:10.1111/cgf.14167

[32] Christophe Lino and Marc Christie. 2015. Intuitive and efficient camera control with the toric space. *ACM Trans. Graph.* 34, 4, Article 82 (July 2015), 12 pages. doi:10.1145/2766965

[33] Feng Liu, Michael Gleicher, Jue Wang, Hailin Jin, and Aseem Agarwala. 2011. Subspace video stabilization. *ACM Trans. Graph.* 30, 1, Article 4 (Feb. 2011), 10 pages. doi:10.1145/1899404.1899408

[34] Hao Lou, Heng Huang, Chaoen Xiao, and Xin Jin. 2021. Aesthetic Evaluation and Guidance for Mobile Photography. In *Proceedings of the 29th ACM International Conference on Multimedia (MM '21)*. Association for Computing Machinery, New York, NY, USA, 2780–2782. doi:10.1145/3474085.3478557

[35] Jiaju Ma, Chau Vu, Asya Lyubavina, Catherine Liu, and Jingyi Li. 2025. Computational Scaffolding of Composition, Value, and Color for Disciplined Drawing. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, Article 161, 15 pages. doi:10.1145/3746059.3747605

[36] Joseph V Mascelli. 1965. *The five C's of cinematography.* Vol. 1. Grafic Publications Hollywood.

[37] Donald A Norman and Stephen W Draper. 1986. *User centered system design; new perspectives on human-computer interaction.* L. Erlbaum Associates Inc.

[38] Thomas Ohanian and Natalie Phillips. 2013. *Digital filmmaking: the changing art and craft of making motion pictures.* Routledge.

[39] OpenAI. 2024. Video Generation Models as World Simulators. https://openai.com/index/video-generation-models-as-world-simulators/. Technical Report, February 2024.

[40] Sambhram Pattanayak, Saad Ullah Khan, Fazal Malik, and Somanath Sahoo. 2024. Automating Camera Movements: AI-Driven PTZ Cameras in Film Production. In *Innovative and Intelligent Digital Technologies; Towards an Increased Efficiency: Volume 1*, Muneer Al Mubarak and Allam Hamdan (Eds.). Springer Nature Switzerland, Cham, 627–639. doi:10.1007/978-3-031-70399-7_48

[41] Pixar Animation Studios. 2016. Universal Scene Description. https://openusd.org. Open source 3D scene interchange framework.

[42] Stefania Sarrubba. 2020. This filmmaker used Barbie dolls to storyboard her debut feature. https://lwlies.com/articles/my-fiona-barbia-doll-story-board-kelly-walker/

[43] Mohamed Sayed, Robert Cinca, Enrico Costanza, and Gabriel Brostow. 2022. LookOut! Interactive Camera Gimbal Controller for Filming Long Takes. *ACM Trans. Graph.* 41, 3, Article 30 (March 2022), 16 pages. doi:10.1145/3506693

[44] Midieum Shin, Byung-soo Kim, and Jun Park. 2005. AR storyboard: an augmented reality based interactive storyboard authoring tool. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*. 198–199. doi:10.1109/ISMAR.2005.12

[45] Filmmaker Staff. 2018. "I literally shot the whole movie with dolls on a miniature set": Director Nicolas Pesce. https://filmmakermagazine.com/104483-i-literally-shot-the-whole-movie-with-dolls-on-a-miniature-set-director-nicolas-pesce-piercing/

[46] J. P. Telotte. 2006. Ub Iwerks' (Multi)Plain Cinema. *Animation* 1, 1 (2006), 9–24. arXiv:https://doi.org/10.1177/1746847706065838 doi:10.1177/1746847706065838

[47] Toon Boom Animation. 2022. Storyboard Pro. https://www.toonboom.com/products/storyboard-pro. Accessed: 2025.

[48] Nhan Tran, Ethan Yang, Angelique Taylor, and Abe Davis. 2024. Personal Time-Lapse. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 56, 13 pages. doi:10.1145/3654777.3676383

[49] Nhan (Nathan) Tran, Ethan Yang, and Abe Davis. 2025. ARticulate: Interactive Visual Guidance for Demonstrated Rotational Degrees of Freedom in Mobile AR. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 28, 8 pages. doi:10.1145/3706598.3713179

[50] Zheng Wei, Hongtao Wu, lvmin Zhang, Xian Xu, Yefeng Zheng, Pan Hui, Maneesh Agrawala, Huamin Qu, and Anyi Rao. 2025. CineVision: An Interactive Pre-visualization Storyboard System for Director–Cinematographer Collaboration. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, Article 18, 18 pages. doi:10.1145/3746059.3747793

[51] Paul Wheeler. 2005. *Practical Cinematography* (2nd ed ed.). Elsevier/Focal Press. Citation Key: book:183517.

[52] Ke Xie, Hao Yang, Shengqiu Huang, Dani Lischinski, Marc Christie, Kai Xu, Minglun Gong, Daniel Cohen-Or, and Hui Huang. 2018. Creating and chaining camera moves for quadrotor videography. *ACM Trans. Graph.* 37, 4, Article 88 (jul 2018), 13 pages. doi:10.1145/3197517.3201284

[53] Ruyu Yan, Jiatian Sun, Longxiulin Deng, and Abe Davis. 2022. ReCapture: AR-Guided Time-lapse Photography. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 36, 14 pages. doi:10.1145/3526113.3545641

A Implementation Details

A.1 Geometric Constraints

To ensure geometrically valid shot plans, our system maintains proper spatial relationships between foreground subjects and background scenes using three constraints:

The size constraint maintains proper visual hierarchy by ensuring the foreground scale (s_f) remains smaller than the background scale (s_b) with a small margin (ϵ): $s_f \le s_b - \epsilon$.

The viewing angle constraint maintains realistic camera perspectives by limiting the maximum viewing angle (θ) based on the foreground scale and distance: $\tan^{-1}\left(\frac{s_f/2}{f_z}\right) \le \theta_{max}$.

The perspective constraint ensures the background remains fully visible through the foreground by calculating proper perspective lines from the camera position. The background boundaries satisfy:

$$\begin{aligned} b_x - s_b/2 &\ge c_x + m_l \cdot b_z \\ b_x + s_b/2 &\le c_x + m_r \cdot b_z \end{aligned} \tag{1}$$

where m_l and m_r are the slopes of the perspective lines.

A.2 Relative Keyframing Algorithms

After users design a shot plan, the system constructs a transformation tree (Algorithm 1) in which each node stores absolute 2D transformations for both the foreground and background layers, encoding their position and scale at each keyframe. In addition to storing these absolute transforms, the system computes and caches the relative transforms between adjacent nodes. For example, the foreground relationship is defined as $R^f_{\text{rel}} = (T^f_{\text{parent}})^{-1} \times T^f_{\text{current}}$, representing how each shot transitions to the next.

During capture, the system compares the planned foreground transformation with the tracked subject position. If the tracked position differs from the plan by more than a threshold, a new keyframe is inserted at the observed time and position.

Algorithm 1: Building a Transformation Tree with Relative Transformations

```
Input: K (a set of keyframes)
Output: nodes (the resulting transformation tree)
foreach k in K do
    $T_{fg}$ ← Transform2D(k.pos, k.fgZoom);
    $T_{bg}$ ← Transform2D(k.bgPos, k.bgZoom);
    node ← new TransformNode($T_{fg}$, $T_{bg}$, k.dur);
    nodes.append(node);
for $i$ ← 0 to |nodes| − 2 do
    nodes[i].children ← [nodes[i+1]];
    nodes[i+1].parent ← nodes[i];
foreach node in nodes where node.parent ≠ null do
    node.rel$T_{fg}$ ← $(\text{node.parent.}T_{fg})^{-1}$ × node.$T_{fg}$;
    node.rel$T_{bg}$ ← $(\text{node.parent.}T_{bg})^{-1}$ × node.$T_{bg}$;
return nodes;
```

Algorithm 2: Reconstruct Shot Plan from New Camera Positions

```
Input: tree, t (current time)
Output: newNode
(cur, accTime) ← FindNodeAtTime(tree, t);
progress ← (t − accTime)/cur.dur;
next ← cur.child[0] if exists, else cur;
fgPos ← Lerp(cur.Tf.pos, next.Tf.pos, $\Delta t$);
fgScale ← Lerp(cur.Tf.scale, next.Tf.scale, $\Delta t$);
bgPos ← Lerp(cur.Tb.pos, next.Tb.pos, $\Delta t$);
bgScale ← Lerp(cur.Tb.scale, next.Tb.scale, $\Delta t$);
relTransforms ← CaptureSubsequentTransforms(cur);
newNode ← TransformNode(;
   Transform2D(fgPos, fgScale),;
   Transform2D(bgPos, bgScale),;
   next.duration, t);
cur.dur ← t − accTime;
cur.child ← [newNode];
newNode.parent ← cur;
ReconstructNodeChain(newNode, relT);
return newNode;
```

The new node is inserted into the transformation tree at the corresponding time (Algorithm 2). The duration of the preceding node is truncated to end at the deviation point, and the system reconstructs all subsequent keyframes by applying the previously stored relative transforms. This preserves the intended spatial and temporal structure of the shot plan, allowing the output sequence to remain consistent with the user's original shot plan despite execution imperfections.

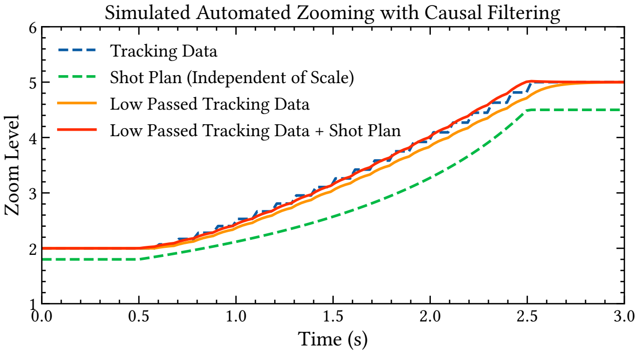


Figure 10: Simulated comparison of zoom control methods during automated camera movement. Noisy tracking data (blue dashed) represents zoom values derived from user position. The shot plan (orange dashed) shows the intended zoom trajectory. A low-pass filter (RC=0.1s) applied to tracking data alone (green solid) smooths noise but introduces temporal lag. Our hybrid approach (red solid) combines filtered tracking data with shot plan derivatives, achieving smooth zoom control that closely follows the intended trajectory while maintaining responsiveness to user movement.

B Real-Time Zoom Filtering

B.1 Tracking-Based Zoom Automation

During a dolly zoom, the foreground subject should remain at a consistent scale while the camera position and zoom parameters change simultaneously. To achieve this, the system computes the bounding-box area of the tracked subject and derives the zoom level required to match the target foreground scale defined in the shot plan.

B.2 Challenges in Automated Zoom Control

Implementing smooth automated zoom during handheld camera movements presents several technical challenges:

Shot Plan Limitations: Relying solely on predetermined shot plan zoom values is insufficient for complex shots like dolly zooms, which require precise synchronization between user movement and zoom speed. Because the required zoom rate increases exponentially with zoom level, any deviation from the planned trajectory—inevitable in handheld operation—breaks the dolly zoom effect.

Tracking Data Noise: Raw position tracking data, while responsive to user movement, suffer from limited refresh rates and sensor noise, producing jittery zoom behavior (blue line in Figure 10). These discontinuities make zoom adjustments visually apparent to viewers.

Filtering Latency: Standard causal filtering techniques (e.g., a single-pole IIR filter with $RC = 0.1$ s) effectively smooth tracking noise but introduce noticeable delay (green line in Figure 10), causing zoom changes to lag behind user movement.

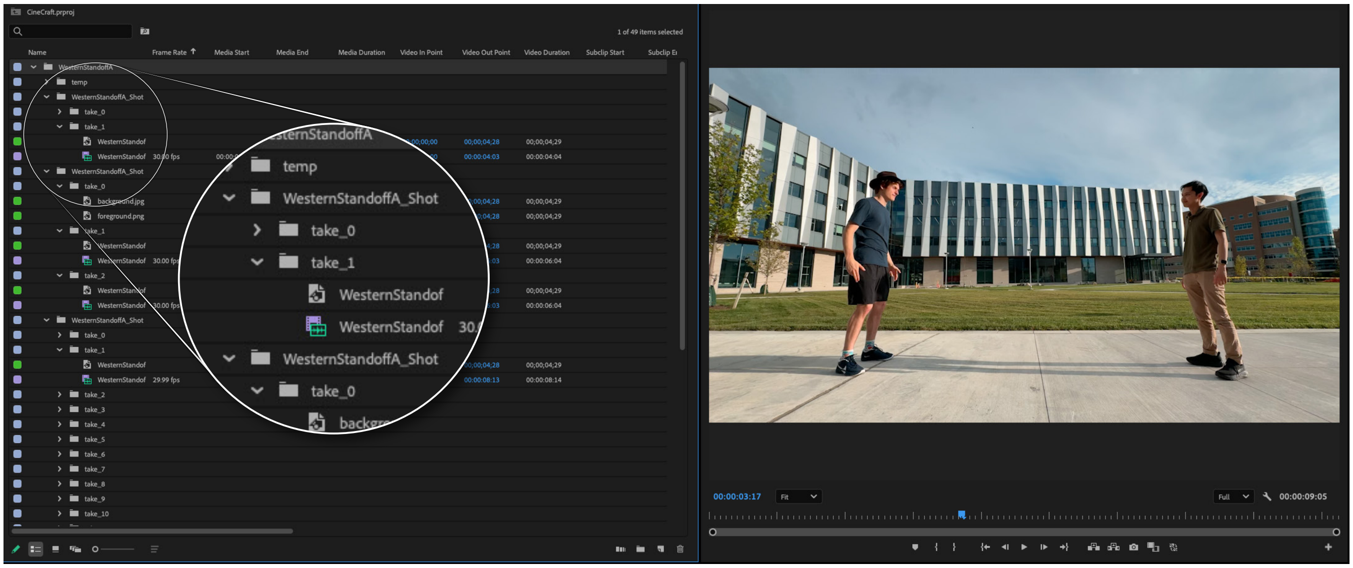
B.3 Hybrid Filtering Solution

We developed a hybrid approach that combines real-time filtered tracking data with shot plan information to achieve both smoothness and responsiveness. The method tracks frame-to-frame changes in the shot plan zoom trajectory and uses these derivatives to predictively update the low-pass filter state, eliminating the characteristic delay of causal filtering while preserving noise reduction.

As shown in Figure 10 (red line), this technique produces zoom control that closely follows the intended trajectory without sacrificing real-time responsiveness. The system maintains smooth zoom transitions even when the user deviates from the planned path, enabling successful execution of complex shots like dolly zooms during handheld operation. An additional benefit is graceful degradation: if tracking data momentarily drop out, zoom automation continues along the shot plan trajectory until tracking recovery restores the signal.

C Takes Management

CineCraft exports captures in a hierarchical folder structure compatible with standard Non-Linear Editors (NLEs), eliminating any conversion process. As shown in Figure 11, the system organizes content as: `Storyboard` → `Shot folders` → numbered `takes` with timestamps. This structure allows users to import their CineCraft captures directly into editing software like Premiere Pro, where they can refine shots beyond the initial rough-cut preview. Each take preserves its metadata, maintaining organizational context throughout the post-production workflow.



CineCraft.prproj

1 of 49 items selected

Name	Frame Rate	Media Start	Media End	Media Duration	Video In Point	Video Out Point	Video Duration	Subclip Start	Subclip Eı
WesternStandoffA									
temp									
WesternStandoffA_Shot									
take_0									
take_1									
WesternStandof					00;00;00	00;00;04;28	00;00;04;29		
WesternStandof	30.00 fps	00:00:0			00	00:00:04:03	00:00:04:04		
WesternStandoffA_Shot									
take_0									
background.jpg						;00;04;28	00;00;04;29		
foreground.png						;04;28	00;00;04;29		
take_1									
WesternStandof						4;28	00;00;04;29		
WesternStandof	30.00 fps					:03	00:00:06:04		
take_2									
WesternStandof						;28	00;00;04;29		
WesternStandof	30.00 fps					5:03	00:00:06:04		
WesternStandoffA_Shot									
take_0									
take_1									
WesternStandof						00;00;04;28	00;00;04;29		
WesternStandof	29.99 fps					00:00:08:13	00:00:08:14		
take_2									
take_3									
take_4									
take_5									
take_6									
take_7									
take_8									
take_9									
take_10									

Figure 11: Adobe Premiere Pro displaying CineCraft's folder structure: storyboards contain shot folders, which contain numbered takes, enabling direct import into video editing software. The sample output video is available at https://megatran.github.io/cinecraft.