Moiré Video Authentication: A Physical Signature Against AI Video Generation

Yuan Qing* Kunyu Zheng* Lingxiao Li* Boqing Gong Chang Xiao
Boston University
https://yuanqing-ai.github.io/llm-hierarchy/

Abstract

Recent advances in video generation have made AI-synthesized content increasingly difficult to distinguish from real footage. We propose a physics-based authentication signature that real cameras produce naturally, but that generative models cannot faithfully reproduce. Our approach exploits the Moiré effect: the interference fringes formed when a camera views a compact two-layer grating structure. We derive the Moiré motion invariant*, showing that fringe phase and grating image displacement are linearly coupled by optical geometry, independent of viewing distance and grating structure. A verifier extracts both signals from video and tests their correlation. We validate the invariant on both real-captured and AI-generated videos from multiple state-of-the-art generators, and find that real and AI-generated videos produce significantly different correlation signatures, suggesting a robust means of differentiating them. Our work demonstrates that deterministic optical phenomena can serve as physically grounded, verifiable signatures against AI-generated video.*

1. Introduction

In February 2024, a finance worker at a multinational firm was tricked into transferring $25 million after attending a video call in which every other participant, including the company's chief financial officer, were impersonated by AI-generated video [8]. In October 2025, an AI-generated video mimicking a news broadcast falsely depicted an Irish presidential candidate announcing her withdrawal from the race, spreading across social media for hours before being removed [2]. Spreading unauthenticated misinformation by generative AI is no longer a hypothetical scenario; it is the new reality. Generative video models such as Sora [5], Runway [32], and Seedance [6] now produce footage so convincing that human viewers, automated detectors, and even forensic analysts struggle to distinguish it from real recordings [10]. As AI-generated video increasingly spreads misinformation and erodes public trust, the need for a reliable, theoretically grounded mechanism to prove that a video is *authentic* has become more urgent than ever.

Figure 1. (a) The Moiré effect, created by overlaying two repetitive layers. (b) Our prototype Moiré signature assembly. Viewing the assembly from slightly different camera angles (left vs. right) causes the fringes to shift in phase according to deterministic physical laws. (c) In a real video, Moiré fringes appear naturally and shift predictably with camera movement. (d) In an AI-generated video, the fringes distort, and their phase shifts do not adhere to the underlying physics.

Existing AI video detection broadly fall into two categories. *Post-hoc* forensic detectors analyze pixel-level artifacts, temporal inconsistencies, or learned statistical fingerprints to flag AI-generated content [17, 31, 39]. While effective against earlier generators, these methods are locked in an arms race: each new model eliminates the artifacts that detectors exploit, demanding constant retraining and offering limited lasting guarantees. *Digital watermarking* approaches, such as C2PA metadata [9] and SynthID [15], embed provenance information into media at creation time. However, metadata can be stripped, re-encoded, or forged, and digital watermarks degrade under common post-processing operations. Crucially, neither paradigm provides an unforgeable physical link between the recorded scene and the resulting video.

We propose a fundamentally different approach: a **physical authentication signature**. A real video is produced by an optical system governed by physical laws, whereas generative models learn statistical correlations from data and rarely model the underlying physics. A physical phenomenon whose appearance is tightly coupled to precise optical geometry therefore serves as a signature that real cameras produce naturally but that current video generative

*Equal contribution.

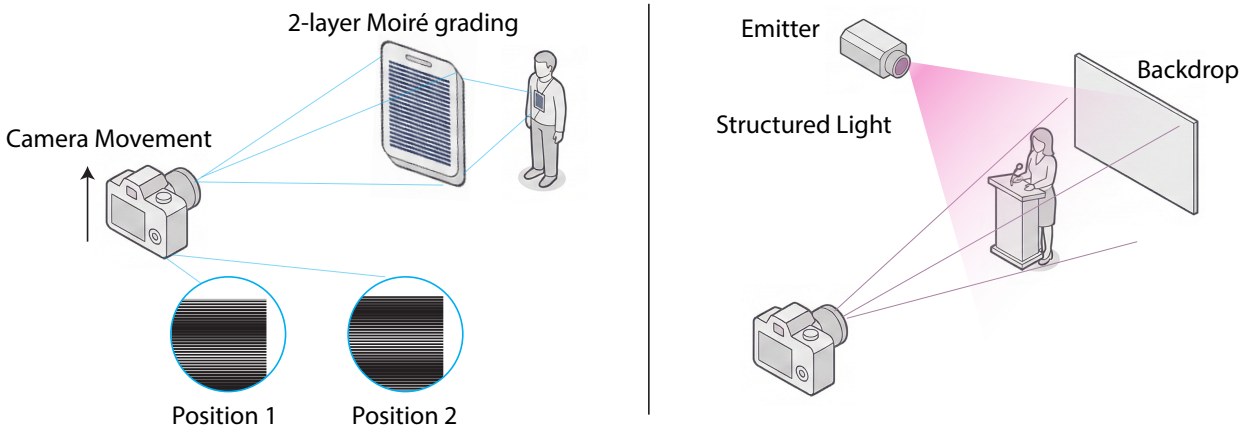

Figure 2. (Left) Our proposed Moiré-based signature requires only a passive 2-layer Moiré structure (e.g., worn as a badge), where standard camera movement naturally captures phase shifts. (Right) Active structured light signatures [29, 33], while also using physical signals for authentication, require specialized external emitter hardware to project patterns onto the scene.

models cannot precisely reproduce.

We instantiate this idea through the Moiré effect [1], an intriguing and widely observed yet often overlooked optical phenomenon. Moiré effect are the interference fringes that appear when two fine periodic structures, such as line gratings, are overlaid (Fig. 1-a). These fringes are highly sensitive to the precise geometric relationship between the camera and the grating assembly: as the camera moves, the fringes shift according to deterministic optical laws (Fig. 1-b). This tight coupling between camera motion and fringe motion forms a physics-based signature that real optical systems produce naturally (Fig. 1-c) but that current video generation models cannot reliably synthesize (Fig. 1-d).

The core assumption underlying our approach is that current video generation models are not grounded in physical simulation. They learn to produce plausible-looking frames from large-scale data, but they do not model the wave-optics interactions that govern Moiré fringe formation and evolution. Accurately reproducing the effect would require solving the exact optical problem for every frame: tiny shifts in camera position produce measurable fringe displacement, and the mapping from camera motion to fringe motion follows a strict mathematical relationship derived from first principles [1]. This stands in sharp contrast to the statistical interpolation that current generative models perform. Even if future models were trained on abundant footage containing Moiré patterns, learning the visual appearance of fringes can hardly, if not impossibly, equip a model to solve the underlying optical equations.

Leveraging this insight, we propose a practical authentication system that works as follows. A compact grating assembly, consisting of a lenticular sheet mounted above a printed stripe pattern, is placed in the scene (*e.g.*, worn by a speaker at a press conference, see Fig. 2). Given a recorded video with relative movement between the assembly and the camera (which occurs naturally from the speaker's body motion even when the camera is static) our algorithm (1) extracts the Moiré fringe pattern and maps it to a canonical coordinate frame, (2) tracks the fringe phase changes over time, and (3) computes the correlation between the observed fringe phase and the displacement of the assembly in the video. A high correlation certifies authenticity; a low or absent correlation flags the video as potentially synthetic. Importantly, analogous to digital watermarking, our method provides positive authentication for videos that contain the signature; it does not make claims about videos lacking one.

This paper contributes a novel idea supported by theoretical grounding and empirical evidence for validating the feasibility. Our contributions are:

- **A novel Moiré-based video authentication framework.** We introduce the concept of using the Moiré effect as a passive, physics-based authentication signature. We derive the Moiré motion invariant: fringe phase displacement and grating image displacement are linearly coupled; the resulting correlation is independent of viewing position and grating structure, yielding a verifiable signal extractable from video alone.
- **A practical implementation and feasibility validation.** We build a proof-of-concept system using off-the-shelf materials (a lenticular sheet, printed patterns, and ArUco markers) and demonstrate that the physics-predicted correlation holds in real captured video, confirming the practical viability of the approach.
- **A threat model analysis.** We systematically examine how our method performs against a spectrum of attack scenarios, characterizing the conditions under which the

Moiré signature remains robust. We also synthesize a large number of AI-generated videos in our best effort to reproduce the Moiré pattern, and the results show that even cherry-picked and carefully engineered generated videos fail to achieve correlation comparable to that of real video.

2. Related Work

DeepFake and AI-generated video. Our work is broadly motivated by the hyper-realistic generative media that calls for reliable systems to distinguish synthetic video from real footage; DeepFake [10, 31] enables the convenient creation and dissemination of misinformation. Unlike authenticity techniques, afterthought defenses [22, 25, 26, 42, 44, 46] are in an arms race with the rapid development of generative models [5, 11, 16, 24, 28].

Physics-based video forensics. A separate line of work detects AI-generated video by identifying violations of physical plausibility, such as irregular eye pupil geometry [18], inconsistent lighting and shadow directions [13], or semantic and anatomical anomalies like extra fingers [38]. These methods exploit incidental physical errors that current generators happen to make. However, as generators improve, these artifacts are progressively eliminated through better training data and architectures. Our approach sidesteps this limitation: rather than searching for the absence of artifacts, we introduce a deliberate physical structure whose optical behavior is governed by deterministic laws that generators cannot learn to replicate statistically.

The work most closely related to ours are methods that actively project temporally modulated light at the recording site to embed a verifiable physical signature into video, such as VeriLight [33] and Michael et al. [29]. Our approach is similar in spirit, grounding authentication in physical-world signals, but differs fundamentally in that the Moiré signature is entirely passive: it arises naturally from the optical geometry of a compact grating structure and ordinary camera motion, requiring no active illumination or dedicated infrastructure (Fig. 2). A complementary line of defense involves digital watermarking [21, 43, 47], which embeds provenance information at the software level.

3. The Moiré Effect

A Moiré pattern is a large-scale interference pattern that emerges when two fine periodic structures are superimposed with a slight difference in their spatial frequencies [1]. Moiré patterns appear ubiquitously in everyday life, from the shimmering bands seen through overlapping fences to the rippling fringes visible when a digital display is photographed at certain positions, and they have found broad application in optical metrology [7, 14, 41], mechanical deformation analysis [1, 23, 34, 37], and computational art [20, 35].

In our system, the Moiré pattern is generated by a compact structure consisting of two layers of line gratings with slightly different spatial periods, separated by a thin gap of thickness g. Because the gap causes each viewing position to reveal a different relative alignment of the two layers, the superposition produces prominent Moiré fringes whose positions are extremely sensitive to camera-grating geometry. The fringes act as a *geometric amplifier*: a microscopic change in viewing position produces a macroscopic fringe shift, yielding a visually prominent response that is easily captured on video yet difficult for generative models to synthesize correctly.

Without loss of generality, we restrict our analysis to **one-dimensional Moiré patterns** (parallel line fringes), which is the configuration used in our implementation. However, two-dimensional Moiré patterns formed by crossed or rotated gratings [1, 41] can be easily implemented and extended in the same principle. As a feasibility study, in this paper we only consider the 1D scenario.

3.1. Theory

We illustrate the relationship between camera motion and Moiré fringe displacement for the one-dimensional two-layer grating structure. Consider two aligned gratings with spatial periods p_f (front) and p_r (rear). From standard Moiré theory [1], their superposition produces fringes with beat period $p_m = p_f p_r / |p_f - p_r|$. A lateral relative shift δ between the layers displaces the fringes by $|\Delta x_m| = M|\delta|$, where $M = p_f/|p_f - p_r|$ is the **Moiré magnification factor**.

In our structure the layers are rigidly fixed; the apparent shift δ arises from viewing-angle parallax. We analyze motion along the axis x perpendicular to the grating lines (the only direction that produces fringe shifts). A camera at distance D undergoing a lateral displacement Δx_c sees an apparent layer shift $\delta = g\,\Delta x_c / D$. Converting fringe displacement to **unwrapped phase** ($\Delta\phi = 2\pi\Delta x_m / p_m = 2\pi\delta/p_r$, where the p_f and $|p_f - p_r|$ terms cancel) and substituting δ yields

$$\Delta\phi = \pm\frac{2\pi g}{p_r \cdot D} \cdot \left(1 + \frac{g}{D}\right) \cdot \Delta x_c, \tag{1}$$

where the sign depends on coordinate convention and the factor $(1 + g/D)$ accounts for the rear grating's projected period on the front grating plane.

3.2. Our Insight: The Correlation Invariant

The preceding relationship shows that Moiré fringe displacement depends on camera motion. We now show that this same quantity governs the apparent motion of the grating structure itself in the camera image, yielding a correlation invariant that is independent of the unknown distance

D. We first derive the invariant for pure camera translation, then generalize to arbitrary camera motion that includes rotation.

In our physical setting, the grating gap g is on the order of millimeters while the observation distance $D \geq 0.5\,\text{m}$, so the factor $(1 + g/D)$ in Eq. (1) is negligible. We therefore approximate Eq. (1) as

$$\Delta\phi = \pm\frac{2\pi g}{p_r \cdot D} \cdot \Delta x_c. \tag{2}$$

Pure translation. Under a pinhole camera model with focal length f, a pure translational camera displacement Δx_c along the perpendicular axis at distance D from the grating structure causes the structure's centroid to shift in the image plane by

$$\Delta u_{\text{trans}} = \pm f \cdot \frac{\Delta x_c}{D},$$

where the sign depends on coordinate convention. Combining this with Eq. (2), we eliminate the common factor $\Delta x_c / D$ to obtain

$$\Delta\phi = \pm\frac{2\pi g}{p_r \cdot f} \cdot \Delta u_{\text{trans}}. \tag{3}$$

General camera motion. In practice, camera motion is a combination of translation and rotation. Under the pinhole model, a small camera rotation $\Delta\theta$ about the axis parallel to the grating lines (*i.e.* pan or tilt that displaces the structure image along the perpendicular axis x) produces an additional image displacement Δu_{rot}. The total observed image displacement of the structure is therefore

$$\Delta u = \Delta u_{\text{trans}} + \Delta u_{\text{rot}}.$$

Critically, rotation does *not* induce parallax between the two grating layers. A pure rotation shifts both layers identically in the image, producing zero apparent relative displacement ($\delta = 0$) and hence zero fringe phase change. Only the translational component Δu_{trans} generates parallax and drives $\Delta\phi$. Therefore Eq. (3) holds in the general case, with $\Delta u_{\text{trans}} = \Delta u - \Delta u_{\text{rot}}$.

For an authentic video, the temporal sequences of fringe phase changes $\{\Delta\phi_t\}$ and translational image displacements $\{\Delta u_{\text{trans},t}\}$ are linearly coupled with near-unity absolute correlation:

$$|\rho| = \left|\text{Corr}\big(\{\Delta\phi_t\},\ \{\Delta u_{\text{trans},t}\}\big)\right| \approx 1. \tag{4}$$

We call this the **Moiré motion invariant**. The proportionality constant need not be known to the verifier: D has cancelled entirely, and although f, g, and p_r may be unknown, they are fixed for a given camera and grating structure. Because correlation measures linear coupling regardless of the slope's value, a verifier simply extracts both sequences from the pixel data and checks whether they are linearly coupled, without requiring any calibration or knowledge of the scene geometry.

In general, isolating Δu_{trans} from the observed image displacement Δu requires estimating and subtracting the rotational component Δu_{rot}, which can be done purely from the video, for instance by tracking distant background features or decomposing camera pose via multi-point correspondence methods such as Perspective-n-Point (PnP) pose estimation [45]. In our implementation, we perform this rotation compensation using the four ArUco fiducial markers already present on the grating assembly as 3D-to-2D correspondences for PnP-based camera pose recovery, allowing us to isolate the translational displacement component. We describe our concrete implementation in Sec. 4.

In an AI-generated video, no physical grating structure exists; the generator must synthesize fringe patterns frame by frame from learned statistics. Because the generator does not solve the optical equation, the synthesized fringe motion will not maintain the strict linear coupling to the structure's position required by the invariant. Even small deviations, a fringe that shifts too fast, too slow, or in the wrong direction by a fraction of a cycle, will depress $|\rho|$ far below 1. This theoretical gap between physics-governed and statistically generated fringe behavior is the foundation of our authentication algorithm.

4. Proof-of-Concept Implementation

We now describe our proof-of-concept system for validating the Moiré motion invariant. We first detail the physical grating assembly, then present the algorithmic pipeline that extracts both the fringe phase signal and the grating assembly displacement signal from a recorded video and computes their correlation.

4.1. Physical Setup

Our grating assembly consists of three layers bonded together into a flat, rigid structure (see Fig. 3-a):

- **Front layer (viewing grating).** A 1 mm-thick lenticular lens sheet (50 LPI, period $p_f = 0.508\,\text{mm}$) whose cylindrical lenses act as angular apertures, so that small lateral camera movements change the visible portion of the rear layer and shift the Moiré fringes.
- **Rear layer (reference grating).** A line pattern printed on paper with pitch $p_r = 0.52\,\text{mm}$. The slight mismatch between p_f and p_r produces Moiré fringes with a beat period of $p_m = p_f \cdot p_r / |p_f - p_r| \approx 22\,\text{mm}$, yielding a Moiré magnification factor of $M = p_f / |p_f - p_r| \approx 42\times$.
- **Base layer.** A flat acrylic sheet that serves as the structural substrate, holding all layers in alignment.

The three layers are assembled and attached to a flat backing surface. For ease of prototyping, we place four

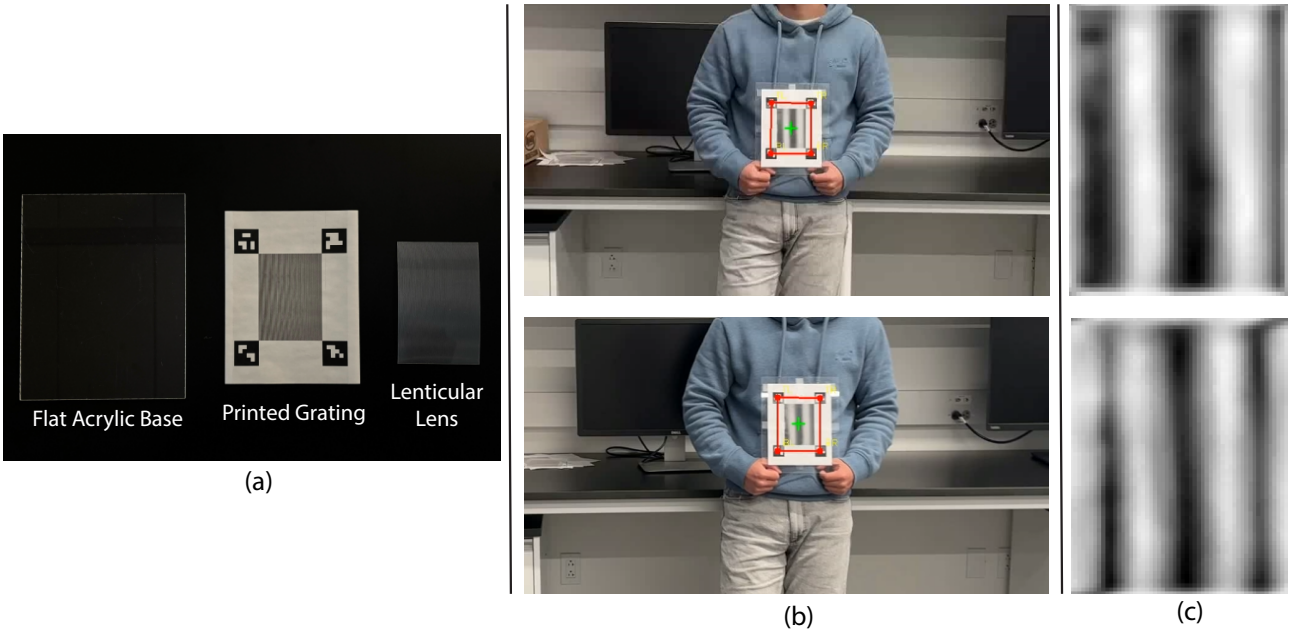

Figure 3. (a) Exploded view of the grating assembly, showing its three constituent layers: a flat acrylic base, a printed grating, and a lenticular lens. (b) Two distinct video frames illustrating our algorithm detecting the ArUco markers to isolate the Moiré fringe region. (c) The extracted Moiré fringes after a canonical transformation, corresponding to the frames in (b). The phase shift of the fringes between the two frames is clearly visible.

ArUco fiducial markers (from the `DICT_4X4_50` dictionary) at the corners of the grating assembly to simplify bounding-box extraction and perspective correction. Any robust localization method could replace them in a production system.

4.2. Verification Pipeline

Given an input video containing the grating assembly, our pipeline extracts two independent temporal signals – the fringe phase and the grating assembly displacement – and tests their linear coupling. The pipeline proceeds in three stages.

Stage 1: Grating Assembly Tracking and Fringe Extraction. We localize the grating assembly in each frame by detecting and tracking the four ArUco corner markers, yielding per-frame corner coordinates and a centroid trajectory $\bar{c}^{(t)}$. Using the four corners, we compute a per-frame homography that warps the grating assembly region into a canonical, front-parallel rectangle and apply CLAHE [48] to enhance fringe visibility. We then determine the fringe orientation via contour analysis [36], rotate the image to align fringes vertically, and collapse the 2D image into a 1D intensity profile by averaging along the fringe direction. Fig. 3-(b, c) shows the tracked Moiré region and the corresponding warped canonical frames from our pipeline.

Stage 2: Phase Tracking. We extract the fringe phase from the 1D profile using a one-dimensional FFT. Because the profile is approximately sinusoidal at the Moiré beat frequency, the phase of the corresponding Fourier component directly encodes the lateral position of the fringe pattern. On the first frame, we identify the dominant frequency peak and lock its bin k^*; on each subsequent frame we read the phase at k^*:

$$\phi^{(t)} = \arg\Big[\mathcal{F}\{I^{(t)}\}(k^*)\Big] .$$

To avoid 2π discontinuities, we track the phase incrementally: we wrap each frame-to-frame difference to $[-\pi, \pi]$ using $\mathrm{wrap}(\theta) = ((\theta + \pi) \bmod 2\pi) - \pi$ and accumulate the deltas into a smooth cumulative phase signal:

$$\Phi^{(t)} = \sum_{\tau=1}^{t} \mathrm{wrap}\Big(\phi^{(\tau)} - \phi^{(\tau-1)}\Big) .$$

Stage 3: Displacement Estimation and Correlation. As derived in Sec. 3.2, camera rotation does not induce parallax between the grating layers and therefore does not drive fringe phase change; only the translational component of camera motion does. The raw centroid trajectory $\bar{c}^{(t)}$ from Stage 1 conflates both components, so we decompose it using PnP-based pose estimation. The four ArUco

marker centers provide four coplanar 3D-to-2D point correspondences per frame, which we pass to a PnP solver (*i.e.*, OpenCV's `cv::solvePnP`) to recover the camera pose (R_i, t_i) for each frame. We use a pinhole camera model with approximate intrinsics computed from image size ($f_x = f_y = \alpha \max(W, H)$). We then isolate the translation-only displacement by re-projecting the 3D centroid of the marker layout using each frame's camera position but the reference frame's orientation, yielding the rotation-compensated translational displacement $\Delta u^{(t)}_{\text{trans}}$. Because the Moiré fringes vary along one axis only, we project $\Delta u^{(t)}_{\text{trans}}$ onto the fringe-sensitive direction (perpendicular to the fringe lines) to obtain the scalar displacement signal.

We test the Moiré motion invariant (Eq. (3)) by computing the Pearson correlation between the cumulative phase signal $\{\Phi^{(t)}\}$ and the projected translational displacement $\{\Delta u^{(t)}_{\text{trans}}\}$. We evaluate the correlation over sliding windows of 30 frames, producing a per-window correlation curve across the video that serves as the authenticity signal. In the next section, we show that real and AI-generated videos produce markedly different correlation curves, suggesting ways to differentiate them.

5. Validation

We validate the Moiré motion invariant across three categories of video: real recordings, physics-based renderings, and AI-generated videos. We apply the same verification pipeline described in Sec. 4 to all categories and compare the resulting correlation signals.

5.1. Real Recorded Video

We record videos across 12 subjects in 29 distinct scenes (25 indoor, 4 outdoor) using an iPhone 15 at 1080p, 60 fps. Each recording falls into one of three motion configurations: (1) grating assembly static with camera moving, (2) grating assembly moving with camera static, and (3) both moving simultaneously (We use notation such as "Outdoor 1" to indicate an outdoor video recorded under the first condition, and so on.). The case where both are static is excluded, as our verification requires relative movement between the camera and the grating assembly to produce fringe shifts. Across subjects, we vary the movement speed and viewing distance (approximately 1.0 to 3.0 m). This yields a total of 87 videos, each approximately 15 seconds in duration.

5.2. Physics-based Rendering Video

To validate the pipeline under fully controlled conditions, we reproduce the grating assembly in Blender [3] and render videos with the Cycles physically-based rendering engine. The virtual assembly replicates the two-layer geometry described in Sec. 4, and the virtual camera matches the iPhone 15's focal length and 1080p resolution. Because Cycles traces light through the two grating layers, it accurately reproduces the parallax-driven fringe formation, providing an ideal-case for verifying the pipeline.

Each rendered sequence places the grating assembly at the world origin, sized to match the physical prototype. The camera begins at $(0.5, 0, z)$ and translates linearly to $(0, 0, z)$ over 120 frames, with small random rotations injected along the trajectory to simulate realistic camera shake. We constrain the rotations so that the grating assembly and all four ArUco markers remain fully visible in every frame. We sweep the camera distance z from 1.1 to 1.7 m in 0.1 m increments, generating 10 sequences per distance with distinct random rotation profiles, for a total of 70 rendered videos (Rendering 1 in Fig. 4). We also created a group for pure translational motion without rotation (Rendering 2 in Fig. 4).

5.3. AI-generated Video

We evaluate AI-generated video as a potential attack against the Moiré motion invariant, testing three state-of-the-art video generation models: Veo 3.1 [16], Grok Imagine [40] (closed-source), and LTX-2 [19] (open-source). We first attempted text-to-video (T2V) generation, providing descriptions of a person holding a grating assembly with visible Moiré fringes (replicating our real video settings). For each of the three motion configurations, we engineered specific text prompts through 15 to 20 rounds of iterative refinement. Across all models and prompt variations, none produced output containing a recognizable or trackable Moiré pattern. The generated videos either omitted the fine periodic structure or rendered it as a static, incoherent texture bearing no resemblance to real Moiré fringes, which can be identified easily by the naked eye.

We therefore focus our quantitative evaluation on the stronger image-to-video (I2V) setting, where the attacker provides one real frame of the grating assembly as conditioning input along with a text prompt describing the desired motion. We extracted the initial frame from each of the 87 authentic videos to serve as a visual condition and refined three corresponding motion prompts over 15 to 20 rounds. For each extracted frame, we use three optimized motion prompts and generate multiple outputs, yielding 321 videos in total. Because this first frame is captured from a real camera, it contains authentic Moiré fringes, giving the generator a visual prior on fringe appearance.

From the initial pool of 321 synthetic videos, manual inspection revealed that the generative models produced severe temporal inconsistencies, such as the sudden disappearance or extreme deformation of the grating assembly. We therefore screened and discarded 229 such corrupted samples, yielding a final curated dataset of 92 high-

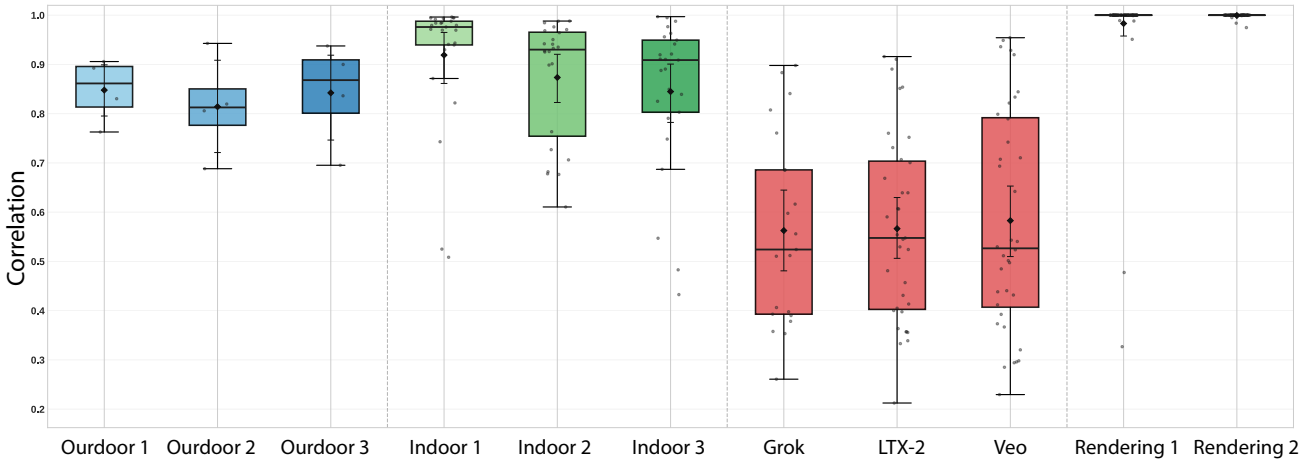

Figure 4. Distribution of Pearson correlation coefficients $|\rho|$ across all video categories. Real recordings (indoor and outdoor) cluster at high correlation, physics-based renderings achieve near-perfect correlation, and AI-generated videos concentrate at markedly lower values.

quality synthetic videos that look authentic to the human eye (see the supplementary video for examples of failed AI-generated videos).

Even in the I2V setting, most current video generation models fail to reproduce the basic visual structure of the grating assembly or the ArUco fiducial markers, and such videos would be immediately flagged as non-authentic. To stress-test the robustness of the Moiré signature itself, we adopt a deliberately generous evaluation protocol: we manually select the corner of Moiré region from the first generated frame, track it using Lucas-Kanade optical flow [4] across subsequent frames, and manually correct the tracking when it deviates. This ensures the extracted signal reflects purely the generated Moiré pattern, independent of preprocessing failures.

5.4. Results

Fig. 4 shows the distribution of Pearson correlation coefficients $|\rho|$ across all video categories. Real recordings cluster at high correlation values ($\mu = 0.87$, $\sigma = 0.14$), confirming that the Moiré motion invariant holds under diverse real-world conditions including varying lighting, distance, and motion configurations. Physics-based renderings achieve near-perfect correlations ($\mu = 0.99$), validating that the pipeline correctly captures the underlying optical relationship under ideal conditions. In contrast, AI-generated videos concentrate at substantially lower values ($\mu = 0.57$, $\sigma = 0.21$).

To test whether the difference is statistically significant, we apply a Welch's t-test, which is appropriate here because the two groups have unequal variances ($\sigma = 0.14$ vs. 0.21) and the sample sizes ($n = 86$ real, $n = 92$ AI-generated) are large enough for the Central Limit Theorem to ensure robustness to non-normality. The test confirms a highly significant separation ($t(160) = 11.6$, $p < 10^{-20}$). To quantify effect size, we compute Cohen's $d = 1.71$, well above the 0.8 threshold conventionally considered a large effect. The gap is consistent across all three generators: Grok ($\mu = 0.56$), LTX-2 ($\mu = 0.57$), and Veo ($\mu = 0.58$) each fall far below the real-video baseline, with no model achieving a meaningful advantage.

Moreover, this results should be read as a conservative lower bound shaped by deliberate choices on both sides of the experiment. On the attack side, our protocol grants every possible advantage: image-conditioned generation with a real first frame, manual Moiré region tracking with human correction, and cherry-picked outputs representing the best case for each model. Under a less generous protocol the attacker would also need to fool ArUco marker detection and automated Moiré tracking. Also, the I2V attack is itself inherently constrained, as the attacker must first obtain a real recording of the target scene to extract a conditioning frame, significantly limiting creative freedom compared to unconstrained T2V generation. On the defense side, our prototype uses a minimal, proof-of-concept pipeline. Our tracking pipeline could be improved to reduce the influence of factors such as reflections and blur in the Moiré assembly, and to use better methods for more accurate rotation estimation. **The key takeaway is therefore not the absolute threshold, but the existence of a physically grounded, measurable signal that distinguishes real from AI-generated video**.

5.5. Other Threat Models

Beyond the generative attacks evaluated above, we consider two additional threat scenarios through conceptual analysis.

These threats do not require new experiments but highlight important boundaries and extensions of our approach.

Threat: Moiré region splicing. The attacker takes a clip of an authentic grating assembly from one real video and composites it into a different (potentially AI-generated) video. The fringes in the spliced region are physically correct because they originate from a real recording.

Analysis. The spliced fringes carry the phase evolution from the *source* recording, but our pipeline measures correlation against the grating assembly displacement in the *target* video. These two signals reflect different camera-assembly geometries, so they will not be linearly related. The transplanted phase signal cannot match a different motion trajectory, and the correlation will be low by construction. We omit empirical validation for this threat because the outcome is self-evident from the formulation of the invariant.

Threat: Face swap and localized editing outside the Moiré region. The attacker takes a real video of person A holding a grating assembly and applies a face-swap model (*e.g.* DeepFaceLab [27], FaceFusion [12]) to replace only the face region, making it appear as person B. The grating assembly region is left entirely untouched.

Analysis. This is the most challenging threat for our system. Because the grating assembly region is unmodified, the Moiré fringes retain their physically correct phase evolution, and the Moiré motion invariant holds. Our correlation check will report high $|\rho|$, authenticating the video. The attacker obtains a video of "person B" apparently holding person A's grating assembly. Since the editing occurs entirely outside the Moiré region, the Moiré correlation check alone cannot detect this manipulation.

However, this class of attack is well addressed by a mature body of work on deepfake and face-swap detection [10, 17, 31, 39]. We view these as complementary: our method authenticates whether the video was produced by a physical optical system, while editing detectors identify localized tampering. The two can be combined in a layered verification pipeline. We additionally outline two forward-looking mitigations:

Mitigation 1: Personalized Moiré identity (Moiré ID). If the grating assembly encodes a unique, verifiable identity, analogous to a public key, the identity encoded in the assembly belongs to person A, not person B. This can be achieved by fabricating the grating with specific spatial frequencies that produce a unique Moiré beat period, acting as a frequency-domain fingerprint. Alternatively, the grating can incorporate non-uniform line spacing so the fringe pattern at different spatial positions encodes distinct identifiers. The verifier then checks not only that the Moiré motion invariant holds but also that the assembly identity matches the claimed identity of the person in the video.

Mitigation 2: Spatial overlap verification protocol. In interactive settings such as video conferencing, the platform can enforce a protocol requiring the participant to move the grating assembly across their face (*e.g.* "please wave your Moiré signature in front of your face now"). This ensures the grating assembly region spatially overlaps with the face region for at least a brief interval. Any face-swap model operating on this footage must either (a) modify the overlapping grating assembly pixels, which breaks the Moiré motion invariant, or (b) leave the assembly intact and distort the face around it, producing visible artifacts.

5.6. Discussion

Our threat analysis reveals a clear division that aligns with the current landscape of video forensics. Localized editing attacks such as face swaps fall outside the scope of what Moiré correlation alone can detect, since the grating assembly region remains untouched. However, existing deepfake and video editing detectors already achieve 82–97% accuracy on such manipulations [10, 17, 30, 31, 39], making them a problem that complementary tools can address. The harder and more urgent challenge lies in detecting fully AI-generated videos: large-scale benchmarks such as GenVidBench [30] show that cross-generator detection accuracy for T2V and I2V content often falls to 42–70%. It is precisely this gap that the Moiré motion invariant fills. No current generative model can synthesize the precise frame-to-frame coupling between fringe phase and grating assembly displacement that real optics produce, and splicing attacks are similarly defeated because the transplanted phase signal cannot match a different camera-assembly geometry. Our method thus complements existing forensic detectors in a layered verification pipeline: editing detectors screen for localized tampering, while the Moiré motion invariant provides a physics-grounded authentication layer against the generative attacks that remain beyond the reach of current detection methods.

6. Conclusion

Limitations. Our approach relies on the assumption that current generative models do not perform physics-based simulation of Moiré optics. Should future models incorporate full ray-tracing of the two-layer grating geometry, they could in principle reproduce the correct fringe-displacement coupling. Additionally, the method requires relative motion between the camera and the grating assembly to generate a measurable phase signal; authentication is not possible in the case where both remain static throughout the recording.

To conclude, we presented a Moiré-based video authentication method that leverages a physics-governed, deterministic link between fringe phase evolution and camera-grating motion. We derived a Moiré motion invariant that

is independent of viewing distance and camera intrinsics, and built a proof-of-concept system that extracts and tests it from video alone. Experiments show significantly different correlation values for real versus AI-generated videos, enabling verifiable discrimination between authentic and synthetic recordings.

Beyond the specific Moiré-based implementation, our work points to a broader principle: physical phenomena governed by deterministic optical laws can serve as authentication signatures that are fundamentally difficult for statistical generative models to forge. We hope this perspective inspires future exploration of other physics-based signatures for media authentication, complementing existing forensic and watermarking approaches to build more robust defenses against AI-generated misinformation.

References

[1] Isaac Amidror. *The Theory of the Moiré Phenomenon, Volume I: Periodic Layers*. Springer, 2nd edition, 2009. 2, 3

[2] BBC. Disgraceful deep-fake AI video condemned by presidential candidate, 2025. 1

[3] Blender Foundation. Blender: Open source 3d creation suite, 2025. 6

[4] Jean-Yves Bouguet. Pyramidal implementation of the Lucas Kanade feature tracker: Description of the algorithm. Technical report, Intel Corporation, Microprocessor Research Labs, 2001. 7

[5] Tim Brooks, Bill Peebles, Connor Holmes, Will DePue, Yufei Guo, Li Jing, David Schnurr, Joe Taylor, Troy Luhman, Eric Luhman, Clarence Ng, Ricky Wang, and Aditya Ramesh. Video generation models as world simulators. OpenAI Technical Report, 2024. 1, 3

[6] ByteDance Seed Team. Seedance 2.0. ByteDance Seed, 2026. 1

[7] Daniel Campos Zamora, Mustafa Doga Dogan, Alexa F Siu, Eunyee Koh, and Chang Xiao. MoiréWidgets: High-precision, passive tangible interfaces via moiré effect. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–10, 2024. 3

[8] CNN. Finance worker pays out $25 million after video call with deepfake chief financial officer, 2024. 1

[9] Coalition for Content Provenance and Authenticity (C2PA). C2PA technical specification. C2PA, 2022. 1

[10] Florinel-Alin Croitoru, Andrei-Iulian Hiji, Vlad Hondru, Nicolae-Catalin Ristea, Paul Irofti, Marius Popescu, Cristian Rusu, Radu Tudor Ionescu, Fahad Shahbaz Khan, and Mubarak Shah. Deepfake media generation and detection in the generative AI era: A survey and outlook. *arXiv preprint arXiv:2411.19537*, 2024. 1, 3, 8

[11] Patrick Esser, Johnathan Chiu, Parmida Atighehchian, Jonathan Granskog, and Anastasios Germanidis. Structure and content-guided video synthesis with diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7346–7356. IEEE, 2023. 3

[12] FaceFusion. FaceFusion: Industry leading face manipulation platform. GitHub, 2024. 8

[13] Hany Farid. Lighting (in) consistency of paint by text. *arXiv preprint arXiv:2207.13744*, 2022. 3

[14] Emin Gabrielyan. The basics of line moiré patterns and optical speedup. *arXiv preprint arXiv:physics/0703098*, 2007. 3

[15] Google DeepMind. SynthID: Identifying AI-generated content. Google DeepMind, 2023. 1

[16] Google DeepMind. Veo 3.1, 2025. 3, 6, 1

[17] Diego Gragnaniello, Davide Cozzolino, Francesco Marra, Giovanni Poggi, and Luisa Verdoliva. Are GAN generated images easy to detect? A critical analysis of the state-of-the-art. In *ICME*, 2021. 1, 8

[18] Hui Guo, Shu Hu, Xin Wang, Ming-Ching Chang, and Siwei Lyu. Eyes tell all: Irregular pupil shapes reveal GAN-generated faces. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2904–2908. IEEE, 2022. 3

[19] Yoav HaCohen, Bar Brazowski, Nisan Chiprut, Yaron Bitterman, Anna Kvochko, Alon Berkowitz, Dor Shalem, Daniel Lifschitz, Dana Moshe, Eyal Porat, Elad Richardson, Guy Shiran, Itai Chachy, Jonathan Chetboun, Matan Finkelson, Michael Kupchick, Nir Zabari, Nadav Guetta, Niv Kotler, Ofir Bibi, Ori Gordon, Philippe Panet, Rami Benita, Shachar Armon, Vladimir Kulikov, Yaki Inger, Yonadav Shiftan, Zeev Melumian, and Zeev Farbman. LTX-2: Efficient joint audio-visual foundation model. *arXiv preprint arXiv:2601.03233*, 2026. 6, 1

[20] Roger D. Hersch and Sébastien Chosson. Band moiré images. *ACM TOG*, 23(3):239–248, 2004. 3

[21] Runyi Hu, Jie Zhang, Yiming Li, Jiwei Li, Qing Guo, Han Qiu, and Tianwei Zhang. Videoshield: Regulating diffusion-based video generation models via watermarking. In *International Conference on Learning Representations (ICLR)*, 2025. 3

[22] Christian Internò, Robert Geirhos, Markus Olhofer, Sunny Liu, Barbara Hammer, and David Klindt. Ai-generated video detection via perceptual straightening. In *Advances in Neural Information Processing Systems*, 2025. 3

[23] Oded Kafri and Ilana Glatt. *The Physics of Moiré Metrology*. Wiley, 1990. 3

[24] Kuaishou Technology. Kling: A pioneering ai video generation model. https://kling.kuaishou.com/, 2024. 3

[25] Rohit Kundu, Hao Xiong, Vishal Mohanty, Athula Balachandran, and Amit K. Roy-Chowdhury. Towards a universal synthetic video detector: From face or background manipulations to fully ai-generated content. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2025. 3

[26] Zongxia Li, Xiyang Wu, Guangyao Shi, Yubin Qin, Hongyang Du, Tianyi Zhou, Dinesh Manocha, and Jordan Lee Boyd-Graber. Videohallu: Evaluating and mitigating multi-modal hallucinations on synthetic video understanding. In *Advances in Neural Information Processing Systems*, 2025. 3

[27] Kunlin Liu, Ivan Perov, Daiheng Gao, Nikolay Chervoniy, Wenbo Zhou, and Weiming Zhang. Deepfacelab: Integrated, flexible and extensible face-swapping framework. *Pattern Recognition*, 141:109628, 2023. 8

[28] Luma AI. Dream machine: High quality, realistic video generation from text and images. https://lumalabs.ai/dream-machine, 2024. 3

[29] Peter F. Michael, Zekun Hao, Serge Belongie, and Abe Davis. Noise-coded illumination for forensic and photometric video analysis. *ACM Transactions on Graphics*, 44(5), 2025. 2, 3

[30] Zhisheng Ni, Qiuyu Yan, Mengqi Huang, Tao Yuan, Yifan Tang, Haiyang Hu, Xin Chen, and Yaowei Wang. GenVidBench: A 6-million benchmark for AI-generated video detection. In *AAAI*, 2026. 8

[31] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Niessner. Faceforensics++: Learning to detect manipulated facial images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 1, 3, 8

[32] Runway. Introducing Runway Gen-4.5. Runway Research, 2025. 1

[33] Hadleigh Schwartz, Xiaofeng Yan, Charles J. Carver, and Xia Zhou. Combating falsification of speech videos with live optical signatures. In *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, pages 3296–3310, 2025. 2, 3

[34] Cesar A. Sciammarella. The moiré method – a review. *Experimental Mechanics*, 22(11):418–433, 1982. 3

[35] Theophanis Sethapakdi, Matteo Perroni-Scharf, Meng Li, Jingyi Li, Justin Solomon, Arvind Satyanarayan, and Stefanie Mueller. FabObscura: Computational design and fabrication for interactive barrier-grid animations. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)*, 2025. 3

[36] Satoshi Suzuki and Keiichi Abe. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing*, 30(1):32–46, 1985. 5

[37] Hiroshi Takasaki. Moiré topography. *Applied Optics*, 9(6): 1467–1472, 1970. 3

[38] Chuangchuang Tan, Xiang Ming, Jinglu Wang, Renshuai Tao, Bin Li, Yunchao Wei, Yao Zhao, and Yan Lu. Semantic visual anomaly detection and reasoning in ai-generated images. *arXiv preprint arXiv:2510.10231*, 2025. 3

[39] Sheng-Yu Wang, Oliver Wang, Richard Zhang, Andrew Owens, and Alexei A. Efros. CNN-generated images are surprisingly easy to spot... for now. In *CVPR*, 2020. 1, 8

[40] xAI. Grok imagine video, 2025. 6, 1

[41] Chang Xiao and Changxi Zheng. Moiréboard: A stable, accurate and low-cost camera tracking method. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pages 881–893, 2021. 3

[42] Fanrui Zhang, Dian Li, Qiang Zhang, Jun Chen, Gang Liu, Junxiong Lin, Jiahong Yan, Jiawei Liu, and Zheng-Jun Zha. Fact-R1: Towards explainable video misinformation detection with deep reasoning. In *Advances in Neural Information Processing Systems*, 2025. 3

[43] Kevin Alex Zhang, Lei Xu, Alfredo Cuesta-Infante, and Kalyan Veeramachaneni. Robust invisible video watermarking with attention. *arXiv preprint arXiv:1909.01285*, 2019. 3

[44] Shuhai Zhang, Zihao Lian, Jiahao Yang, Daiyuan Li, Guoxuan Pang, Feng Liu, Bo Han, Shutao Li, and Mingkui Tan. Physics-driven spatiotemporal modeling for ai-generated video detection. In *Advances in Neural Information Processing Systems*, 2025. 3

[45] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000. 4

[46] Chende Zheng, Ruiqi Suo, Chenhao Lin, Zhengyu Zhao, Le Yang, Shuai Liu, Minghui Yang, Cong Wang, and Chao Shen. D3: Training-free ai-generated video detection using second-order features. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12852–12862, 2025. 3

[47] Zihang Zou, Boqing Gong, and Liqiang Wang. Anti-neuron watermarking: Protecting personal data against unauthorized neural networks. In *European Conference on Computer Vision (ECCV)*, pages 449–465. Springer, 2022. 3

[48] Karel Zuiderveld. Contrast limited adaptive histogram equalization. In *Graphics Gems IV*, pages 474–485. Academic Press, 1994. 5

Moiré Video Authentication: A Physical Signature Against AI Video Generation

Supplementary Material

Overview

This supplementary material provides additional details, experimental results, and discussion to support the main manuscript. It is organized as follows:

- Sec. A: Detailed camera configurations, experiment setup, and instructions for reproducing our physical device.
- Sec. B: A detailed description of the attack process.
- Sec. C: Additional experimental results, including the ROC curves.
- Sec. D: Extended discussion and future work, addressing potential limitations and real-world scalability.

A. Camera Configurations and Hardware Setup

Data were acquired using an iPhone 15 (1080p, 60 fps). To capture the Moiré effect, we employed dynamic camera movements (translation and varying viewing distance) relative to the Moiré board. This intentional motion is essential, as the captured interference patterns are inherently dependent on the perspective and relative displacement between the camera and the dual-grating layers. We utilized the camera's auto-exposure mechanism to adapt to the varying luminance of the interference fringes throughout the recording, ensuring consistent signal visibility without manual exposure locking.

A.1. Reproducing the Device

To facilitate the reproduction of the Moiré board in any laboratory environment, we provide the following hardware specifications and assembly instructions.

A.1.1. Materials and Digital Assets

The device consists of a three-layer optical stack. The primary components are as follows:

- **Base Layer (Support):** A 3mm (1/8 inch) thick clear plexiglass sheet (8" $\times$ 10") provides structural support. [Amazon Link]
- **Rear Layer (Reference Grating):** A printed pattern consisting of 160 horizontal lines with a pitch of $p_r = 0.52$ mm.
- **Front Layer (Viewing Grating):** A 1mm-thick lenticular lens sheet (50 LPI, $p_f = 0.508$ mm). [Amazon Link]
- **Digital Assets:** The reference grating is provided as a printable PDF (`reference grating.pdf`) in the root directory of the attached supplementary folder. To maintain the designed geometry (pitch $p_r = 0.52$ mm), the document must be printed at 100% scale (actual size) on A4 paper.

A.1.2. Assembly Instructions

The assembly requires careful layer alignment:

1. Place the printed Rear Layer (A4 paper) onto the 3mm Base Layer.
2. Overlay the Front Layer (Lenticular Sheet) ensuring the grating lines of both layers are parallel.
3. Secure the layers together to minimize the air gap between the two gratings, as this gap is critical for maintaining high-contrast Moiré interference across varying viewing angles and distances.

B. Detailed Description of the Attack Process

To comprehensively evaluate the robustness of our proposed Moiré video authentication method against state-of-the-art AI generation, we meticulously designed an adversarial video synthesis pipeline. This section details our model selection, prompt engineering strategy, dataset curation statistics, and the full prompt templates used to synthesize the fake videos.

B.1. Model Selection

We selected three state-of-the-art video generation models for both Text-to-Video (T2V) and Image-to-Video (I2V) capabilities. To ensure a comprehensive evaluation across different model architectures and availability levels, we chose:

- **Veo 3.1 Pro [16]**: Representing the industry-leading, closed-source commercial models with high fidelity and temporal consistency.
- **Grok Imagine Video [40]**: Representing a highly competitive, alternative closed-source generative architecture.
- **LTX-2 Video [19]**: Representing the current state-of-the-art in open-source video generation models, allowing us to evaluate publicly accessible attack vectors.

B.2. Adversarial Video Synthesis and Dataset Curation

To synthesize the adversarial attack videos, we targeted three distinct kinematic paradigms: (1) static subject with a moving camera, (2) moving subject with a static camera, and (3) simultaneous movement of both subject and camera.

Controlled Prompt Strategy. Designing prompts for video generation is inherently subjective. Rather than heavily engineering idiosyncratic prompts tailored to the specific inductive biases of each individual model, we deliberately

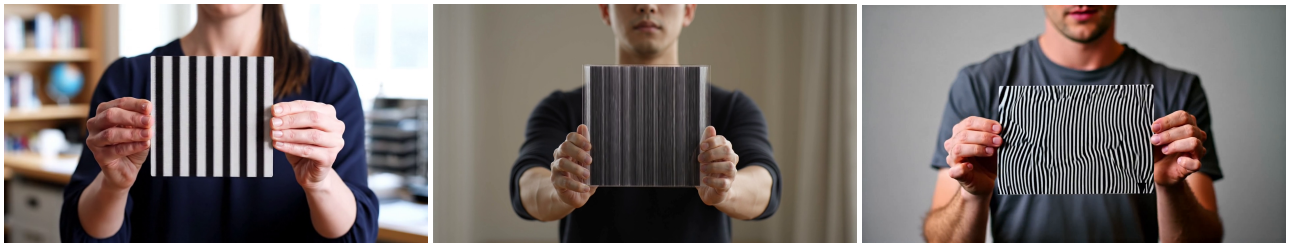

Figure 5. Examples of pure Text-to-Video (T2V) generation failures. The models struggle to synthesize accurate, rigid Moiré patterns from scratch, frequently resulting in unnaturally thick stripes or severely wobbly, non-physical deformations.

formulated a unified, rigorous set of prompt templates applied uniformly across all three models. This standardizes the evaluation baseline, acting as a controlled zero-shot test of each model's intrinsic ability to adhere to complex physical and geometric constraints (i.e., maintaining the Moiré board's spatial integrity and phase structures) without model-specific prompt overfitting.

Generation Yield and Stringent Filtering. To construct a robust evaluation foundation, we first collected a dataset comprising 87 authentic Moiré videos captured across both indoor (75 videos) and outdoor (12 videos) environments.

Using the three selected state-of-the-art models, we initially explored pure Text-to-Video (T2V) generation. For each motion paradigm, we meticulously engineered specific text prompts and refined them over 15 to 20 iterations. Using the best-performing prompts, we generated 20 videos per model. However, empirical observations revealed that these 60 purely text-generated videos lacked physical realism and could be easily distinguished by the naked eye (see Fig. 5 for representative artifacts).

Consequently, we shifted our focus to a first-frame-conditioned Image-to-Video (I2V) strategy. To significantly enhance visual fidelity and physical consistency, we extracted the initial frame from the authentic videos to serve as a visual anchor. We then randomly assigned an optimized motion prompt to guide the synthesis. This approach yielded 257 I2V generated videos, culminating in an initial synthetic pool of 317 videos (60 T2V and 257 I2V videos).

Final Curated Dataset for Evaluation. As discussed in the main manuscript, manual inspection revealed that generative models frequently produced severe temporal inconsistencies, such as the sudden disappearance, detachment, or extreme deformation of the Moiré grating assembly (Fig. 6 illustrates examples of these persistent artifacts).

We screened and discarded 225 such corrupted samples. This curation yielded a final dataset of **92 high-quality synthetic videos** that appear deceptively authentic to the human eye. **All subsequent algorithmic evaluations, threshold analyses, and ROC metric computations were conducted exclusively on this challenging 92-video subset** alongside the 87 authentic videos.

Tab. 1 summarizes the dataset generation and curation pipeline, detailing the exact yield per generative model. Additionally, Tab. 2 provides a comprehensive statistical breakdown of the video metadata for the final evaluation dataset, reflecting the native generation formats of the respective models.

B.3. Full Prompt Templates

This section provides the complete text of the prompt templates engineered for the adversarial video synthesis pipeline. For each of the three kinematic paradigms, we developed a prompt for the first-frame-conditioned Image-to-Video (I2V) generation, and a corresponding prompt for the pure Text-to-Video (T2V) generation.

As discussed in Section B.2, the I2V prompts proved significantly more effective at preserving the required physical geometry of the grating structure across all tested models.

B.3.1. Paradigm 1: Moving Camera, Static Subject

This paradigm evaluates the generative model's ability to render natural optical flow and Moiré interference driven solely by camera translation, while maintaining the rigid structure of a stationary subject and grating. The corresponding prompts are detailed in Fig. 7 and Fig. 8.

B.3.2. Paradigm 2: Static Camera, Moving Subject

This condition challenges the model to synthesize smooth human motion while dynamically altering the Moiré effect based strictly on the subject's translation relative to a fixed viewpoint. The prompts are shown in Fig. 9 and Fig. 10.

B.3.3. Paradigm 3: Simultaneous Movement

This represents the most complex kinematic scenario, requiring the model to maintain the relative scale and perspective of the subject while simultaneously generating coherent background translation and optical interference. The prompts are provided in Fig. 11 and Fig. 12.

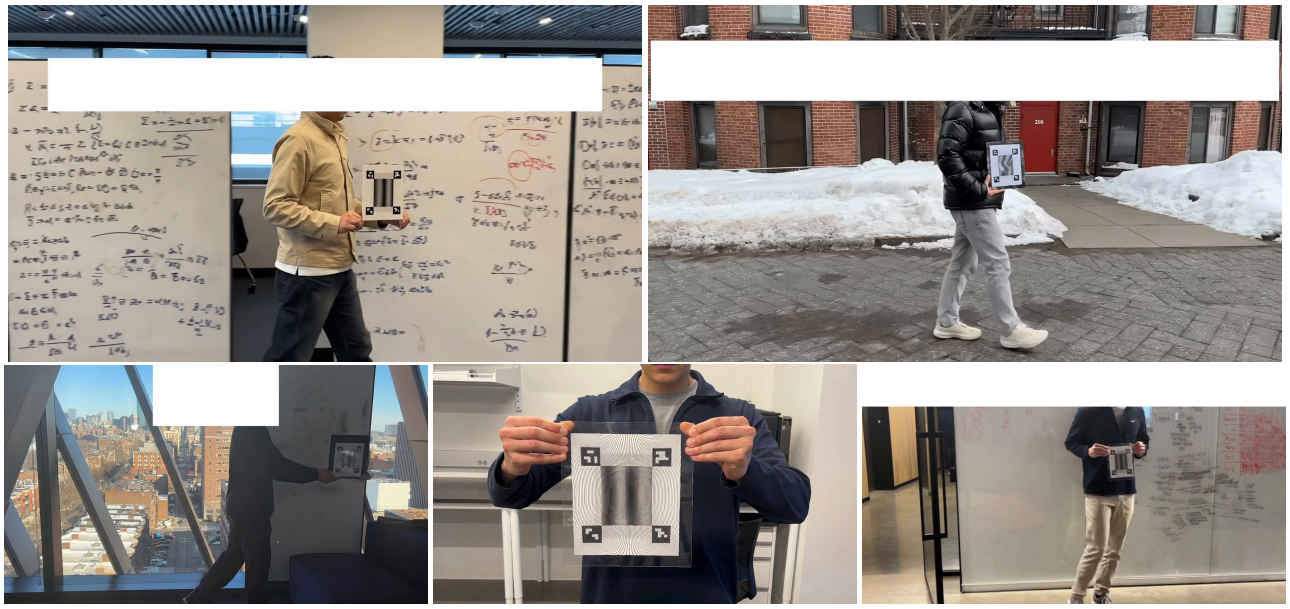

Figure 6. Representative frames from Image-to-Video (I2V) generation exhibiting temporal and physical inconsistencies. While first-frame conditioning significantly improves overall fidelity, models still struggle with maintaining the rigid geometry of the grating during motion, resulting in structural warping, blurring, and detachment from the subject's hands.

Table 1. Dataset Generation and Curation Pipeline. The initial synthetic pool of 317 videos was rigorously filtered to remove 225 samples with severe temporal or physical collapses, yielding 92 high-quality adversarial videos.

Pipeline Stage	**Veo 3.1 Pro**	**Grok Imagine**	**LTX-2 Video**	**Total**
T2V Generated	20	20	20	60
I2V Generated (First-Frame)	83	87	87	257
Initial Synthetic Pool	103	107	107	**317**
Discarded (Severe Artifacts)	-67	-86	-72	-225
Final Curated Synthetic	**36**	**21**	**35**	**92**

C. Additional Experimental Results

C.1. Threshold-Based Real/Fake Classification

Problem Setup and Notation. Let each video sample be represented by a scalar score $s \in \mathbb{R}$, denoting the maximum correlation (`best_correlation`) produced by our detector. We define the binary ground-truth labels as $y = 1$ for authentic real videos (captured across all indoor and outdoor settings) and $y = 0$ for AI-generated fake videos. Following our data curation protocol (detailed in Sec. B), generated samples that exhibited catastrophic physical failure (yielding exactly $s = 0$) or invalid non-finite values (NaN/Inf) are excluded from this threshold analysis. These samples are trivially detectable by the naked eye and are therefore treated as invalid for continuous score-distribution analysis.

Given a detection threshold τ, the prediction rule is de-

Table 2. Metadata Summary of the Final Evaluation Dataset. The statistics reflect the authentic captures and the inherent output configurations of the respective AI models used for the 92 curated fake videos.

Video Property	**Authentic (Real)**	**Veo 3.1 Pro**	**Grok Imagine**	**LTX-2 Video**
Video Count	87	36	21	35
Resolution	Mostly 1920×1080	1280×720	848×480	1920×1080
Average FPS	59.8	24.0	24.0	25.0
Average Duration	14.4 s	8.0 s	8.0 s	10.3 s

```
Using the provided first frame, generate a photorealistic video of the same person clearly holding a rectangular moire
     board in their hands. The board is physically grasped by the person and remains in their hands for the entire
    video.
The moire board is a real, solid object with vertical black and white stripes printed on its surface. The stripes
    exist only on this physical board.
The person and the board remain completely stationary in the scene.
The board must stay fixed in the person's hands.
The board cannot move independently, cannot drift, cannot expand, cannot shrink, cannot rotate, and cannot deform.
Its size and position in the frame must remain consistent with the first frame.
Only the camera moves. The camera translates smoothly and strictly horizontally from left to right at constant speed.
No vertical motion. No forward or backward motion. No zoom. No tilt. No rotation.
The vertical stripes on the board must remain perfectly vertical.
Any visible moire interference must occur only on the surface of the board as a natural optical result of horizontal
    camera movement.
No stripes, patterns, or interference effects may appear anywhere outside the board.
Do NOT generate any new moving pattern in the background or across the room.
There must be no global animated overlay.
The background must remain stable and pattern-free.
Lighting, shadows, and perspective must stay physically realistic and consistent with the first frame.
Maintain sharp detail and strong temporal coherence across all frames.
Duration: 8 seconds. Frame rate: 24-30 fps. Style: natural documentary realism.
```

Figure 7. I2V Prompt 1 (First-Frame Conditioned) used for Paradigm 1: Moving Camera, Static Subject.

```
Generate a photorealistic video from scratch of a person clearly holding a rectangular moire board in their hands. The
     board is physically grasped by the person and remains in their hands for the entire video.
The moire board is a real, solid object with vertical black and white stripes printed on its surface. The stripes
    exist only on this physical board.
The person and the board remain completely stationary in the scene.
The board must stay fixed in the person's hands.
The board cannot move independently, cannot drift, cannot expand, cannot shrink, cannot rotate, and cannot deform.
Its size and shape must remain constant throughout the entire video.
Only the camera moves. The camera translates smoothly and strictly horizontally from left to right at constant speed.
No vertical motion. No forward or backward motion. No zoom. No tilt. No rotation.
The vertical stripes on the board must remain perfectly vertical.
Any visible moire interference must occur only on the surface of the board as a natural optical result of horizontal
    camera movement.
No stripes, patterns, or interference effects may appear anywhere outside the board.
Do NOT generate any new moving pattern in the background or across the room.
The background must remain stable and pattern-free.
Lighting, shadows, and perspective must stay physically realistic.
Maintain sharp detail and strong temporal coherence across all frames.
Duration: 8 seconds. Frame rate: 24-30 fps. Style: natural documentary realism.
```

Figure 8. T2V Prompt 1 (Text-Only) used for Paradigm 1: Moving Camera, Static Subject.

fined as:

$$\hat{y}(\tau) = \begin{cases} 1, & \text{if } s \geq \tau \quad \text{(Predict Real)} \\ 0, & \text{if } s < \tau \quad \text{(Predict Fake)} \end{cases} \tag{5}$$

Evaluation Metrics. By sweeping the threshold τ, we compute the standard components of the confusion matrix: True Positives (TP), False Negatives (FN), True Negatives (TN), and False Positives (FP). From these, we derive the True Positive Rate ($\text{TPR} = \frac{\text{TP}}{\text{TP+FN}}$) and True Negative Rate ($\text{TNR} = \frac{\text{TN}}{\text{TN+FP}}$).

We report both standard *Accuracy* and *Balanced Accuracy*. While standard Accuracy measures the overall proportion of correct predictions, it can be biased by class imbalance. Balanced Accuracy, defined as $\frac{1}{2}(\text{TPR}(\tau) + \text{TNR}(\tau))$, equally weights the positive (real) and negative (fake) recall, providing a more robust metric for imbalanced datasets.

ROC Curve and Threshold Selection. The Receiver Operating Characteristic (ROC) curve is generated by plotting the TPR against the False Positive Rate ($\text{FPR} = 1 - \text{TNR}$) across all possible operational thresholds. The global separability of the two distributions, independent of any specific threshold, is summarized by the Area Under the Curve (AUC), approximated numerically via trapezoidal integration.

For practical deployment scenarios requiring a binary decision, we select the optimal threshold τ^* by maximizing the standard accuracy:

$$\tau^* = \arg\max_{\tau} \text{Accuracy}(\tau) \tag{6}$$

If multiple thresholds yield identical accuracy, we break ties by selecting the one with the higher Balanced Accuracy.

Results and Analysis. Fig. 13 presents the ROC curve of our detection algorithm, achieving a strong AUC of 0.8453.

```
Using the provided first frame, generate a photorealistic 8-second video of the same person clearly holding a
     rectangular moire board in their hands. The board is physically grasped by the person and remains in their hands
     for the entire video.
The moire board is a real, solid object with vertical black and white stripes printed on its surface. The stripes
     exist only on this physical board.
The camera remains completely stationary for the entire video.
Only the person moves.
The person moves smoothly and strictly horizontally (left or right) at a constant speed.
No vertical movement. No forward or backward movement.
No rotation of the body toward or away from the camera. No tilting of the board.
The board must stay fixed in the person's hands.
The board cannot move independently, cannot drift, cannot expand, cannot shrink, cannot rotate, and cannot deform.
Its size, orientation, and shape must remain consistent with the first frame.
The board must always face the camera directly and remain upright.
The vertical stripes on the board must remain perfectly vertical.
Any visible moire interference must occur only on the surface of the board as a natural optical result of horizontal
     movement.
No stripes, patterns, or interference effects may appear anywhere outside the board.
Do NOT generate any new moving pattern in the background.
No global animated overlay.
The background must remain stable and pattern-free.
Lighting, shadows, and perspective must stay physically realistic and consistent with the first frame.
Maintain sharp detail and strong temporal coherence across all frames.
Duration: 8 seconds. Frame rate: 24-30 fps. Style: natural documentary realism.
```

Figure 9. I2V Prompt 2 (First-Frame Conditioned) used for Paradigm 2: Static Camera, Moving Subject.

```
Generate a photorealistic 8-second video of a person clearly holding a rectangular moire board in their hands. The
     board is physically grasped by the person and remains in their hands for the entire video.
The moire board is a real, solid object with vertical black and white stripes printed on its surface. The stripes
     exist only on this physical board.
The camera remains completely stationary for the entire video.
Only the person moves.
The person moves smoothly and strictly horizontally (left or right) at a constant speed.
No vertical movement. No forward or backward movement.
No rotation of the body toward or away from the camera. No tilting of the board.
The board must stay fixed in the person's hands.
The board cannot move independently, cannot drift, cannot expand, cannot shrink, cannot rotate, and cannot deform.
The board must always face the camera directly and remain upright.
The vertical stripes on the board must remain perfectly vertical.
Any visible moire interference must occur only on the surface of the board as a natural optical result of horizontal
     movement.
No stripes, patterns, or interference effects may appear anywhere outside the board.
Do NOT generate any new moving pattern in the background.
No global animated overlay.
The background must remain stable and pattern-free.
Lighting, shadows, and perspective must stay physically realistic.
Maintain sharp detail and strong temporal coherence across all frames.
Duration: 8 seconds. Frame rate: 24-30 fps. Style: natural documentary realism.
```

Figure 10. T2V Prompt 2 (Text-Only) used for Paradigm 2: Static Camera, Moving Subject.

It is crucial to note that these metrics are computed *exclusively* on the heavily curated subset of 92 high-quality adversarial videos (as detailed in Sec. B.2), deliberately excluding 225 easily detectable generative failures. Consequently, this 0.8453 AUC reflects the algorithm's discriminative capability under a highly challenging, "hard-negative" evaluation setting, ensuring that the performance is not artificially inflated by trivially flawed generations.

Furthermore, as illustrated in Fig. 14, our threshold selection criterion identifies an optimal decision boundary at an exceptionally high correlation value of $\tau^* \approx 0.927$, where both standard and Balanced Accuracy reach their peak. The high value of this optimal threshold fundamentally underscores the strictness of our proposed Moiré motion invariant: authentic physical captures consistently maintain near-perfect phase-displacement correlations, whereas even the most visually deceptive AI generations fail to satisfy this rigid mathematical and optical constraint.

D. Discussion and Future Work

While our proof-of-concept validates the Moiré motion invariant and demonstrates its effectiveness against state-of-the-art video generative models, there remain important limitations and avenues for future research before deployment at scale. We discuss these aspects and outline potential system-level solutions.

```
Using the provided first frame, generate a photorealistic 8-second video of the same person clearly holding a
     rectangular moire board in their hands. The board is physically grasped by the person and remains in their hands
     for the entire video.
The moire board is a real, solid object with vertical black and white stripes printed on its surface. The stripes
     exist only on this physical board.
Both the camera and the person move smoothly and strictly horizontally (left or right) at the same constant speed and
     in the same direction.
The relative position between the camera and the person remains constant.
There is no change in framing, scale, or perspective.
The person stays centered in the frame.
No vertical movement. No forward or backward movement. No zoom. No tilt. No rotation.
The board must stay fixed in the person's hands.
The board cannot move independently, cannot drift, cannot expand, cannot shrink, cannot rotate, and cannot deform.
Its size, orientation, and shape must remain identical to the first frame.
The board must always face the camera directly and remain upright.
The vertical stripes on the board must remain perfectly vertical.
Any visible moire interference must occur only on the surface of the board as a natural optical result of horizontal
     motion.
No stripes, patterns, or interference effects may appear anywhere outside the board.
Do NOT generate any new moving pattern in the background.
No global animated overlay.
The background may translate horizontally due to motion, but must remain physically realistic and stable.
Maintain sharp detail and strong temporal coherence across all frames.
Duration: 8 seconds. Frame rate: 24-30 fps. Style: natural documentary realism.
```

Figure 11. I2V Prompt 3 (First-Frame Conditioned) used for Paradigm 3: Simultaneous Movement.

```
Generate a photorealistic 8-second video of a person clearly holding a rectangular moire board in their hands. The
     board is physically grasped by the person and remains in their hands for the entire video.
The moire board is a real, solid object with vertical black and white stripes printed on its surface. The stripes
     exist only on this physical board.
Both the camera and the person move smoothly and strictly horizontally (left or right) at the same constant speed and
     in the same direction.
The relative position between the camera and the person remains constant.
There is no change in framing, scale, or perspective.
The person stays centered in the frame.
No vertical movement. No forward or backward movement. No zoom. No tilt. No rotation.
The board must stay fixed in the person's hands.
The board cannot move independently, cannot drift, cannot expand, cannot shrink, cannot rotate, and cannot deform.
The board must always face the camera directly and remain upright.
The vertical stripes on the board must remain perfectly vertical.
Any visible moire interference must occur only on the surface of the board as a natural optical result of horizontal
     motion.
No stripes, patterns, or interference effects may appear anywhere outside the board.
Do NOT generate any new moving pattern in the background.
No global animated overlay.
The background may translate horizontally due to motion, but must remain physically realistic and stable.
Maintain sharp detail and strong temporal coherence across all frames.
Duration: 8 seconds. Frame rate: 24-30 fps. Style: natural documentary realism.
```

Figure 12. T2V Prompt 3 (Text-Only) used for Paradigm 3: Simultaneous Movement.

Toward Real-World Deployment and Multi-directional Patterns. Currently, our empirical evaluation demonstrates strong discriminative performance in controlled laboratory and standard indoor/outdoor settings. However, large-scale, "in-the-wild" deployment will require robustness against extreme environmental factors, such as severe motion blur, low-light noise, and aggressive social media compression (e.g., H.264/H.265 transcoding on platforms like WhatsApp or X). To strengthen the signature against such degradation, future hardware iterations will incorporate *multi-directional Moiré patterns*. A one-dimensional grating produces fringes sensitive to camera motion along a single axis. Incorporating multiple grating orientations (e.g., orthogonal or hexagonal grids) within a single assembly extends sensitivity to arbitrary motion trajectories. This not only ensures a strong correlation signal regardless of how the user moves the camera but also exponentially increases the dimensionality of the authentication signature. A generative model would need to simultaneously reproduce the correct phase-displacement coupling along every grating axis, compounding the difficulty for an attacker.

Advanced Generative Attacks and Fundamental Limits. Our threat model assumes that current generative architectures (Diffusion and Transformers) learn statistical mappings rather than performing explicit physics-based simulations of Moiré optics. While we utilized meticulously optimized prompts to generate the strongest possible ad-

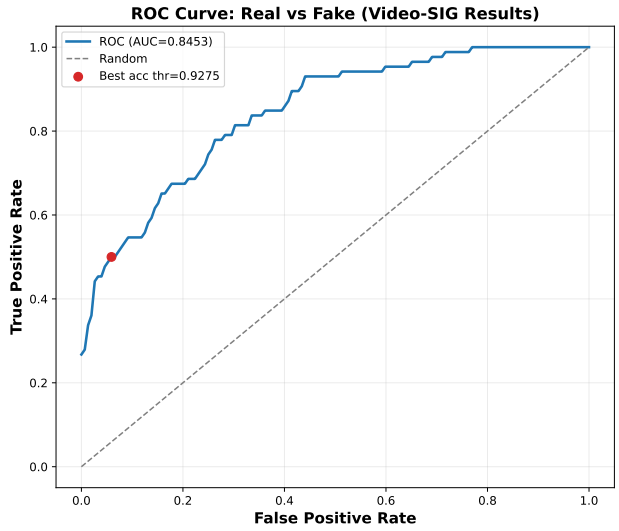

Figure 13. ROC Curve for real vs. fake video classification. Our method achieves an AUC of 0.8453, demonstrating strong global separability based on the Moiré correlation signature.

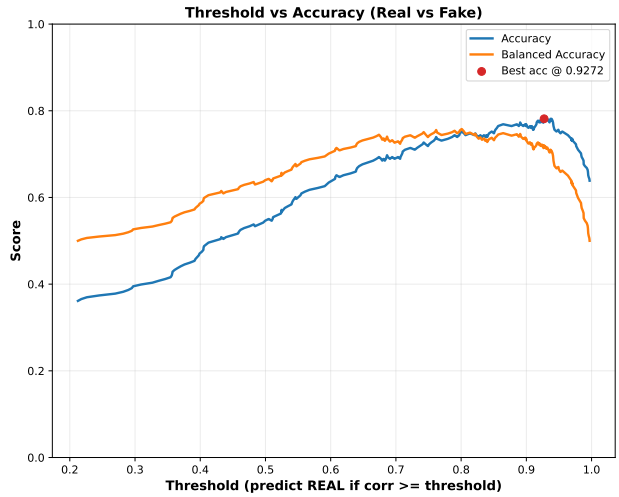

Figure 14. Impact of the decision threshold (τ) on Accuracy and Balanced Accuracy. The optimal operational point is identified at $\tau^* \approx 0.927$.

versarial videos (as detailed in Sec. B), a highly sophisticated attacker might attempt to bypass the system by rendering the Moiré board using a traditional 3D graphics engine (e.g., Unreal Engine with full ray-tracing) and compositing it into an AI-generated scene. While theoretically possible, achieving photorealistic coherence between a ray-traced object and a generative background—including lighting, shadows, and perspective matching—requires immense manual effort, thereby neutralizing the primary threat of "effortless, large-scale AI generation." Finally, our method strictly requires relative motion between the camera and the grating assembly to produce a measurable phase-displacement signal. Authentication is fundamentally impossible if both the camera and the subject remain entirely static throughout the recording. Future software interfaces must require the user to perform a slight scanning motion to trigger the verification protocol.