Tempora:
Recovering Occluded Background from Time in Monocular-to-Stereoscopic Conversion

Thomas Varin

CineFlight — Diorama 3D, Montréal, Canada

Public version — see Appendix A. Patent application filed. Engine version 0.7.0, schema `cine-flight.stereo-render.v0.2`.

Abstract

We present Tempora, a disocclusion-filling method for converting ordinary monocular video into stereoscopic video. In pursuit of fidelity over plausibility, Tempora rests on one observation: the background hidden behind a subject at frame *N* is almost always *visible elsewhere in the same shot*, because the camera or the subject moved. A single pre-pass accumulates, per shot, the mean of every pixel classified as background across all frames — a *plate* — and disoccluded regions are then resampled from that plate rather than extrapolated from neighbouring colour. The method therefore recovers real footage instead of synthesising texture, and two independent gates refuse the plate wherever it cannot be trusted. On our reference sequence, 80.4 % of disoccluded pixels (median over frames, both eyes) are filled with observed background, and inter-frame flicker measured on synthetic pixels alone falls from 5.85 to 4.98 (0–255 scale). We further report an instrumented, falsification-oriented evaluation protocol: the renderer writes 60 measured quantities per frame, refuses to publish a render that fails its own guards, and its benchmark tool ships with an explicit list of forbidden vocabulary. Two competing fill strategies were implemented, measured, and rejected on that evidence; both rejections are documented here. All measurements come from a single sequence without ground truth, and we state throughout what that does and does not license us to claim.

Correspondence: 3D@cineflight.ca
Project page: cineflight.ca/3d

This report describes what the code does, and only that. Where a quantity is not measured, it is named as not measured. Where an earlier claim of ours was falsified by a later measurement, the falsification is reported rather than the claim.

Contents

1 Introduction

Converting a monocular film into a stereoscopic pair requires inventing a viewpoint that was never recorded. Estimating per-pixel depth is now a solved-enough problem [9, 12, 6] that the difficulty has moved downstream: once the pixels of frame N are displaced to synthesise a second eye, the regions they vacate — the disocclusions — contain no information at all. Everything a viewer will call an artefact happens there. Halos around people, edges that smear, subjects that look cut out of cardboard, backgrounds stretched into taffy: these are not depth errors, they are fill errors.

The prevailing response is to synthesise. Neighbouring colour is extrapolated, or a generative model is asked to imagine what was behind the subject. Both approaches share a property that matters when the footage belongs to somebody: what appears in the hole is *plausible*, not *true*. For a client who is watching their own wedding, the distinction is not academic. An invented detail in a memory is noticed by exactly one person — the person whose memory it is.

In this work we step back from the framing of disocclusion as an inpainting problem and return to a more elementary question: is the information actually missing? For a static image it

is. For a *shot* it usually is not. Over the two or three seconds of a typical shot, the camera drifts, the subject turns, someone walks through frame — and the wall, the foliage, the table that sat behind the subject at frame N was directly photographed at frame $N \pm k$. Our work provides a mechanism for retrieving it.

Tempora accumulates, once per shot and before any rendering, the mean of every pixel classified as background over all frames of that shot, in a half-resolution buffer we call the *plate*. At render time, a disoccluded pixel is resampled from the plate at the background coordinate obtained by unit-slope extrapolation from its donor. Because the plate contains background by construction, the extrapolation that condemns the equivalent single-frame method does not apply. Two gates then decide, per pixel, whether the plate may be used: a minimum number of contributing observations, and a colour-coherence check evaluated *at the donor* rather than inside the hole. Rejected pixels fall back to a smoothed donor colour; the method never leaves a hole.

A second contribution is methodological, and in our judgement the more transferable one. The renderer is built to refuse. It writes sixty measured quantities per frame; it publishes its outputs atomically only after its own guards pass; it records a SHA-256 of the raw pixel stream separately from the container hash, because a Matroska container is not deterministic and a bit-exactness guarantee that rests on the container is not a guarantee; and its companion depth benchmark ships with a `vocabulaire_interdit` — a list of phrases the tool refuses to emit, among them "more accurate" and "metric error", because a single sequence without ground truth does not license them. Section 6.4 reports two fill strategies that this instrumentation killed.

2 Related work

Depth-image-based rendering. The classical pipeline warps a source image by a disparity field derived from depth, then repairs the resulting cracks; Fehn's formulation [3] remains the reference statement of the problem for broadcast 3D, and post-rendering 3D warping [7] established the forward-warping formulation we adopt. Backward warping — resampling the destination from the source — is trivially hole-free but cannot represent occlusion: two surfaces at different depths competing for the same destination pixel are blended rather than ordered. Forward warping represents occlusion correctly and produces holes. We take the second route and treat the holes as the actual research problem.

Monocular depth. The input side of the problem has moved quickly. Dense prediction transformers [9] and the Depth Anything family [12, 6] now produce relative depth of sufficient quality that, in our experience, the residual visual defects of a conversion are dominated by what happens in the disoccluded regions rather than by depth error. This report is about that second half; the engine consumes a depth array and records its SHA-256, and is agnostic to which estimator produced it.

Inpainting and background extrapolation. The dominant family fills disocclusions from spatial context — by fast marching from the hole boundary [11], by exemplar-based patch synthesis [2], or, in the layered-depth formulations built for 3D photography, by a learned generative prior [10, 8]. Video methods propagate along optical flow and attend across a temporal window [13]. All of them produce content that did not exist in the source. The failure mode we care about is not visual implausibility but silent fabrication: the fill looks correct and is not the film. Our position is narrower than a criticism of these methods, which are strong at what they

do — it is that a client watching their own wedding is the one viewer able to detect an invented detail, and the only one whose judgement matters.

Temporal background modelling. Accumulating a per-pixel background model from video is long-established in surveillance, and the observation that a moving foreground reveals its own background is older still. What we contribute is not that observation but its use as a *disocclusion fill under a depth-ordered forward warp*, with acceptance gates strict enough that the method degrades to a conservative fallback instead of degrading into invention — and with the instrumentation to tell which of the two happened, per pixel, per frame. Registration of the accumulation uses enhanced correlation coefficient maximisation [4], warm-started frame to frame.

Disparity and comfort. Nonlinear disparity mapping [5] established that the perceptually correct operation on a stereoscopic disparity range is a non-uniform remapping rather than a global scale; our soft knee (§4.2) is a deliberately minimal instance of that idea, chosen because it is exactly the identity over the bulk of the range and therefore verifiable. Vergence-accommodation conflict and its comfort zone [1] motivate the floating window of §4.3, though we stress in that section that this engine enforces no angular criterion of its own.

3 Tempora

We tackle the synthesis of a stereoscopic pair from a monocular sequence and a per-frame depth map, under a hard constraint: any pixel presented to the viewer must either come from the film or be declared as a fallback.

3.1 Formulation

Let $I(u,y)$ be the source frame on its useful area of width L, and $p(u,y) \in [0,1]$ the depth map normalised per shot, with the convention *near_high* (1 is nearest). The total inter-ocular disparity is

$$d(u,y) = A \cdot (p(u,y) - c) \tag{1}$$

where A is the shot's amplitude in pixels and c its convergence plane — the normalised depth that receives zero disparity. Geometry is symmetric: each eye receives half the total disparity, in opposite directions, so that

$$x_{left} = u + ½\cdot d(u,y), \quad x_{right} = u - ½\cdot d(u,y) \tag{2}$$

This halving is not cosmetic. It is the reason the floating-window criterion of §4.3 must be written against d and not $d/2$, a distinction that cost us a measurably wrong window (§4.3).

Depth domain. The engine accepts a metric domain in [0,1] and a relative inverse domain stored as affine `uint16`. In the inverse domain **no reciprocal is applied**, because screen disparity is affine in that domain already. An earlier version took the reciprocal first and then applied an affine disparity, which cost a factor of 14.6 in relief on a close-up face — measured 2026-08-04. The manifest records which domain was used and the SHA-256 of the depth array itself.

3.2 Forward projection under a total-order z-buffer

Each source pixel is projected to its destination and competes for it. When several sources claim one destination, that is an occlusion and not an ambiguity: with a *near_high* map, the nearest wins. To make the outcome independent of evaluation order we pack every contribution into a single signed 64-bit key,

```
key = round(p·(2²⁰−1)) << (bits_source + 8)
    | round(w·255)        << bits_source
    | ((L−1) − u)
```

and reduce with a maximum. Depth dominates; ties are broken by sub-pixel weight w, then by the smallest source index — the last field is stored inverted so that a single maximum implements the whole order. The key is *total*: no two contributions to one destination can be equal, because the source field separates them. Consequently the reduction is commutative and associative, and the winner is identical whether it is computed by `np.maximum.at` on CPU or by an atomic `scatter_reduce(amax)` on GPU.

Bit budget. `bits_source` is computed from the actual width rather than fixed. A constant of 11 bits suits 1920 px but overflows at 3840, where the index needs 12; the extra bit would encroach on the weight field and could change the winner of a collision. The engine raises rather than silently truncate if the fields exceed 63 bits.

Sub-pixel handling. The z-buffer decides *visibility only*. The winner's source index is then corrected by its sub-pixel residual to build a source-coordinate map, and the final colour is obtained by a single bilinear `remap` of the original frame through that map. This yields the correct occlusion of forward projection together with the resampling quality of backward warping.

What an uncovered destination means. With one tap, an empty destination means no *rounded* source landed there — which conflates true disocclusion with columns skipped by rasterisation. Our original documentation claimed the former; measurement on a hand region-of- interest falsified it, and the claim was struck. Adding a second bilinear tap removes 57.9 % of the holes in that ROI, which places the residual much closer to genuine disocclusion without proving it on its own. The engine records the tap count in the manifest and warns that hole percentages must not be compared across tap settings.

3.3 The temporal background plate

Before rendering, `construire_plaques()` decodes the range a second time, strictly sequentially, and accumulates one plate per shot. Both passes call the same normalisation, so that they see the same world. A pixel contributes to its shot's plate iff

$$p(u,y) \le c_{shot} \tag{3}$$

that is, iff it lies behind the shot's own convergence plane. The temporal window is the *entire shot*; there is no bounded $\pm N$ search and no per-hole candidate list. Aggregation is the arithmetic mean of all background observations, per plate pixel, at half resolution (≈8 MB per shot).

Sub-pixel purity. Image and background mask are reduced with `INTER_AREA`; the reduced mask therefore yields a *fraction* of background sub-pixels per plate pixel, and accumulation

happens only where that fraction reaches a high threshold. A plate pixel straddling the subject's edge would be precisely the contamination the method exists to forbid.

Motion compensation. With `--plaque-compensation affine`, an affine transform is estimated per frame against the shot's first frame by ECC [4] on a strongly reduced grayscale image, **initialised from the previous frame's transform** — camera motion is continuous, and ECC converges poorly from the identity over a large accumulated displacement. Accumulation happens in reference coordinates; out-of-frame borders fall to fraction 0 and contribute nothing. ECC failures are caught, counted per shot, and printed.

Sampling. At render time, a hole pixel x with donor x_D is assigned the background coordinate obtained by unit-slope extrapolation,

$$u(x) = s(x_D) + (x - x_D) \tag{4}$$

where $s(\cdot)$ is the source coordinate of the donor. This is exactly the formula of a single-frame variant we rejected (§6.4) — but what condemned that variant was that it sampled the *occluder* in the current frame, and the plate contains background by construction. The coordinate passes through the current frame's inverse ECC transform, is clamped inside the plate, and is sampled bilinearly; the number of observations behind the sample is taken as the minimum over the four corners.

3.4 Two acceptance gates

Gate 1 — enough observations. A plate pixel is usable only above a minimum number of contributing observations. A single observation may be a specular highlight, a motion blur, or a piece of subject that the depth threshold misclassified. The test is applied both at the hole coordinate and at the donor coordinate.

Gate 2 — colour coherence, evaluated at the donor. The plate, sampled *at the donor's position*, is compared with the donor's real colour; the worst of the three BGR channels must stay within a tight tolerance. This location is the whole point. Comparing the plate to the donor colour *inside the hole* condemns every textured background, because there the plate legitimately differs from the donor — that difference is its reason for existing. Truth exists only at the donor: if the plate reproduces the donor's real colour at the donor's position, it is locally registered, and its content inside the hole is trustworthy. A lying plate — strong camera motion, failed registration — fails at the donor and falls back. This correction was made on 2026-08-05 after the verifier caught the original formulation; the first render under the corrected gate is the one reported in §6.2.

Fallback. Tempora never leaves a hole. Rejected pixels keep the donor colour already written by the colour fill — the farther-depth of the two horizontal neighbours — and only those fallback pixels then receive a vertical normalised smoothing. Plate-filled pixels are deliberately **not** smoothed: they carry real texture, the vertical comb artefact does not exist for them, and smoothing would only destroy it. The manifest counts `pixels_plaque`, `pixels_repli_couleur`, `plaque_rejets_echantillons` and `plaque_rejets_coherence` separately, so the provenance of every filled pixel is auditable after the fact.

The only true residual holes are segments wider than a fixed maximum width, refused upstream and reported as `residu_trous_*_pourcent`. We do not attempt them: a hole wider than the background's own pattern would be an extrapolation, not a reconstruction.

3.5 Implementation details

Depth input. Depth is estimated by Depth Anything 3 in its metric Large configuration (`depth-anything/DA3METRIC-LARGE`) [6], run at full frame rate on GPU, and stored as the reciprocal of the metric prediction — a relative inverse-depth domain in which screen disparity is affine, encoded as affine `uint16` with its decode parameters in a sidecar file. The pre-analysis stage probes the same model on sampled frames of each shot to build the per-shot profile of §4. The renderer itself remains agnostic: it consumes a depth array, verifies its declared domain and orientation, and records its SHA-256.

Should that model fail to load, the pipeline falls back to Depth Anything V2 Small [12] rather than stopping — an assessment that never arrives serves no one. The two are not interchangeable in quality, so the substitution is printed as a blocking-style warning, flagged in the pre-analysis report, and reproduced in the delivery report's provenance section alongside the file digests. We report this because a reader who discovered a silent model substitution after fourteen pages on measurement discipline would be right to discount everything else in them.

The engine is a single Python file of 4 621 lines with no proprietary dependency — NumPy, OpenCV and FFmpeg only. Decoding is strictly sequential and never seeks: positioning by `CAP_PROP_POS_FRAMES` is approximate on streams with bidirectional frames and would make the frame-to-depth pairing codec-dependent. Frame indices remain absolute in both video and depth. Output is Full side-by-side, piped to FFmpeg as raw `bgr24`, 8-bit `yuv420p` or 10-bit `p010le`. All pixel-domain constants are expressed at a 1920-px reference width and scaled by the actual useful width. Appendix A lists them.

4 Per-shot calibration

A close-up and a wide landscape do not take the same treatment, and a single global setting is wrong for both. The renderer computes nothing about the content: a separate pre-analysis has already seen every shot, and the engine *refuses a report without an exploitable profile rather than inventing values*. Amplitude and convergence are then derived per shot from the profile's `force_3d` and `pop_out`:

$$A_{shot} = A_{max} \cdot force\_3d \,/\, 100, \quad c_{shot} = 1 - pop\_out \,/\, 100 \tag{5}$$

Shots whose measured temporal drift crosses a threshold are flagged fragile and receive a reduced amplitude factor on top.

4.1 Robust normalisation on the useful area

Bounds are the 2nd and 98th percentiles over a fixed number of sampled frames per shot — and they are computed on the **cropped useful area only**. Letterbox bars would otherwise dominate the distribution. Measured on a case where the useful area occupies [0.40; 0.60] and the bars are 0: bounds became 0.000/0.593 instead of 0.404/0.596, the normalised median 0.841 instead of 0.497, and at convergence 0.5 that projects 100 % of the shot in front of the screen with a median disparity of +19.7 px instead of −0.15.

4.2 Soft knee at the depth extremes

Percentiles followed by a hard clip flatten roughly 2 % of the nearest and 2 % of the farthest pixels to a constant, and all internal relief there disappears — a very close face becomes card-

board. Following the principle that a disparity range should be remapped non-uniformly rather than globally scaled [5], we replace the clip by a knee that is strictly the identity on $[w;\ 1-w]$ and exponential outside it:

$$g(t) = w \cdot e^{(t-w)/w} \text{ for } t \le w; \quad g(t) = 1 - w \cdot e^{-(t-(1-w))/w} \text{ for } t \ge 1-w \tag{6}$$

for a narrow shoulder width w. The function is C^1 at both junctions in value and in slope, so no gradient discontinuity is introduced and no artificial contour appears in the relief; it is monotone, and the asymptotes 0 and 1 are never reached. The falsifiable criterion was stated in advance: ≈2 %/2 % of flat-clipped pixels under hard clip, ≈0 under the soft knee. The 2026-08-04 headset verdict was a net gain with no palpitation, and the knee was adopted for the production pipeline; the engine default remains `off` so that the legacy path stays bit-exact.

4.3 Floating window

An object in crossed parallax that touches the lateral frame edge sends two contradictory cues: stereopsis says it is in front, occlusion by the edge says it is behind. This conflict is a major cause of discomfort in a headset [1]. A pixel at abscissa x has binocular correspondence iff $0 \le x - d \le L-1$, so the monocular band at the left edge is $x < d$.

> **CORRECTION OF 2026-08-04**
>
> The first implementation tested $x < d/2$. It measured the band disoccluded by a half-warp, that is half the real band. The mask derived from it gave the window edge half the disparity it needed, so the window stayed *behind* the object violating it — which corrects nothing.

Negative disparities are explicitly not violations: looking at a distant object through a window frame is a natural situation. Violation is aggregated as the 99th percentile across lines within a frame, then the maximum across a fixed number of probed frames per shot. Masking is asymmetric — only the differential part carries window disparity — over a common base band, with a half-cosine feather that reaches exactly 1. A feather that stops short of 1 leaves a vertical contour of very low contrast but perfectly straight and constant in time, hence far more visible than its amplitude suggests. A safety cap flags an amplitude too strong for the content, and a probe that saturates aborts the render: a measurement equal to the probe is not a measurement, it is a bound.

An independent per-frame verification re-applies the correspondence criterion pixel by pixel rather than reusing the pre-pass. Its own first version tested $x - d < 0$ instead of $x - d <$ base and returned zero while six columns per line remained monocular. Under `--fenetre-stricte`, any offending frame deletes the partial outputs and raises.

> **ON ANGULAR COMFORT THRESHOLDS**
>
> The engine computes no angular quantity. There is no arc-minute, no viewing distance, no interpupillary distance and no headset field of view anywhere in it, and it enforces no angular parallax limit. A 15-arc-minute threshold does exist in our A/B reporting tool, at a declared 45° field of view (1 px = 1.40625′), where it is applied to the *width of the widest filled disocclusion band* — what the headset will show of the defect — and not to parallax. That tool states its own thresholds as declared, adjustable heuristics rather than measurements. Any statement that this pipeline holds parallax under an angular comfort limit would misdescribe it.

4.4 Transitions at cuts

A vergence jump at a cut is visible. Amplitude and convergence are linearly interpolated from the previous shot's values over the first 8 frames of each shot — about one third of a second at 24 fps, the time of a vergence adjustment. Outside the ramp the current profile is returned exactly, with no floating-point arithmetic, which is what preserves the bit-exactness of the legacy path.

5 Instrumentation, determinism and refusal

What is measured. Every frame writes a CSV row of roughly sixty fields. The ones that matter for this report are the 95th percentile of absolute total disparity; the percentage of pixels flat-clipped at each depth extreme; the percentage of destination pixels uncovered before filling and still uncovered after; window violation before and after masking, and the count of monocular pixels outside the applied mask; and the scintillation block, measured *only between consecutive frames of the same shot* — including `scintillement_rempli`, the mean absolute inter-frame difference restricted to pixels that are *synthetic in both frames*. That restriction is what makes it a measurement of the fill rather than of the film.

Fill provenance. In plate mode the manifest reports

$$plaque\_part\_reelle = median_{frames}\,[\ plate\ /\ (plate + fallback)\ ] \tag{7}$$

summed over both eyes per frame, together with the per-shot fraction of the plate carrying at least three observations.

A diagnostic that contradicted a metric. The Jacobian of the warp, $J = 1 \mp ½{\cdot}d'(x)$, detects folds ($J \le 0$) and severe compression ($0 < J < 0.25$). On one edge it reported 2.22 % of folded pixels where only 0.18 % of disocclusion was declared. The two formulations disagree by construction — a centred difference spreads a discontinuity over both neighbours, roughly doubling the affected fraction and halving the severity: on the synthetic case, 0.42 % at −10.5 centred against 0.21 % at −22.0 discrete. The engine records both and states that any guard or quoted figure must rest on the discrete measure.

Determinism. There is no random number generator in the engine, and nothing to seed: a search for `random`, `seed`, `shuffle`, `Thread` and `multiprocessing` returns nothing. Parallelism is external and cuts at shot boundaries, where shots are independent by construction; the one global coupling — the constant window — is injected rather than recomputed, so segments concatenate without re-encoding and the result is bit-identical. The parameter fingerprint is a SHA-256 over a JSON dump with `sort_keys=True`, so key order cannot change it.

> **TWO OUTPUT HASHES, AND WHY**
>
> A non-regression check in v0.3.8 compared the container hash of two renders and failed, while the decoded frames were byte-identical: the Matroska container is not deterministic. The lesson was recorded in the code and in the schema — **any bit-exactness guard must bear on pixels** — hence `sortie_brute_sha256`, taken over the raw side-by-side bytes as they are piped to the encoder, alongside the container's own hash.
>
> A second lesson followed in v0.4.0: an earlier frame-hash comparison failed because OpenCV 4.13 had been replaced by 5.0 between the two reference renders. Bit-exactness is defined only at constant environment, which is why the manifest now carries a Python/NumPy/OpenCV block.

Refusal. The three outputs are written as `.part` files and renamed only after the guards pass; a refused render deletes them, because a refused result left on disk is eventually mistaken for an accepted one. Every opt-in mode defaults to `off`, and each `off` path is documented as bit-exact with the previous version.

6 Experiments

6.1 Protocol, and what it does not license

All figures below come from a single reference sequence — a wedding film, 24000/1001 fps, rendered at 3840×1080 Full-SBS from an inverse-domain depth array, with the pre-analysis profile active. There is no ground-truth stereo pair. Our depth benchmark states this in its own output rather than in a caption:

```
"mode": "benchmark instrumental sans vérité terrain",
"limites": {
  "verite_terrain": "absente",
  "sequences": 1,
  "comparaison_absolue": "impossible",
  "dispersion_rapportee": "inter-images, non inter-sequences",
  "intervalle_confiance_sequence": {
    "statut": "non_calculable",
    "raison": "une seule séquence disponible" } },
"vocabulaire_interdit": [
  "plus exact", "meilleure profondeur réelle",
  "erreur métrique", "intervalle de confiance séquence" ]
```

The consequence is stated plainly. The numbers that follow are *internal, relative and instrumental*. They support statements of the form "this configuration produces fewer synthetic pixels than that one, on this sequence". They do not support any statement about accuracy against a true depth, nor any confidence interval across sequences, and the tooling refuses to emit the vocabulary that would imply otherwise. The same benchmark refuses one comparison outright, with its reason recorded in the output: comparing an independent depth map to our reference by a fidelity metric would presuppose that our reference is a truth, which it is not.

One further caveat is built into the tool. A more coarsely quantised depth map obtains a *mechanically* better temporal-stability score, because every variation below one quantisation step disappears. The benchmark therefore reports the quantisation step alongside the stability figure and the ratio of error to step, and flags the reading as interpretable only with reserve.

6.2 Fill provenance and flicker

Table 1 reports the first render under the corrected coherence gate (§3.4).

Quantity	Value	Meaning
Plate coverage, per shot	68 – 100 %	fraction of the plate carrying ≥ 3 observations
Fill from observed background	**80.4 %**	median over frames, both eyes summed, plate ÷ (plate + fallback)
Flicker on synthetic pixels, before	5.85	mean |Δ| between consecutive frames, 0–255, pixels synthetic in both
Flicker on synthetic pixels, after	4.98	same measure, plate fill

Table 1 Tempora on the reference sequence. Coverage is a property of the plate; fill provenance and flicker are properties of the render. The flicker measure is restricted to pixels that are synthetic in both frames of a pair, so it measures the fill and not the film.

Roughly four filled pixels in five carry background that was photographed. The remaining fifth is not a hole and is not invention either: it is the donor colour, vertically smoothed, and it is counted separately in the manifest with its rejection reason — too few observations, or failed coherence at the donor.

6.3 Ablations

Configuration	Effect	Verdict
a. Second bilinear tap in the projection	−57.9 % holes	adopted; holes on the hand ROI, one tap → two
b. Coherence gate inside the hole	—	rejected: condemns every textured background
c. Coherence gate at the donor	80.4 % fill	adopted (§3.4)
d. Hard clip at the depth extremes	≈2 % / 2 % flat	rejected: internal relief lost, a close face flattens
e. Soft knee	≈0 % flat	adopted for production; engine default remains off
f. Bounds on the full frame	+19.7 px median	rejected: letterbox bars dominate; see §4.1
g. Bounds on the useful area	−0.15 px median	adopted
h. Reciprocal before affine disparity	×14.6 relief loss	rejected: disparity is affine in the inverse domain
i. Vectorised z-buffer reduction	−28 % render time	adopted; 13.3 → 9.6 s / 10 frames, projection 7.4 → 4.5 s
j. GPU plate sampling	−2 % render time	rejected: 0.112 s on 4.6 s does not buy back bit-for-bit comparability

Table 2 Ablations, on the reference sequence. Rows b–c, d–e and f–g are paired: the rejected configuration and the adopted one. Row j is a working, measured, deliberately disconnected optimisation — the code remains in the file, disabled, with its measurement.

6.4 Two falsified hypotheses

Both were implemented in full, measured, and killed by their own numbers. Both remain in the source with their measurements attached, because a rejected approach without its evidence gets proposed again.

Single-frame background extrapolation. The same unit-slope formula of Eq. (4), applied to the *current* frame's source coordinates rather than to a plate. On frame 45, 69.9 % of the filled pixels sampled the foreground, 58.5 % of them at a depth above 0.7, and the pattern of the groom's jacket repeated visibly across the hole. This is precisely the failure that the plate does not have, because the plate contains background by construction — and it is why we consider the two methods different despite the shared formula.

Texture-directed fill. The hypothesis was that a hole in a textured background should be filled by propagating the local texture. The measurement falsified it: on 63.7 % of the segments treated by texture, the mean vertical gradient inside the zone was identical to the reference to four significant figures (13.88 against 13.88), and the difference image was indistinguishable from zero — 2.53/255 of variation at one pixel, 9.85/255 at five. The mode is retained in the code, unused, with the measurement in the docstring.

6.5 GPU parity

The GPU path mirrors the same 64-bit key and the same tie-break; because the reduction is a maximum over a total key, atomics produce the same winner regardless of execution order. The only divergence is the final resampling, where the code emulates OpenCV's 5-bit (1/32) fractional quantisation before sampling so that the weights match. Measured outcome: integer decisions identical on both eyes; resampling within ±1 LSB on ≥ 99.5 % of pixels, 0.000 % beyond; ×16.9 throughput on an RTX 3090. The GPU path is therefore **declared not bit-exact** with the CPU path, traced as `_gpu` in the render signature and in the parameter fingerprint — a mode, never a silent substitution.

7 Limitations

We list what the pipeline does not do, in the words its own manifest uses.

- **One sequence, no ground truth.** Every figure in this report is internal and relative. No inter-sequence confidence interval is computable, and the tooling refuses to print one.
- **Unit-slope fill.** Eq. (4) is exact for a background of locally constant disparity and approximate on a steeply sloping background.
- **No texture synthesis.** A hole wider than the background's own pattern remains an extrapolation, not a reconstruction. Segments above the maximum width are refused outright and reported.
- **No temporal stabilisation of disparity.** Provisioned, not adopted: the measured inter-frame variation is 0.000 px at the median, which does not justify it.
- **Constant floating window** over the whole film. Dynamic convergence and the three-band shift are read from the pre-analysis, recorded in the manifest, and never applied.
- **Uniformly black bars only** for crop detection; coloured mattes are not handled.
- **No angular comfort computation** in the engine (§4.3).

- **Amplitude and convergence are not clamped.** The pre-analysis supplies `force_3d` and `pop_out`; the engine applies them without bounds.
- One reported quantity, `contamination_premier_plan_pourcent`, is structurally zero in the colour and plate modes because the colour comes from the donor itself. It is retained only to compare modes on equal footing and is not an empirical result.

8 Conclusion

Tempora shows that the hardest part of monocular-to-stereoscopic conversion — the regions where no information exists in the source frame — is largely a retrieval problem rather than a synthesis problem, provided one is willing to look along the time axis of the shot instead of along the spatial axes of the frame. A per-shot background plate, a unit-slope coordinate, and two gates that verify registration where truth exists, are enough to give four filled pixels in five a real photographic origin, while the fifth degrades to a declared, separately counted fallback rather than to invention.

The second result is about method rather than pixels. An engine that measures sixty quantities per frame, hashes its own raw output, refuses to publish when its guards fail, and ships a list of words its benchmark may not use, is an engine that can falsify its author. It did so three times in this report — on what an uncovered destination means, on where the coherence gate belongs, and on whether texture propagation does anything at all. We regard those three reversals as the most substantive evidence in the document, and we would distrust a conversion pipeline that had never produced any.

A Note on numerical values

The tuned constants of the method — observation minima, colour tolerances, band widths, probe counts — are omitted from this public version. They are the outcome of the calibration effort described in §6, they are not derivable from the method itself, and the patent application does not turn on their particular values. Every quantity presented as a *result* is complete and unaltered: the fill provenance, the flicker measurement, the ablations and the two falsified hypotheses of §6.4 are reported exactly as measured.

The full version of this report, including the table of named constants, is available on request to parties with a technical or commercial interest. Write to the correspondence address on the first page.

References

[1] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of Vision*, 11(8):11, 2011. doi:10.1167/11.8.11.

[2] A. Criminisi, P. Pérez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*, 13(9):1200–1212, 2004. doi:10.1109/TIP.2004.833105.

[3] C. Fehn. Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV. In *Proc. SPIE 5291, Stereoscopic Displays and Virtual Reality Systems XI*, pages 93–104, 2004. doi:10.1117/12.524762.

[4] G. D. Evangelidis and E. Z. Psarakis. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10): 1858–1865, 2008. doi:10.1109/TPAMI.2008.113.

[5] M. Lang, A. Hornung, O. Wang, S. Poulakos, A. Smolic, and M. Gross. Nonlinear disparity mapping for stereoscopic 3D. *ACM Transactions on Graphics* (Proc. SIGGRAPH), 29(3):75:1–75:10, 2010.

[6] H. Lin, S. Chen, J. H. Liew, D. Y. Chen, Z. Li, G. Shi, J. Feng, and B. Kang. Depth Anything 3: Recovering the visual space from any views. ByteDance Seed, arXiv preprint, November 2025.

[7] W. R. Mark, L. McMillan, and G. Bishop. Post-rendering 3D warping. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics*, pages 7–16, 1997.

[8] S. Niklaus, L. Mai, J. Yang, and F. Liu. 3D Ken Burns effect from a single image. *ACM Transactions on Graphics*, 38(6):184:1–184:15, 2019. doi:10.1145/3355089.3356528.

[9] R. Ranftl, A. Bochkovskiy, and V. Koltun. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12179–12188, 2021.

[10] M.-L. Shih, S.-Y. Su, J. Kopf, and J.-B. Huang. 3D photography using context-aware layered depth inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8025–8035, 2020.

[11] A. Telea. An image inpainting technique based on the fast marching method. *Journal of Graphics Tools*, 9(1):25–36, 2004. doi:10.1080/10867651.2004.10487596.

[12] L. Yang, B. Kang, Z. Huang, Z. Zhao, X. Xu, J. Feng, and H. Zhao. Depth Anything V2. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pages 21875–21911, 2024. arXiv: 2406.09414.

[13] S. Zhou, C. Li, K. C. K. Chan, and C. C. Loy. ProPainter: Improving propagation and transformer for video inpainting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10477–10486, 2023.