The American Journal of Engineering and Technology
ISSN 2689-0984 | Open Access

TYPE Original Research
PAGE NO. 259-267
DOI 10.37547/tajet/Volume07Issue08-21

Check for updates

Adaptation of Cinematographic Techniques for Mobile and Vertical Video Formats

Herasymiuk Heorhii
Videographer Miami, USA

OPEN ACCESS

SUBMITED 25 July 2025
ACCEPTED 07 August 2025
PUBLISHED 21 August 2025
VOLUME Vol.07 Issue 08 2025

CITATION
Herasymiuk Heorhii. (2025). Adaptation of Cinematographic Techniques for Mobile and Vertical Video Formats. The American Journal of Engineering and Technology, 7(8), 259–267. https://doi.org/10.37547/tajet/Volume07Issue08-21

COPYRIGHT
© 2025 Original content from this work may be used under the terms of the creative commons attributes 4.0 License.

Abstract: The article describes how, in the context of the rapid growth of video consumption on mobile devices and the dominance of portrait screen orientation (9:16), classical cinematographic aspect ratios and techniques must be reconsidered and adapted. While traditional aspect ratios (4:3, 2.39:1) reflected historical technological and aesthetic constraints, the modern user experience dictates new rules of composition, blocking, editing, and technical implementation arising from the characteristics of touch controls and continuous scrolling. The study aims to identify and systematize the principles for applying classical cinematographic techniques in the mobile vertical format, to substantiate their effectiveness, and to develop a new grammar of visual storytelling for the narrow frame. The relevance of this work is driven by the unprecedented growth in mobile video traffic and changes in audience engagement models, which demand scholarly reflection on the aesthetic and technical transformations. Its novelty lies in the comprehensive integration of historical-technical analysis, ergonomic experiments, neurophysiological measurements, and content-analytic data to build a unified model of vertical video language. The methodological foundation combines comparative format analysis, ergonomic tests, a systematic review of recommendations on composition and camera movement, as well as empirical EEG-response studies and statistical analysis of audience behavior on TikTok and Reels. Results show that the central vertical axis of the frame serves as the path of least resistance for gaze and interaction; the rule of thirds transforms into vertical dynamics; and the Z-axis and dolly-in/out techniques become crucial for dramaturgy. Vertical split-screen significantly increases retention;

gyrostabilized POV reduces fatigue; light gradients and pinpoint color accents guide attention from top to bottom; and binaural and tactile sound expand perception of the narrow frame. The study's conclusions establish the theoretical and practical basis of a new cinematographic grammar for mobile vertical video: each modality—composition, camera movement, editing, lighting, color, sound, and interface effects—interacts to precisely direct attention within the 9:16 frame. This article will be particularly helpful to video production specialists, mobile app UX designers, and media aesthetics researchers.

Keywords: adaptation; cinematographic techniques; vertical format; mobile video; composition; editing; stabilization; lighting; color; sound

Introduction

Over the past five years, the mobile screen has ceased to be secondary and has become the primary media carrier. According to Vidico, 75% of all video views now occur on smartphones, which users hold vertically almost all the time, rather than rotating them for a horizontal frame [1]. The shift is also visible in everyday media diets: the British TouchPoints-2025 survey recorded that adults spend 3h 21min per day on their phones—for the first time exceeding time spent watching linear television (3 h 16 min)—and do so evenly throughout the day, whereas TV viewing still peaks in the evening [2].

Such a redistribution of attention is supported by network infrastructure: Ericsson estimates that by the end of 2024, video applications will have generated 111 exabytes of the 147 exabytes of monthly smartphone traffic, i.e., over three-quarters of the total data volume, with a six-year growth exceeding sevenfold [3]. In practical terms, this means that the energy consumption of mobile video streams alone over one month is comparable to London's five-month needs, underlining not only the cultural but also the infrastructural significance of this format.

Against this backdrop, the evolution of cinematographic aspect ratios takes on a new dimension. Early film stock and the 4:3 television standard—formally codified as the Academy ratio of 1.37:1 in 1932—provided visual continuity with the theatrical projection window and the technological limits of the first sound track [4]. In the 1950s, in response to the challenge posed by television, studios transitioned to anamorphic CinemaScope. They established the 2.39:1 wide frame, which remains one of the standard screen ratios in digital cinema distribution.

However, the logic of the wider, the more cinematic no longer applies in an environment dominated by portrait device orientation. Snapchat was the first to demonstrate this: as early as the mid-2010s, the company reported that vertical advertisements were watched to completion more often than horizontal ones, provided users were not required to rotate their screens [5].

With the advent of TikTok and short Reels, the vertical frame ceased to be merely a cropped version of a film. It became a new foundational 9:16 stock, dictating its own rules of composition, blocking, and pacing, and demanding scholarly reflection on the aesthetic and technological changes at the intersection of mobile UX and classical cinematography.

Materials and Methodology

This study is based on the analysis of 17 key sources, including academic articles, industry reports, empirical experiments, and user cases. The theoretical foundation comprises studies of consumer habits and infrastructural trends. The Vidico report [1] and the TouchPoints-2025 survey [2] revealed the dominance of vertical viewing and the even distribution of mobile consumption throughout the day. The DataReportal report [3] confirmed the rapid growth of video-app traffic. The historical-technical background of pre-cinematographic aspect ratios was provided by SMPTE standards [4], and practical conclusions on the effectiveness of the vertical format were drawn from the Snapchat case [5]. Compositional and ergonomic aspects were drawn from classic methodologies by Perry and Hourcade [6] and modern vertical framing recommendations by Parisi [7], supplemented with theoretical insights on split-screen engagement metrics by Clayton [8] and Spyrou [9].

Methodologically, the research combines several approaches. A comparative format analysis juxtaposed classical cinematic aspect ratios (4:3, 2.39:1) with the vertical 9:16, considering historical constraints and current user preferences [4, 5]. Ergonomic experiments—conducted following Perry and Hourcade [6] and Parisi [7]—demonstrated the path of least resistance along the vertical screen axis and the transformation of the rule of thirds in the vertical format. A systematic review of theoretical

recommendations encompassed studies on the influence of shooting angle on perceived trust (Baranowski & Hecht [10]), the effects of various camera-movement techniques (Yilmaz et al. [11]), and the algorithmic aspects of video stabilization (Luchetti et al. [13]).

Content and statistical analyses included: investigation of duet and reaction videos' impact on audience retention in TikTok/Reels (Spyrou [9]); analysis of the pocket-gimbal market—sales forecasts and CAGR through 2033 (Archive Market Research [12]); neurophysiological EEG measurements of principal components in experiments by Huttunen [14] and Mattavelli et al. [15] to assess emotional and cognitive responses to lighting schemes; and quantitative analysis of color accents in mobile feeds—click-through rates for warm palettes (Sokolik et al. [16]) and the U-shaped relationship between saturation and conversion (Wen et al. [17]). The integration of these methods provided a comprehensive picture of how the vertical format reconfigures composition, camera movement, and technical solutions in the mobile context.

Results and Discussion

Reorienting the frame from landscape to portrait has led to a reassembly of the classical compositional grid. Ergonomic studies of touch input show that, when holding the device with one hand, users tap elements located in the central band of the screen more quickly and accurately. In contrast, objects at the top and bottom edges require greater joint movement—this reinforces the value of the longitudinal axis as a path of least resistance for gaze and thumb interaction [6].

At the same time, the rule of thirds does not disappear; instead, it transforms: instead of the familiar displacement of the protagonist to the left or right intersection point, attention is distributed along the vertical axis. Creators of vertical videos recommend placing the dominant object within the middle third of the width, yet smoothly sliding it between the upper and lower horizontal dividers to engage the eye's natural scroll movement and preserve interface symmetry [7].

The limited horizontal span necessitates a reevaluation of spatial relationships within the frame: horizontal mise-en-scène gives way to depth staging. The vertical format more easily accommodates a full-length human figure, enabling storytelling through movement toward or away from the camera rather than lateral tracking shots. Theorists of vertical cinema note that it is the Z-axis, not the X-axis, that becomes the principal carrier of dramaturgy: the foreground jumps interactively toward the viewer, the midground sustains the action, and the background provides context—forming a three-layer column instead of a panoramic horizon [8].

The lack of width is compensated for by multi-level editing. The most popular technique is the vertical split-screen. TikTok's and Reels' algorithms rank duets and reaction videos higher because the simultaneous stimulation of two screen regions increases average watch time by ≈ approximately 40% and boosts information retention by up to 60%, thanks to concurrent visual and auditory channels [9]. However, cognitive experiments also record increased working-memory load. With an excess of parallel tracks, users reproduce the core message less accurately, demanding a strict hierarchy of layers and a contrasting rhythm of shot changes.

Thus, the central axis establishes a fulcrum for rapid perception, depth becomes the primary tool for plot construction, and the multi-level frame expands perceived space without widening the screen's boundaries. These three principles operate in concert and form the basis of a cinematographic grammar for the vertical 9:16 format.

When the vertical format fixes the gaze on a narrow axis, camera movement assumes the function of expanding space and redistributing semantic emphasis. Vertical pans and classic tilt-up/tilt-down movements prove especially effective, as they scan the frame along the same vector in which the viewer's finger scrolls the feed. Here, filmic language merges with gesture navigation: a camera rising from the floor to the subject's face shifts attention from the interface's comments zone to the protagonist. At the same time, the reverse movement prepares the viewer for the next swipe. Simultaneously, the vertical angle functions as a semantic filter: an experiment [10] demonstrated that even a 30° tilt of the camera from above or below alters trust ratings for the character; audiences ascribed maximum reliability to eye-level shots, whereas a high angle statistically reduced trust scores by 0.8 points on a nine-point scale, as shown in Fig. 1. Thus, tilt-up—as a marker of status elevation—and tilt-down—as a distancing effect—become formal antonyms within the same portrait geometry.

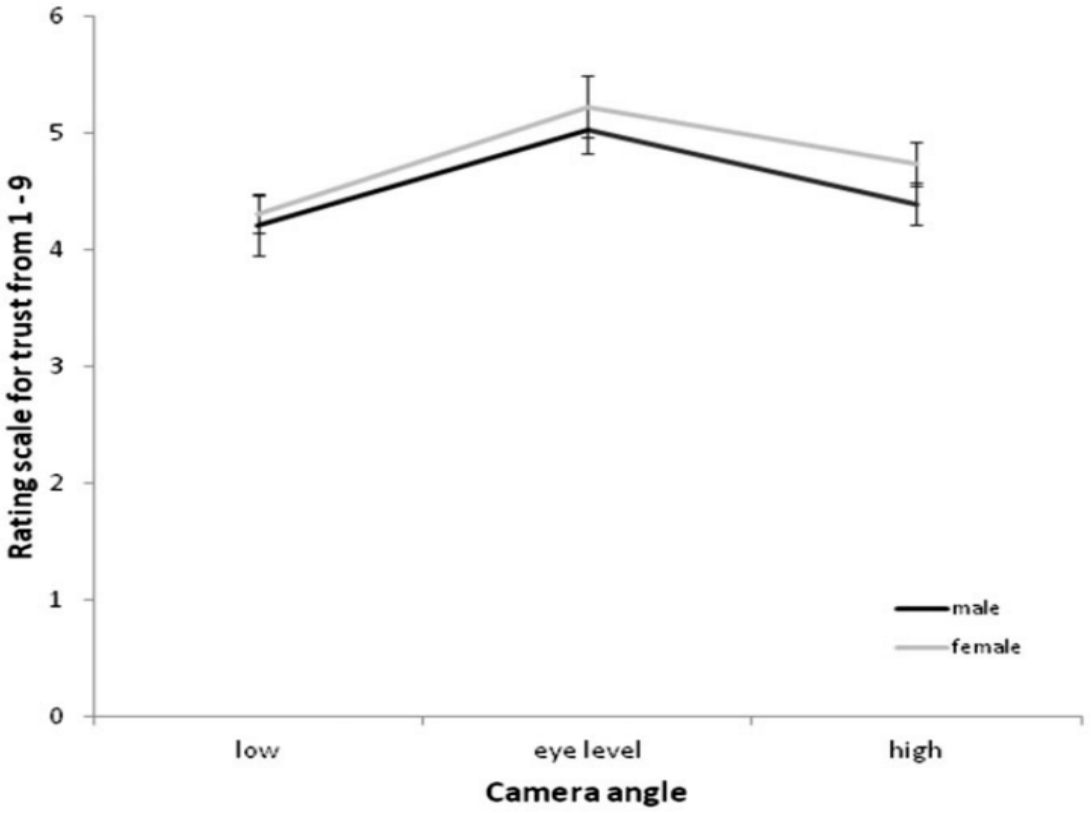

Fig. 1. Ratings for trust by camera angle and sex [10]

Depth movement along the Z-axis, which in horizontal cinema was supported by complex multi-plane mise-en-scène, is simplified in the vertical format to dolly-in and dolly-out. When shooting on a smartphone, these techniques visually draw the viewer closer to or farther from the event, allowing one to experience approach and withdrawal without lateral displacement. A neurocinematographic study with 44 respondents compared dolly, steadicam, handheld camera, and static shots. Smooth forward movement yielded a median immersion index higher than static shots, although emotional valence remained unchanged [11]. For the mobile viewer, this means that dolly-in can serve as a universal emotional accent. By minimizing editing, creators achieve significant engagement gains through pure kinetics, along the same vertical tunnel that houses the protagonist.

The third component of vertical cinematographic grammar is gyrostabilized POV. The small amplitude of micro-vibrations, barely perceptible on a cinema screen, becomes visual noise on a six-inch display, increasing the risk of acute fatigue and motion sickness. Consequently, demand for pocket three-axis stabilizers is growing faster than the mobile-video market itself; an industry report estimates global sales of such devices at USD 1.5 billion in 2025, with a projected CAGR of 15% through 2033, as illustrated in Fig. 2 [12].

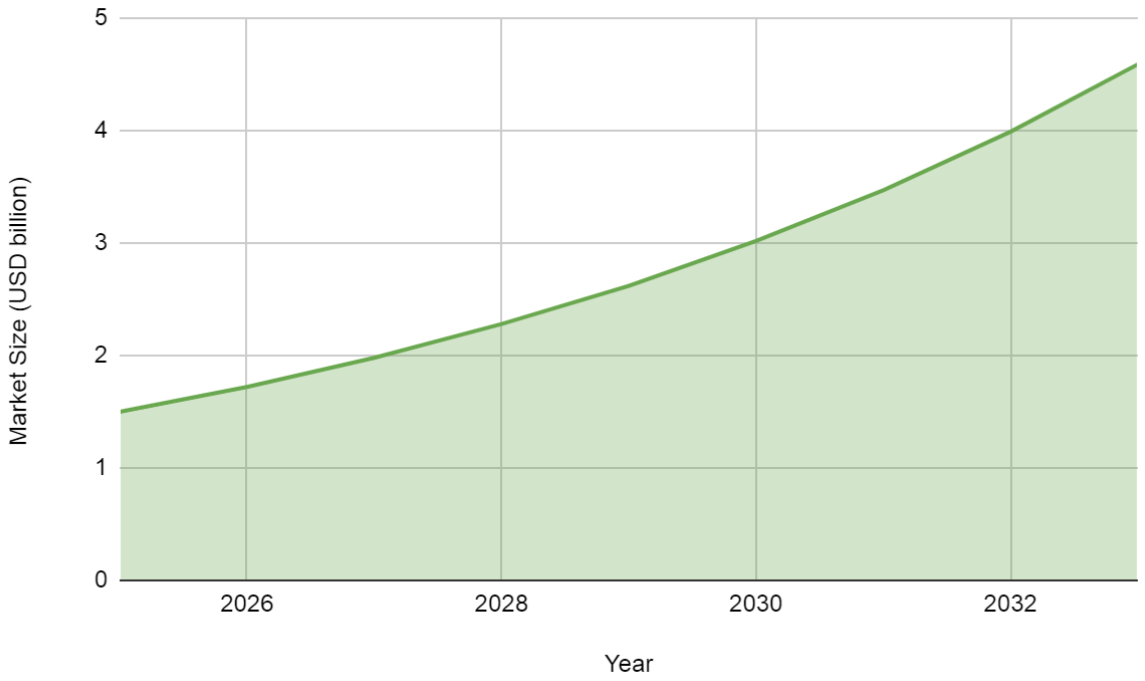

Fig. 2. Smartphone Gimbal Stabilizers Market Size [12]

The stabilization effect is measured not only commercially, but also in other ways. In a sample subjected to the standard Simulator Sickness Questionnaire, the subjective severity of symptoms after viewing the same clips, but with offline trajectory stabilization, was statistically lower than the baseline level [13]. Thus, gyrostabilized POV simultaneously enhances physiological comfort and preserves presence through slight correction of hand-held camera motion, making it the primary technical solution for mobile narrative.

Together, vertical pans, depth-axis dolly passes, and stabilized POV create a virtual volume in the narrow frame: the first establishes a hierarchy along the Y-axis, the second adds emotional rhythm along the Z-axis, and the third holds the viewer within this structure without kinetic noise. In subsequent sections, it will be shown how light, color, and sound amplify this kinetic matrix, turning the 9:16 constraint into a tool for precise direction of attention.

In the vertical frame, light becomes not only an exposure tool but also a compositional one: a gradient along the Y-axis replaces missing width. A classic Rembrandt-style technique—placing the key light high above the lens and slightly to one side, with a fill light below—creates a sliding vertical transition from a bright forehead to a darker torso. Such a gradient structures the scene much like the tilt-up discussed previously: the eye first reads the upper headline details, then naturally descends to the action while the background remains in mid-tone and does not compete with the app interface. Portrait photography has long demonstrated that an asymmetrical diagonal of light in the upper quarter of the frame renders the face both three-dimensional and clickable, and requires only one lamp and reflector, critical for on-the-move mobile shooting.

Neurophysiological data reinforce this technological logic. In a pilot EEG experiment where faces were lit from above, below, and in silhouette, underlighting that slipped beneath the lower-panel icons—and complete eye-region darkness—elicited a more negative early posterior negativity than the classic 45° top key, interpreted as a subconscious vigilance signal [14], as illustrated in Fig. 3.

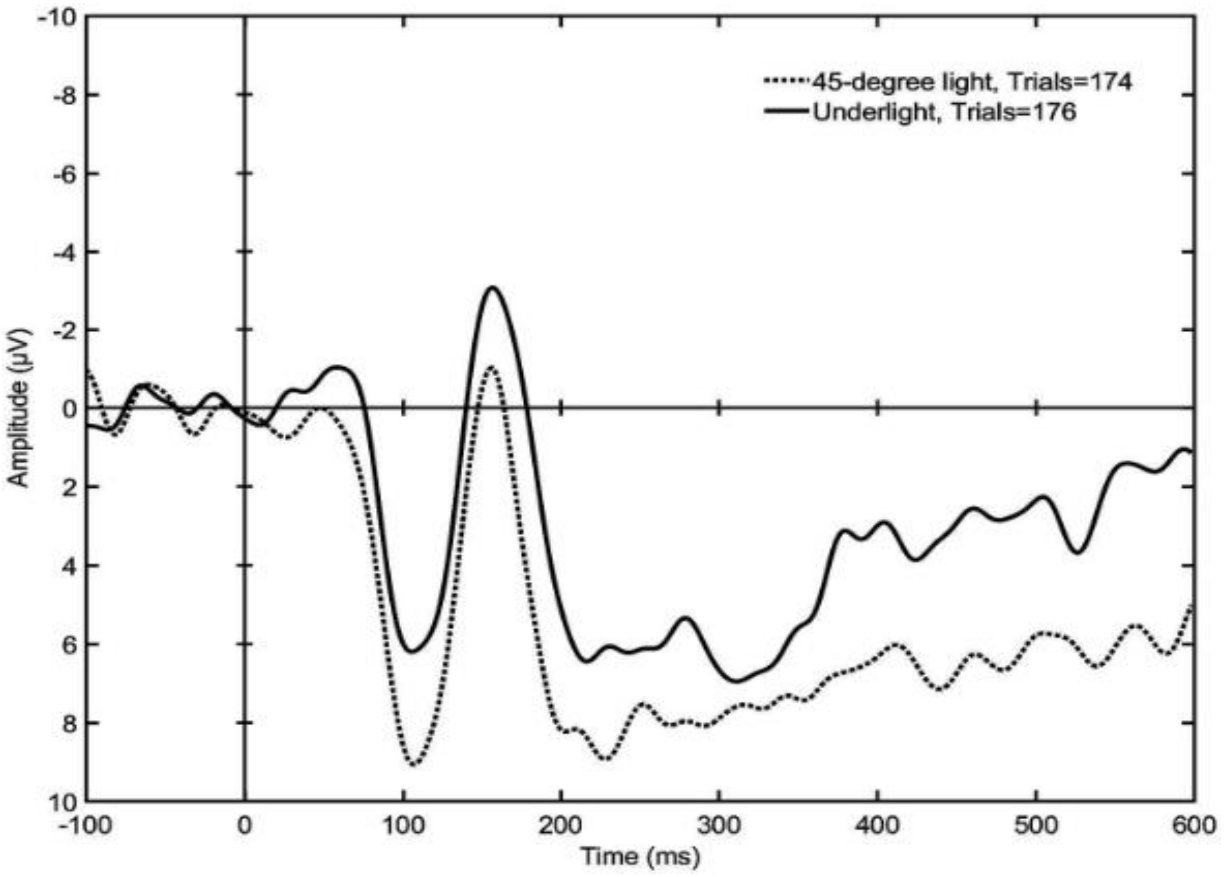

Fig. 3. The mean signal amplitudes elicited by underlight and 45-degree light [14]

The graph shows a noticeable difference in the mean negativity within both the 150–300 and 200–600 ms after-stimulus intervals, although the latter did not reach statistical significance. This suggests that lighting that originates from below the face, distorting the facial features, elicits a stronger early emotional response than lighting that reveals the entire face.

In other words, an error in vertical-contrast distribution is instantaneously reflected in the viewer's emotional baseline before the narrative even begins. A psychophysical study [15] confirmed the practical finding: with a central gaze model, trust indices statistically decreased under a split-light condition (a harsh horizontal shadow), whereas a moderate top-down Rembrandt gradient increased perceptions of trustworthiness and attractiveness, especially in younger faces.

The vertical gradient also applies to color. In an endless scroll, attention is captured by small spectral bursts—localized but not aggressive accents. A field sample of 1.5 million impressions showed that a warm red palette in a narrow vertical box generated clicks significantly more often than a cool blue one, confirming that a warm hue shouts from a monotone feed [16]. A more recent machine analysis of 2,984 short videos revealed a U-shaped relationship: moderate or very high saturation increased subsequent purchases, whereas mid-range values were lost against neighboring clips [17]. Thus, for the mobile-vertical format, the point-color rule applies: a bright accent in one third of the screen amplifies the dopaminergic hook in the first milliseconds, but excessive colorfulness disintegrates into noise and reduces retention.

Altogether, a top-to-bottom light gradient, calibrated facial contrast, and measured color guide the viewer's gaze along the same path as the camera movement described earlier: from above down to the focal point and then toward the interface zone. This transforms the 9:16 limitation into a multi-layered channel of directed sensory information, where lighting not only paints the subject but also synchronizes the viewer's physiological rhythm with the edit pace of the clip.

Sound in vertical video functions as an invisible extension of the frame, filling what is bounded by the narrow screen. Spatial, binaural, recording creates the illusion of volume: the viewer perceives footsteps behind or a voice to the side, even though only a portrait plane is visible. Through such acoustic completion, the director is freed from constantly changing camera angles, and the audience receives a convincing three-

dimensional impression even with a static camera. The smooth movement of the sound source around the head directs the movement of attention, preparing the viewer for the appearance of an object in the center of the frame.

The vertical experience is amplified not only auditorily but also somatically. Low frequencies, transmitted through the speaker or via vibration of the phone's body, create a tactile background that underscores the scale or emotional weight of a scene. A brief bass impulse reinforces the close-up of the protagonist, while prolonged sub-bass masks location changes by smoothing abrupt edits. The user perceives a light pulsing pressure in the palm, transforming the spectacle into a kinesthetic event.

Editing in the mobile context relies on the speed of perception. Retention is aided by a sharp hook: the opening moments present the primary visual stimulus, followed by a rapid sequence of short shots. The rhythm is calibrated so that each shot delivers a fresh piece of information but leaves no room for the viewer to skip to the following clip. Whereas traditional editing revolved around the complete thought of a scene, here it is oriented toward micro-transitions that integrate into the endless feed and synchronize with the scroll gesture.

Text elements adhere to the same principle of time economy. Titles do not appear as separate cards but as animated typography embedded in the trajectory of the camera or finger movement. Letters grow from the horizon line, assemble, and disappear at the moment the visual emphasis shifts. Such living text does not hijack attention for long nor obscure the face, yet it registers in memory as part of the unified motion.

An additional layer of engagement is created by augmented reality. Filters that respond to facial expressions or gestures connect the narrative to the viewer: a mask that mirrors one's expression or an interactive object caught by a swipe transform passive viewing into a personal event. A similar effect is achieved by the comic-panel composition, where the main vertical frame remains fixed while auxiliary layers scroll independently, revealing alternative angles or plot details.

Together, spatial sound, tactile bass, accelerated editing, kinetic typography, and the AR layer form a synesthetic system in which each channel supports the screen's confined geometry. The light gradients described earlier map the visual route from top to bottom; audio markers and vibration anchor the stages of this route; visual effects hold attention on key points. Thus, the vertical frame becomes not a limitation but a platform where all modalities work in concert to direct the viewer to the center of the story.

Conclusion

The conducted study demonstrates that the shift to the mobile vertical 9:16 video format has triggered a comprehensive set of aesthetic and technical transformations of classical cinematographic language. Within the narrow frame, the central longitudinal axis serves as the path of least resistance for gaze and interface interaction, defining a new compositional paradigm. Instead of the habitual horizontal distribution of emphasis, we observe a vertical zone of interest in which positioning the dominant object within the middle third of the width and dynamically sliding it along the frame's height harmonize with the user's natural scroll, facilitating perception and enhancing attention retention.

The depth Z-axis becomes the primary dramaturgical tool: dolly-in and dolly-out replace complex multi-plane compositions, creating the effect of pushing the object toward the viewer or detaching it from the background. Simultaneously, multi-level editing—such as vertical split-screen—expands the narrow frame's perceived space without losing image coherence, stimulating visual and auditory perception through parallel information streams.

Technical solutions, such as gyrostabilized POV, ensure the necessary smoothness of movement despite the limitations of handheld smartphone shooting, thereby minimizing visual noise and reducing the risk of fatigue. The combination of vertical pans, axial movements, and stabilization forms a virtual volume within the portrait frame, in which lighting, color, and sound perform not only utilitarian but also compositional functions. Directed light gradients structure perception from top to bottom; point-color accents shout from the endless feed; and binaural and tactile components complement the visual track, creating a synesthetic system of interaction.

Thus, the adaptation of classical cinematographic techniques to the 9:16 format goes beyond a mere rotation of the frame and constructs a new grammar of mobile narrative, in which each modality—from

composition and editing to lighting, color, sound, and interface effects—acts in concert, turning the vertical screen constraint into a tool for precise attention direction. These conclusions provide the theoretical and practical foundation for the further development of vertical film language and the optimization of user experience in mobile video applications.

References

1. "20 Amazing Video Consumption Trends To Watch Out For," Vidico, Aug. 29, 2024. https://vidico.com/news/video-consumption-trends/ (accessed May 26, 2025).
2. M. Sweney, "Adults in Great Britain now spending more time on mobiles than watching TV," The Guardian, 2025. https://www.theguardian.com/media/2025/jun/25/adults-great-britain-time-mobiles-watching-tv-screen-ipa-survey (accessed May 27, 2025).
3. "Digital 2025: Mobile Data Consumption Trends," DataReportal, Feb. 05, 2025. https://datareportal.com/reports/digital-2025-sub-section-mobile-data-consumption (accessed May 28, 2025).
4. "SMPTE ST 430-16:2017," SMPTE, 2017. https://pub.smpte.org/doc/st430-16/20170714-pub/st0430-16-2017.pdf (accessed May 29, 2025).
5. S. Fatah, "Snapchat launches new program, highlights impact of vertical video," Media in Canada, 2016. https://mediaincanada.com/2016/06/14/snapchat-launches-new-program-highlights-impact-of-vertical-video/ (accessed May 30, 2025).
6. K. B. Perry and J. P. Hourcade, "Evaluating one-handed thumb tapping on mobile touchscreen devices," Graphics Interface, pp. 57–64, May 2008, Accessed: Jun. 26, 2025. [Online]. Available: https://www.researchgate.net/publication/221474717_Evaluating_one_handed_thumb_tapping_on_mobile_touchscreen_devices
7. D. Parisi, "Vertical Video for Social Media: Best Practices for Reels in 2025," LocalEyes, Apr. 08, 2025. https://localeyesit.com/vertical-video-for-social-media/ (accessed Jun. 01, 2025).
8. R. Clayton, "This is a repository copy of Filmmaking Theory for Vertical Video Production," University of Leeds, 2019. Accessed: Jun. 01, 2025. [Online]. Available: https://eprints.whiterose.ac.uk/id/eprint/180677/1/EuroMedia2019_52556.pdf
9. Stéphanos Spyrou, "Discover how split-screen videos on TikTok are doubling engagement rates," LinkedIn, Dec. 03, 2024. https://www.linkedin.com/pulse/split-screen-phenomenon-double-content-distraction-st%C3%A9phanos-spyrou-bbg3e (accessed Jun. 02, 2025).
10. A. M. Baranowski and H. Hecht, "Effect of Camera Angle on Perception of Trust and Attractiveness," Empirical Studies of the Arts, vol. 36, no. 1, pp. 90–100, May 2017. https://experimental.psychologie.uni-mainz.de/files/2024/06/2018BaranowskiHecht_Camera_Angle.pdf (accessed Jun. 02, 2025).
11. M. B. Yilmaz, E. Lotman, A. Karjus, and P. Tikka, "An embodiment of the cinematographer: emotional and perceptual responses to different camera movement techniques," ProQuest, vol. 17, Jul. 2023, doi: https://doi.org/10.3389/fnins.2023.1160843.
12. "Smartphone Gimbal Stabilizers Market Size," Archive market research, 2024. https://www.archivemarketresearch.com/reports/smartphone-gimbal-stabilizer-260840#summary (accessed Jun. 05, 2025).
13. A. Luchetti, M. Zanetti, D. Kalkofen, and M. D. Cecco, "Stabilization of spherical videos based on feature uncertainty," The Visual Computer, vol. 39, no. 9, pp. 4103–4116, Jul. 2022, doi: https://doi.org/10.1007/s00371-022-02578-z.
14. S. Huttunen, "Faces in shadows: silhouette light, underlight and toplight elicit increased early posterior negativity," Frontiers in Neuroscience, vol. 19, Mar. 2025, doi: https://doi.org/10.3389/fnins.2025.1553977.
15. G. Mattavelli, I. Gorrino, E. Cesana, J. D. Angelis, and P. Ricciardelli, "Illumination and gaze effects on face evaluation: The Bi-AGI database," Frontiers in Psychology, vol. 13, Oct. 2022, doi: https://doi.org/10.3389/fpsyg.2022.948142.
16. K. Sokolik, R. G. Magee, and J. D. Ivory, "Red-Hot and Ice-Cold Web Ads: The Influence of Web Ads' Warm and Cool Colors on Click-Through Rates," Journal of Interactive Advertising, vol. 14, no. 1, pp. 31–37, Jan.

2014, doi: https://doi.org/10.1080/15252019.2014.907757.

17. X. Wen, Y. Niu, and C. Liao, "U-shaped Effect of Color in Short-form Video: Combination of Text, Audio, Image ML and Generative AI," Journal of Business Research, vol. 192, pp. 115278–115278, Mar. 2025, doi: https://doi.org/10.1016/j.jbusres.2025.115278.