Multimodal 3D Monitoring and Visual Analytics via Dynamic Frequency Residual Splatting

Yongfeng Shan
School of CS
UTS, Australia
yongfeng.shan@student.uts.edu.au

Christy Liang
School of CS
UTS, Australia
jie.liang@uts.edu.au

Daming Luo
School of CS
UTS, Australia
daming.luo@student.uts.edu.au

Chenxuan Zhou
School of CS
UTS, Australia
chenxuanzhou90@gmail.com

Xiaoru Yuan
School of EECS
Peking University, China
xiaoru.yuan@pku.edu.cn

Robert Wu
School of CS
UTS, Australia
mingxuan.wu@uts.edu.au

Jun Li
School of CS
UTS, Australia
jun.li@uts.edu.au

***Abstract*—Multimodal packet streams—video synchronized with telemetry, pose estimates, and runtime events—increasingly support real-time monitoring in robotics, digital twins, and sensor networks. Frequency-domain diagnostics are attractive in this setting because they can localize where a rendered view deviates from the observation, yet they are often validated under simplified previous-frame proxies whose conclusions may not transfer to real rendering pipelines. We present *Coherent Frequency Packet Splatting for Dynamics* (CFPS-DYN), a visual analytics methodology that computes an Explainable Frequency Heatmap via a tiled, phase-sensitive spectral residual between the incoming observation and the current prediction. We evaluate five diagnostic representations—pixel residual, FFT amplitude, phase coherence, and two packet-based variants—under both a previous-frame *proxy* and *real 3D Gaussian Splatting* (3DGS) renders on TUM RGB-D sequences. Under the proxy, the pixel baseline exhibits an inflated correlation with temporal instability (r=0.920), but this oracle advantage collapses under real 3DGS renders (r=0.041) when temporal information is no longer leaked into the residual. We empirically observe a decoupling between temporal instability and reconstruction fidelity under real 3DGS renders, with stable-but-incorrect outputs occurring in high-error regions. In this regime, frequency-domain packet scalar energy is the strongest predictor of reconstruction error (r=0.236, $p<10^{-7}$), outperforming pixel residuals. A linked-view interface exposes heatmap overlays and timeline prioritization to support monitoring and intervention.**

***Index Terms*—CFPS-DYN, Dynamic Frequency Residual Splatting (DFRS), Visual Analytics, Multimodal Streams, Explainable Frequency Heatmaps, Adaptive Sampling, Online Scheduling, Temporal Stability**

I. Introduction

Visual analytics (VA) systems are increasingly deployed in streaming settings rather than offline analysis [1]–[4]. In robotics, sensor networks, and digital twins, analysts receive time-ordered packets that mix video with telemetry signals such as pose, inertial measurements, timestamps, and runtime events [5], [6]. These systems are used for monitoring and diagnosis, which means the visualization must provide stable and reliable evidence while the backend continuously updates a live 3D state. In this regime, the bottleneck is not peak rendering quality; it is maintaining trustworthy evidence and allocating limited computation as packets arrive under strict per-frame constraints on latency, memory, and bandwidth [7], [8].

Recent real-time neural rendering pipelines, especially splatting-based methods such as 3D Gaussian Splatting (3DGS), enable interactive view synthesis for these monitoring scenarios [9]–[12]. However, a recurring operational failure mode is *temporal instability*—shimmer, flicker, and boundary inconsistency—which can distract analysts, trigger false alarms, or hide weak but critical signals in time-oriented workflows [13]–[16]. For example, in drone-based inspection, thin structures or specular surfaces often exhibit shimmering when successive pose estimates or illumination updates disagree. These artifacts are strongly tied to phase-sensitive structure: small misalignments may preserve magnitude spectra while producing phase inconsistencies that correlate with visible temporal artifacts [17]. This motivates diagnostics that (i) *localize* failure regions for inspection and (ii) remain reliable under real rendering pipelines rather than simplified surrogates.

A key challenge is that diagnostic validity is commonly evaluated under simplified *previous-frame proxy* protocols, where the "prediction" is set to the last observation (e.g., $\hat{I}_t = I_{t-1}$). While convenient, this practice can introduce an **oracle advantage**: temporal information is implicitly leaked into the residual signal, inflating correlations with temporal-instability metrics and masking how diagnostics behave under real 3DGS renders. In other words, a diagnostic can appear strong in proxy evaluation simply because it measures frame-to-frame change, not because it captures rendering failures. Moreover, real 3DGS failures are not always flickery: we empirically observe a structural **decoupling** between temporal instability and reconstruction fidelity, where the renderer can produce *stable-but-incorrect* (over-smoothed) outputs that remain temporally consistent yet perceptually wrong. This decoupling makes proxy-only validation insufficient for monitoring systems that must flag *fidelity* failures, not just flicker.

To address this gap, we present *Coherent Frequency Packet*

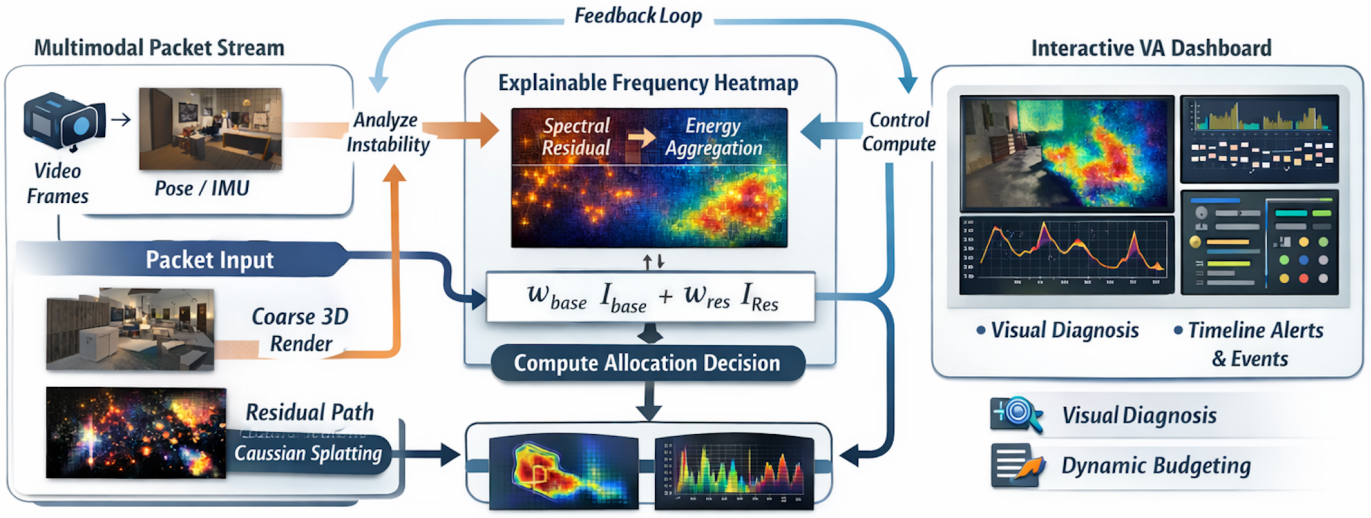

Fig. 1. **DFRS overview.** A multimodal stream (video + pose/IMU + events) is rendered by a fast base path; a phase-sensitive tiled spectral residual yields an *Explainable Frequency Heatmap* that localizes unstable regions and drives closed-loop compute allocation (adaptive sampling and routing to a residual channel). Linked VA views expose overlays, diagnostics, and timeline alerts for analyst inspection and steering.

Splatting for Dynamics (CFPS-DYN), a VA-first diagnostic framework that computes an *Explainable Frequency Heatmap* online via a tiled, phase-sensitive spectral residual between the incoming observation and the current prediction [17]–[20]. The heatmap localizes where the rendered view deviates from the observation in a frequency- and phase-aware manner, producing an interpretable overlay that supports inspection. We summarize the heatmap into a per-frame diagnostic energy for timeline ranking and triage, enabling analysts to retrieve and compare high-risk frames in time-oriented monitoring [16], [21], [22]. While CFPS-DYN is primarily a diagnostic contribution, it is also compatible with stream-adaptive pipelines: the localized signal can be consumed by optional budgeting and routing components to concentrate computation where failures persist [23].

We instantiate the framework with Dynamic Frequency Residual Splatting (DFRS), which integrates packet-based residual representations into splatting-style pipelines [9], [10]. In particular, we study two packet variants: a scalar packet energy (magnitude accumulation) and a coherent packet energy (complex-valued accumulation with interference). Crucially, we evaluate diagnostics under both proxy protocols and *real 3DGS renders*, explicitly testing transfer once the oracle advantage is removed. This evaluation yields a diagnostic "failure taxonomy" over temporal instability and reconstruction error, exposing stable-but-wrong regimes that temporal-only monitoring would miss.

Contributions are as follows:

- We propose the **Explainable Frequency Heatmap**, a tiled phase-sensitive spectral residual that localizes rendering failures as an interpretable diagnostic overlay.
- We identify and quantify an **oracle advantage** in previous-frame proxy evaluation, and empirically show that temporal instability and reconstruction fidelity are decoupled in real 3DGS monitoring.
- We introduce and analyze **packet-based frequency diagnostics**, demonstrating that frequency-domain packet energy remains predictive of reconstruction error (r=0.236) under real renders where pixel baselines collapse (r=0.041).
- We provide a **linked-view VA interface** and an evaluation protocol spanning proxy and real settings, supporting timeline-driven monitoring and intervention.

II. Related Work

Our work lies at the intersection of streaming visual analytics, real-time 3D view synthesis under bounded updates, and frequency-domain diagnostics. Across these areas, a shared challenge is maintaining temporally stable and interpretable evidence when computation must be allocated online under strict per-frame constraints.

A. Streaming Visual Analytics

Visual analytics research emphasizes continuous decision support, where automated analysis and interactive visualization are tightly coupled around monitoring, diagnosis, and intervention tasks. Design study methodology and empirical VA research highlight that such systems are evaluated by how well they support reasoning, validation, and action in real operational contexts, rather than isolated interaction performance [24], [25].

In streaming settings, time-oriented workflows dominate. Analysts scan timelines for change points, validate events, and compare local evidence across neighboring timesteps, motivating dashboards that link a primary view with diagnostics and temporal context [15], [26]. These systems further demonstrate that exposing uncertainty and anomaly localization is critical for trust and triage. Our interface design follows these principles, while extending them by making the diagnostic signal not only visible to the analyst, but also directly consumed by the backend scheduler.

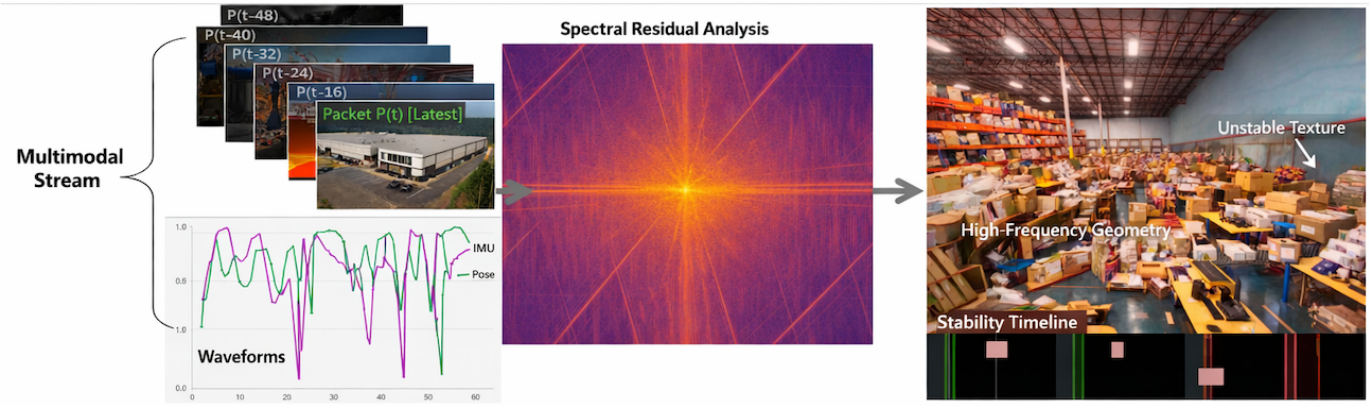

Fig. 2. DFRS treats real-time 3D monitoring as *stream visual analytics*. Left: a multimodal packet stream (video + telemetry such as pose/IMU + optional logs). Middle: a stream-adaptive pipeline where an *Explainable Frequency Heatmap* functions as both (i) an analyst-facing diagnostic overlay and (ii) a control signal for budgeting and routing. Right: heatmap overlays highlight information-dense / unstable regions (motion boundaries, thin structures, specular anomalies), triggering adaptive sampling, residual updates, and alerts on the timeline.

B. Temporal Stability in Real-Time Neural Rendering

Neural view synthesis methods such as NeRF enable high-quality reconstruction and novel-view rendering, but are primarily designed for offline or batch optimization and can be sensitive to sampling and update decisions when deployed in streaming pipelines [27]. Real-time neural graphics systems such as Instant-NGP highlight the importance of computation-aware representations, yet temporal instability artifacts—shimmer, flicker, and boundary inconsistency—remain prominent under bounded updates [28].

Recent splatting-based approaches, most notably 3D Gaussian Splatting, represent scenes using explicit Gaussian primitives and achieve interactive rendering without volumetric ray marching [9]. While highly efficient, such methods can still exhibit temporal instability when updates are sparse, unevenly budgeted, or driven by rapidly changing input streams. Existing work largely addresses stability through representation design or spatial anti-aliasing, whereas our focus is complementary: treating temporal stability as an operational monitoring problem that requires localized diagnosis and proactive compute allocation.

C. Frequency-Domain Diagnostics

Frequency-domain analysis provides compact descriptors of structure and change, and phase-sensitive methods are particularly relevant when instability arises from misalignment rather than magnitude error. Classical phase correlation demonstrates that phase agreement captures translational consistency that magnitude-only measures can miss [17], [29]. This insight motivates diagnostics that are sensitive to temporal coherence rather than static appearance.

Our approach adopts this principle in a streaming VA context. We compute a localized, tiled frequency residual with explicit phase information and expose it as an Explainable Frequency Heatmap. Unlike prior frequency-based saliency or offline analysis, this diagnostic serves a dual role: it provides an interpretable explanation of instability to the analyst and simultaneously acts as a control signal that guides online budgeting and routing under strict per-frame constraints.

III. Methodology

We present DFRS as a closed-loop method for real-time multimodal 3D monitoring. The key idea is to treat the diagnostic as the scheduling primitive: the same signal shown to the analyst is also used to allocate compute under per-frame constraints.

A. Explainable Frequency Heatmap

1) Phase-Sensitive Tiled Spectral Residual: We compute a windowed, overlapping tiled FFT to preserve spatial locality, following short-time Fourier analysis with overlap-add windowing [19], [30], [31]. For each tile u with Hann window $w_u(\mathbf{p})$, define $T_u = w_u \odot I_t$ and $\hat{T}_u = w_u \odot \hat{I}_t$. Let $\mathcal{F}(\cdot)$ be the 2D Fourier transform on the tile grid. We form tile spectra:

$$F_u(f) \triangleq \mathcal{F}(T_u)(f), \qquad \hat{F}_u(f) \triangleq \mathcal{F}(\hat{T}_u)(f). \tag{1}$$

We use a phase-aware residual:

$$\mathcal{R}_u(f) = \mathcal{R}_u^{\text{amp}}(f) + \beta\, \mathcal{R}_u^{\text{phase}}(f). \tag{2}$$

The amplitude term is

$$\mathcal{R}_u^{\text{amp}}(f) = \left| \log\big(\left|\hat{F}_u(f)\right| + \epsilon \big) - \log\big(|F_u(f)| + \epsilon \big) \right|, \tag{3}$$

and the phase term is

$$\mathcal{R}_u^{\text{phase}}(f) = 1 - \Re\left(\frac{\hat{F}_u(f)\, \overline{F_u(f)}}{\big(\left|\hat{F}_u(f)\right| + \epsilon \big)\big(|F_u(f)| + \epsilon \big)} \right). \tag{4}$$

We weight frequencies by $q(f)$ and define tile energy:

$$e_u = \sum_{f \in \Omega} q(f) \left(\mathcal{R}_u^{\text{amp}}(f) + \beta\, \mathcal{R}_u^{\text{phase}}(f) \right). \tag{5}$$

Windowing and averaging follow standard spectral estimation practice [20]. Phase coherence reduces false negatives from misalignment that can preserve magnitude [17].

TABLE I
SYSTEM MODULES AND DATA INTERFACES

Module	Inputs / Outputs	Purpose (VA-facing)
Packet Ingest	I_t, pose/IMU, logs $\to \mathcal{P}_t$	Synchronize modalities; align packets on a timeline.
Base Path	$\mathcal{P}_t \to I_{\text{base}}$	Fast visualization backbone (low-risk regions).
Spectral Residual	$(I_t, \hat{I}_t) \to H_t$	Heatmap: localize where prediction is unstable or hard.
Router + Budgeter	$H_t, \mathbf{v}_t \to w_{\text{base}}, w_{\text{res}}, n(\mathbf{p})$	Allocate compute and expose the rationale.
Residual Packets	High-H_t regions $\to I_{\text{res}}$	Stabilize phase-sensitive high-frequency artifacts.
VA Views	$\hat{I}_t, H_t, s_t, \ell_t$	Linked views: overlay, diagnostics panel, timeline alerts.

2) Heatmap Aggregation and Temporal Smoothing: We splat tile energies back to pixels and apply temporal smoothing:

$$\tilde{H}_t(\mathbf{p}) = \sum_{u:\mathbf{p}\in u} \alpha_u(\mathbf{p})\, e_u, \qquad H_t = (1-\rho)H_{t-1} + \rho\tilde{H}_t, \tag{6}$$

where α_u are overlap weights and ρ controls EMA smoothing [32]. The interface shows H_t as an overlay, while the backend treats H_t as the control input for budgeting.

3) Multimodal Modulation (Telemetry and Logs): We compute a per-frame telemetry anomaly score s_t (e.g., pose jitter spikes, IMU outliers) and optionally parse log events ℓ_t into timeline markers. Telemetry modulates sensitivity:

$$H_t^\star(\mathbf{p}) = H_t(\mathbf{p}) \cdot \left(1 + \alpha\, \frac{s_t - \mu_s}{\sigma_s + \epsilon}\right), \tag{7}$$

which is consistent with streaming change-detection and robust anomaly scoring [33], [34].

B. Problem Setting and System Overview

1) Multimodal Packet Stream: We operate in streaming VA settings where data arrive as time-ordered packets [4], [35]. A packet at time t is $\mathcal{P}_t = (I_t, \boldsymbol{\pi}_t, \boldsymbol{\omega}_t, \tau_t, \ell_t)$, where I_t is the frame, $(\boldsymbol{\pi}_t, \boldsymbol{\omega}_t)$ are telemetry signals, τ_t is the timestamp, and ℓ_t are optional runtime markers. The system maintains a live 3D state and produces $\hat{I}_t$, diagnostics, and online scheduling decisions under per-frame budgets [1], [2], [36].

2) Architecture (Closed Loop): Fig. 3 shows the loop: predict $\hat{I}_t$, compute residual vs. I_t, aggregate into H_t, and use H_t (or $H_t^\star$) to allocate the next update. The objective is stable evidence under strict budgets, measured by monitoring usefulness rather than peak rendering quality [14].

3) Analyst Tasks and VA-facing Interfaces: The interface supports monitoring workflows where the analyst needs to localize instability, understand why compute is concentrated in specific regions, and intervene by adjusting sensitivity or pinning ROIs. We expose H_t in the viewport and time-aligned summaries (e.g., E_t and event markers) in linked views [16], [21], [22], [35], [37]–[39].

C. Stream-Adaptive Routing with Coherent Residual Packets

1) Motivation and Residual Channel: Scalar accumulation in splatting-style pipelines can behave like a low-pass mixture for phase-sensitive microstructure, producing shimmer and flicker in thin structures and specular events. DFRS activates a residual channel only where $H_t^\star$ indicates persistent difficulty, keeping low-risk regions on the base path to preserve throughput.

2) Coherent Residual Packets: We use a compact complex-valued residual parameterization based on Gaussian primitives.

Definition 1 (Coherent Frequency Packet)**.** *A coherent frequency packet is a complex-valued Gaussian wave packet parameterized by mean $\boldsymbol{\mu} \in \mathbb{R}^3$, covariance $\Sigma \succ 0$, carrier $\mathbf{k}_0 \in \mathbb{R}^3$, phase $\phi \in \mathbb{R}$, and amplitude $A \in \mathbb{R}_+$:*

$$\begin{aligned}\psi(\mathbf{x}) = A \exp\!\big(-\tfrac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})\big) \\ \cdot \exp\!\big(i\big(\mathbf{k}_0^\top(\mathbf{x}-\boldsymbol{\mu}) + \phi\big)\big)\,.\end{aligned} \tag{8}$$

3) Coherent Accumulation: Let a ray be sampled at points $\{\tilde{\mathbf{x}}_i\}_{i=1}^N$ with weights $\{\omega_i\}$. We accumulate complex amplitudes and convert to intensity:

$$\Psi_{\text{ray}} = \sum_{i=1}^{N} \omega_i \sum_{j\in\mathcal{A}(\tilde{\mathbf{x}}_i)} \psi_j(\tilde{\mathbf{x}}_i), \tag{9}$$

$$I_{\text{res}} = |\Psi_{\text{ray}}|^2\,. \tag{10}$$

4) Hybrid Routing: We blend a fast base path (I_{base}) with the coherent residual:

$$\hat{I}_t(\mathbf{p}) = w_{\text{base}}(\mathbf{p})\, I_{\text{base}}(\mathbf{p}) + w_{\text{res}}(\mathbf{p})\, I_{\text{res}}(\mathbf{p}). \tag{11}$$

Routing logits and weights are:

$$\begin{aligned}\mathbf{z}_t(\mathbf{p}) &= g\big(H_t^\star(\mathbf{p}),\, \mathbf{v}_t\big), \\ \big[w_{\text{base}}(\mathbf{p}),\, w_{\text{res}}(\mathbf{p})\big] &= \mathrm{softmax}\big(\mathbf{z}_t(\mathbf{p})\big).\end{aligned} \tag{12}$$

The router increases w_{res} where $H_t^\star$ is high, so the same overlay used by the analyst directly drives compute allocation.

5) Heatmap-Driven Budgeting and Capacity Management: We couple diagnostics to three online decisions under bounded per-frame budgets. Sampling is allocated by

$$n(\mathbf{p}) = \mathrm{clip}\left(n_0 \left[1 + \eta \frac{H_t^\star(\mathbf{p})}{\overline{H_t^\star} + \epsilon}\right], n_{\min}, n_{\max}\right). \tag{13}$$

Residual packet parameters are updated more frequently for high-H tiles. Tiles with persistently high H spawn new packets, while low-contribution packets are pruned using $\mathcal{C}_j = \sum_i |\psi_j(\tilde{\mathbf{x}}_i)|^2$ under a threshold for L iterations.

IV. EXPERIMENTS AND ABLATIONS

We evaluate DFRS as a frequency-domain diagnostic for rendered-view quality assessment. Our key question is whether conclusions drawn under a previous-frame *proxy* transfer to *real 3DGS* renders, where the oracle advantage (temporal leakage) is removed.

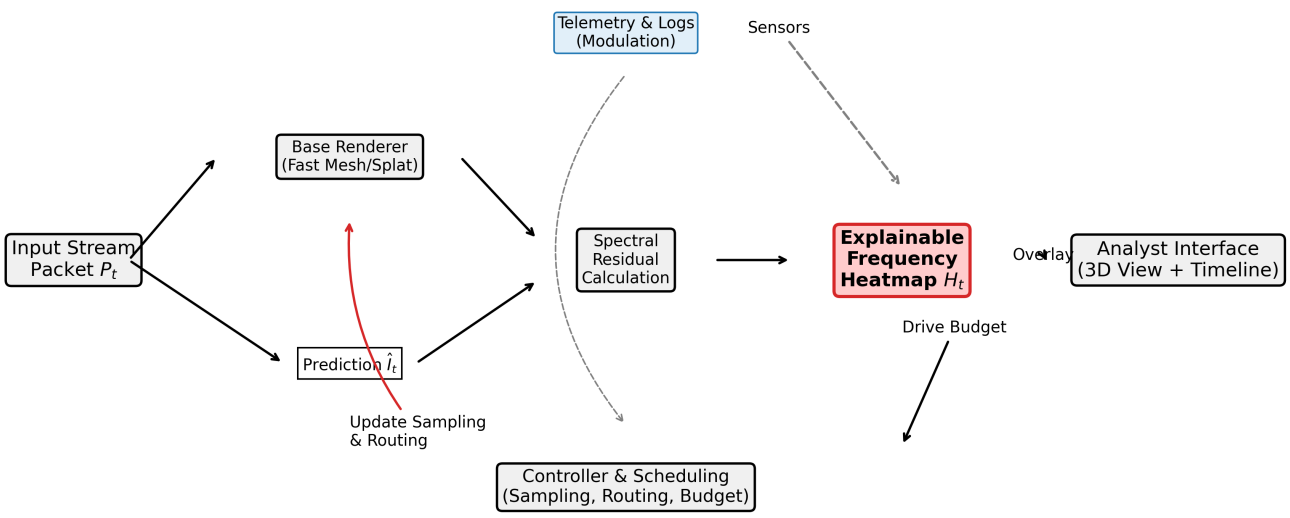

Fig. 3. System overview. Packets feed a hybrid 3D visualization backend. A tiled spectral residual produces an Explainable Frequency Heatmap H_t, displayed to analysts and used to control sampling, routing, and residual packet management. Telemetry and log signals contribute to timeline alerts and can modulate heatmap sensitivity.

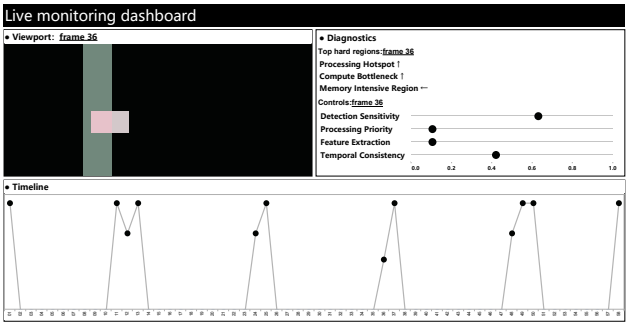

Fig. 4. Interactive visual analytics dashboard. The interface comprises three coordinated views: (a) Viewport, (b) Diagnostics panel, and (c) Timeline analysis.

A. Datasets and Evaluation Settings

We evaluate on three TUM RGB-D sequences [40] (RGB frames at 640×480) under two complementary settings (Table II). In the *proxy* setting, the prediction is the previous frame ($\hat{I}_t = I_{t-1}$), a common simplified evaluation protocol. In the *real 3DGS* setting, $\hat{I}_t$ is rendered by a Splatfacto model [41] trained on the sequence, removing any temporal overlap between prediction and observation.

B. Evaluated Methods

We compare five diagnostic representations, all computed from the same observation–prediction pair $(I_t, \hat{I}_t)$ with identical tiling (32×32, stride 16) and temporal smoothing:

TABLE II
DATASET AND EVALUATION SUMMARY.

Sequence	Frames	Setting	Prediction $\hat{I}_t$
TUM fr1_desk	612	Proxy	Previous frame I_{t-1}
TUM fr1_360	755	Proxy	Previous frame I_{t-1}
TUM fr1_room	549	Real 3DGS	Splatfacto render

- **M0 – Pixel residual**: spatial absolute difference $|I_t - \hat{I}_t|$, no frequency transform.
- **M1 – FFT amplitude**: log-magnitude residual $\big|\log|F_{\text{obs}}| - \log|F_{\text{pred}}|\big|$, radially weighted toward high frequencies.
- **M2 – Phase coherence**: normalised cross-spectrum $1 - \mathrm{Re}\big(F_{\text{pred}}\,\overline{F_{\text{obs}}}/|F_{\text{pred}}|\,|F_{\text{obs}}|\big)$.
- **M3a – Packet (scalar)**: per-tile sum of radially-weighted magnitude residuals accumulated across overlapping windows—the scalar energy of our coherent frequency packet without interference.
- **M3b – Packet (coherent)**: complex-valued accumulation with constructive/destructive interference across overlapping tiles, followed by magnitude extraction.

All heatmaps are summarised into a per-frame scalar energy E_t via top-10% mean aggregation (Sec. IV-G2).

C. Targets and Metrics

We evaluate E_t against two complementary targets:
Target A – Reconstruction error:

$$D_t^{\text{recon}} = \text{LPIPS}(\hat{I}_t,\ I_t), \tag{14}$$

TABLE III
PROXY EVALUATION (TUM FR1_DESK): E_t VS. TEMPORAL INSTABILITY D_t^{temp}.

Method	Pearson r	Spearman ρ	Hit@10%
M0: Pixel	**0.920**	**0.901**	**0.742**
M1: FFT amplitude	0.573	0.715	0.548
M2: Phase coherence	−0.500	−0.548	0.000
M3a: Packet scalar	0.612	0.715	0.548
M3b: Packet coherent	−0.479	−0.504	0.000

measuring per-frame fidelity against the ground-truth observation.

Target B – Temporal instability:

$$D_t^{\text{temp}} \;=\; \text{LPIPS}(\hat{I}_t,\; \hat{I}_{t-1}), \tag{15}$$

measuring perceptual flicker between consecutive predictions.

We report Pearson r, Spearman ρ, and Hit@Top-10% (fraction of true top-10% target events retrieved by the top-10% of E_t).

D. Proxy Evaluation

Table III reports diagnostic correlation with temporal instability under the proxy setting on TUM fr1_desk. The pixel baseline (M0) achieves r=0.920, far exceeding all frequency-domain methods. Scalar packet accumulation (M3a) and FFT amplitude (M1) reach moderate positive correlation ($r\approx$0.6), while phase-based methods (M2, M3b) show negative correlation.

This dominance of M0 is expected: in the proxy setup the prediction *is* the previous frame, so the pixel residual directly encodes the temporal change that LPIPS also measures. We call this the **oracle advantage**—the proxy leaks the temporal signal into the input, inflating correlation for any method sensitive to frame differences.

E. Real 3DGS Evaluation: Oracle Advantage Removal

On TUM fr1_room, where $\hat{I}_t$ comes from a trained Splatfacto model, the oracle advantage vanishes: the prediction no longer contains the previous frame's content.

1) Temporal Instability (Target B): Table IV shows that *all* methods exhibit negative correlation with temporal instability under the real setting. This reveals a structural property of 3DGS: the model tends to produce **stable-but-incorrect** renders (blur, over-smoothing) in high-error regions, while low-error regions exhibit more temporal variation from view-dependent effects and fine geometric detail.

2) Reconstruction Quality (Target A): The more relevant question for quality assessment is whether the diagnostic can identify frames with poor reconstruction fidelity. Table V correlates E_t with per-frame reconstruction error D_t^{recon}.

Two findings emerge. First, the pixel baseline collapses from r=0.920 (proxy) to r=0.041 (real, p=0.34, not significant), confirming that its proxy-setting dominance was entirely due to the oracle advantage. Second, packet scalar energy (M3a) achieves the strongest positive correlation with reconstruction error (r=0.236, p=2.3×10^{-8}), followed by FFT amplitude (r=0.110, p=0.01). This demonstrates that frequency-domain diagnostics maintain predictive power for reconstruction quality after the oracle advantage is removed—precisely the setting relevant to deployed monitoring systems.

TABLE IV
REAL 3DGS (TUM FR1_ROOM): E_t VS. TEMPORAL INSTABILITY. CORRELATIONS FLIP NEGATIVE, REVEALING THAT RECONSTRUCTION ERROR AND TEMPORAL INSTABILITY ARE *decoupled* IN REAL RENDERS.

Method	Pearson r	Spearman ρ	Hit@10%
M0: Pixel	−0.184	−0.241	0.109
M1: FFT amplitude	−0.160	−0.282	0.073
M2: Phase coherence	−0.103	−0.093	0.091
M3a: Packet scalar	−0.195	−0.280	0.109
M3b: Packet coherent	−0.556	−0.638	0.000

TABLE V
REAL 3DGS (TUM FR1_ROOM): E_t VS. RECONSTRUCTION ERROR. FREQUENCY METHODS OUTPERFORM THE PIXEL BASELINE WHEN THE ORACLE ADVANTAGE IS REMOVED.

Method	Pearson r	p-value
M0: Pixel	0.041	0.34
M1: FFT amplitude	0.110	0.01
M2: Phase coherence	−0.080	0.06
M3a: Packet scalar	**0.236**	2.3×10^{-8}
M3b: Packet coherent	−0.112	0.01

3) Summary: Proxy vs. Real: Table VI consolidates the three evaluation columns. The systematic sign flip from proxy to real reveals that proxy-based evaluation overestimates diagnostic validity through temporal signal leakage. Under real 3DGS renders, temporal instability and reconstruction fidelity are fundamentally decoupled, and frequency-domain methods—particularly scalar packet energy—remain effective for the reconstruction quality task that proxy evaluation obscures.

F. Diagnostic Failure Taxonomy

To make the decoupling visually explicit, Fig. 5 plots each frame in a 2D diagnostic space: $x = D_t^{\text{temp}}$ (temporal instability) and $y = D_t^{\text{recon}}$ (reconstruction error), with points coloured by packet coherent energy (M3b). The plot reveals that real 3DGS frames concentrate in the *stable-but-wrong* quadrant (low instability, high error), confirming that the dominant failure mode is over-smoothing rather than flicker. M3b assigns high energy to these frames, demonstrating that the coherent packet representation is sensitive to phase-structure anomalies characteristic of blur—a failure mode invisible to temporal-instability-only monitoring.

G. Ablations

1) Phase Term Contribution: To isolate the contribution of the phase residual (Eq. (4)), we compare the full phase-sensitive diagnostic (β=0.5) against an amplitude-only variant (β=0) on both proxy TUM sequences. As shown in Table VII, removing the phase term consistently degrades Pearson and

TABLE VI
ORACLE ADVANTAGE REMOVAL: PROXY VS. REAL 3DGS. PROXY INFLATES CORRELATION VIA TEMPORAL LEAKAGE; REAL EVALUATION REVEALS THE TRUE DIAGNOSTIC LANDSCAPE.

	Proxy	Real 3DGS	
Method	r (instab.)	r (instab.)	r (recon.)
M0: Pixel	+0.920	−0.184	0.041
M1: FFT amplitude	+0.573	−0.160	+0.110
M2: Phase coherence	−0.500	−0.103	−0.080
M3a: Packet scalar	+0.612	−0.195	**+0.236**
M3b: Packet coherent	−0.479	−0.556	−0.112

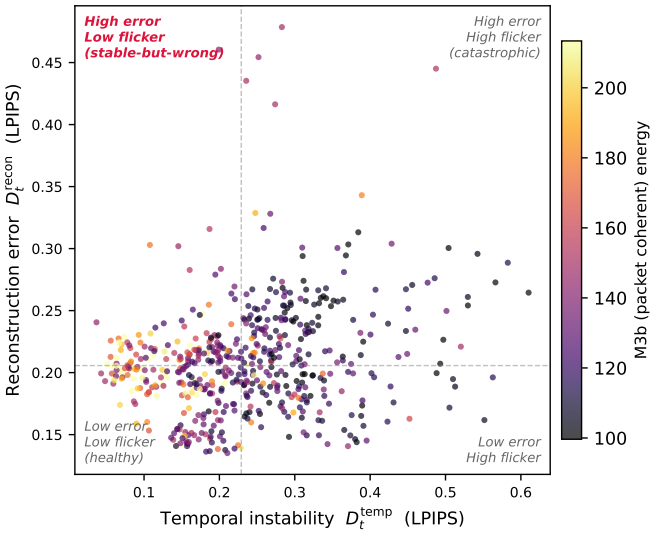

Fig. 5. Diagnostic failure taxonomy on TUM fr1_room (real 3DGS). Each dot is one frame. The spread across quadrants confirms that temporal instability and reconstruction error are decoupled. The *stable-but-wrong* quadrant (upper-left) contains frames that flicker-only monitoring would miss.

Spearman correlation. The improvement is modest but systematic, confirming that phase information captures instability modes—particularly shimmer from sub-pixel misalignment—that amplitude-only residuals miss.

2) Aggregation Strategy and Pilot Monitoring Study: We validate the aggregation design on a 59-frame in-house monitoring stream (Table VIII). Using a global mean over H_t suppresses sparse high-risk signals and reverses the diagnostic direction (r=−0.592), while top-10% mean aggregation yields positive correlation (r=0.625) and practical retrieval performance (AUC=0.739). A top-p% sweep confirms that moderate tail aggregation (p=10–20%) is robust (Table IX). The lag correlation analysis further shows that the diagnostic energy leads measured instability by approximately one frame (k=1, r=0.673), supporting its use as a proactive scheduling signal in the visual analytics interface.

H. Discussion

The proxy-to-real comparison yields three contributions:

First, proxy-based validity conclusions are unreliable. The previous-frame proxy inflates correlation by leaking temporal information, producing misleadingly high scores (Pixel r=0.920) that collapse under real rendering (r=0.041, not significant).

Second, temporal instability and reconstruction fidelity are structurally decoupled in 3DGS. A monitoring dashboard that relies solely on flicker detection will miss the dominant failure mode—stable-but-incorrect renders caused by over-smoothing (Fig. 5).

Third, frequency-domain diagnostics maintain predictive power for reconstruction quality when the oracle advantage is removed. Packet scalar energy (r=0.236, p<10^{-7}) provides a lightweight, interpretable signal that the visual analytics interface exposes as heatmap overlays, timeline rankings, and scheduling priorities—capabilities absent from opaque pixel-domain residuals.

TABLE VII
PHASE ABLATION (β=0 VS. β=0.5) UNDER PROXY EVALUATION.

	TUM fr1_360		TUM fr1_desk	
Metric	β=0	β=0.5	β=0	β=0.5
Pearson r	0.449	**0.463**	0.725	**0.733**
Spearman ρ	0.563	**0.585**	0.826	**0.831**
Hit@Top-10%	0.276	0.276	0.532	0.532

TABLE VIII
AGGREGATION PILOT ON A 59-FRAME MONITORING STREAM: DIAGNOSTICS VALIDITY AND EVENT RETRIEVAL.

Metric	Value
# Frames (N)	59
Event threshold quantile	0.9
Pearson correlation (r)	0.625 ($p = 1.21 \times 10^{-7}$)
Spearman correlation (ρ)	0.392 ($p = 2.14 \times 10^{-3}$)
Event retrieval AUC (rank by E_t)	0.739
Precision / Recall @ top-10%	0.600 / 0.500
Precision / Recall @ top-5%	0.667 / 0.333
Lag peak (Pearson)	$k = 1$, $r = 0.673$

V. INTERACTIVE VISUAL ANALYTICS INTERFACE

The DFRS interface (Fig. 4) exposes the diagnostic–control loop through three coordinated views, following the visual information-seeking mantra [13] and coordinated multiple views (CMV) principles [21], [22]. The design is organized around three primary monitoring tasks that connect the Explainable Frequency Heatmap to analyst reasoning and intervention.

a) Viewport Overlay (T1: Localize Instability): The primary view renders the current prediction $\hat{I}_t$ with the Explainable Frequency Heatmap H_t composited as a semi-transparent overlay. Saturated regions highlight spatial clusters where the stream is information-dense or temporally unstable. This design makes the scheduler's internal allocation of residual-path compute directly legible, providing the rationale behind the system's resource distribution without costly context switching.

b) Diagnostics Panel (T2: Attribute and Validate): This view supports rapid triage by tracking per-tile energy trends

TABLE IX
LEAD–LAG CORRELATION AND AGGREGATION ABLATIONS ($H_t \to E_t$) ON THE 59-FRAME PILOT STREAM.

Item	Value
Lag correlation (Pearson r for E_{t+k} vs D_t)	$k = -3, -2, -1, 0, 1, 2$ $r = 0.354, 0.420, 0.522, 0.625, 0.673, 0.654$
Aggregation ablation	Global mean over H_t: $r = -0.592$, $\rho = -0.392$, AUC $= 0.258$ Top-10% mean over H_t: $r = 0.625$, $\rho = 0.392$, AUC $= 0.739$
Top-p% sweep (Pearson r)	$p = 5\%, 10\%, 20\%$ $r = -0.592, 0.625, 0.595$

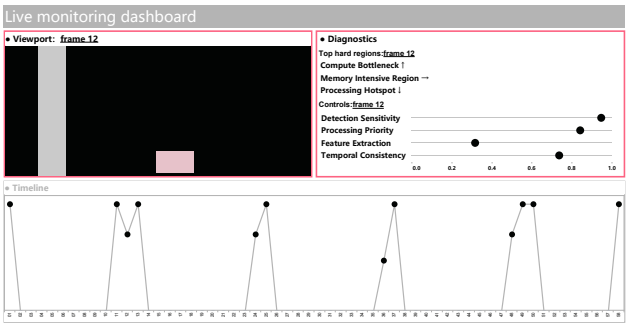

Fig. 6. Dashboard state when frame 12 is selected. Note the pink detection marker in the Viewport and the energy trends: tile #1 decreasing, tile #2 stable, tile #3 increasing.

and normalized control-feature cues [42]. By observing these trends, analysts can distinguish transient spikes from persistent drift. This facilitates the validation of scheduling interventions, confirming whether the activated residual paths successfully mitigate instability over successive frames.

c) Timeline and Alerts (T3: Identify Episodes and Retrieve): A temporal view plots the per-frame diagnostic energy E_t to provide a macro-level overview of stream health, consistent with established time-oriented monitoring workflows [16]. Selecting peaks in the timeline synchronizes the entire system state to that specific timestep, enabling a rapid transition from episodic anomalies to localized spatial evidence and diagnostic context.

d) Linked Interaction: The views share a unified selection state. This linkage ensures that an event detected in the timeline is immediately contextualized by its spatial heatmap and diagnostic trends, supporting forensic inspection across temporal points (e.g., comparing Fig. 6 and Fig. 7) without repetitive manual reconfiguration [43].

VI. CONCLUSION

We presented CFPS-DYN, a visual analytics framework that couples frequency-domain diagnostics with a linked-view monitoring interface for multimodal 3D streaming pipelines. Our proxy-versus-real evaluation on TUM RGB-D sequences reveals that previous-frame proxies create an **oracle advantage** that inflates diagnostic validity (pixel r=0.920) and collapses under real 3DGS renders (r=0.041); more importantly for monitoring practice, temporal instability and reconstruction fidelity are structurally decoupled in real renders, meaning that dashboards relying solely on flicker detection will miss the dominant "stable-but-wrong" failure mode. Frequency-domain packet scalar energy remains a reliable predictor of reconstruction error in this regime (r=0.236, $p<10^{-7}$), and the linked-view interface exposes this signal as heatmap overlays, timeline prioritization, and a diagnostic failure taxonomy that enables analysts to distinguish rendering errors from geometric misalignment—capabilities absent from scalar pixel-residual displays. Ablations confirm that the phase term and top-tail aggregation are both essential design choices. Future work will validate on additional 3DGS architectures, investigate adaptive tiling for the coherent packet representation, and conduct formal user studies to measure how the interface supports triage and intervention decisions in deployed inspection workflows.

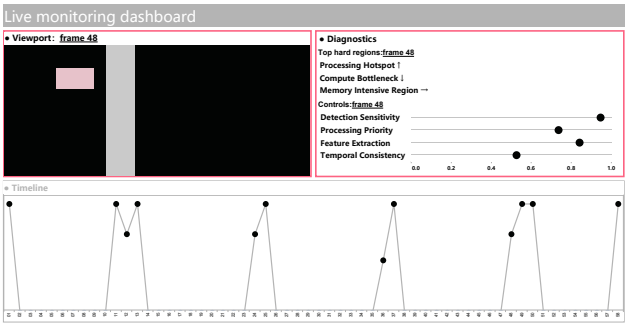

Fig. 7. Interactive coordination mechanism demonstration. These figures show how selecting different frames in the Timeline (bottom panel) triggers synchronized updates in all dashboard components. The Viewport displays frame-specific detection results, the Diagnostics panel shows energy trends for three processing tiles, and the Controls section presents parameter configurations. The system maintains the last selected state, enabling efficient comparative analysis across frames.

REFERENCES

[1] D. A. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, "Visual analytics: Definition, process, and challenges," in *Information Visualization: Human-Centered Issues and Perspectives*, ser. Lecture Notes in Computer Science. Springer, 2008, vol. 4950, pp. 154–175.

[2] J. Heer and B. Shneiderman, "Interactive dynamics for visual analysis," *Communications of the ACM*, vol. 55, no. 4, pp. 45–54, 2012.

[3] T. Munzner, *Visualization Analysis and Design*. CRC Press, 2014.

[4] J. J. Thomas and K. A. Cook, Eds., *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society, 2005.

[5] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[6] A. Fuller, Z. Fan, C. Day, and C. Barlow, "Digital twin: Enabling technologies, challenges and open research," 2019.

[7] M. Stonebraker, U. Çetintemel, and S. Zdonik, "The 8 requirements of real-time stream processing," *SIGMOD Record*, vol. 34, no. 4, pp. 42–47, 2005.

[8] T. Akidau, R. Bradshaw, C. Chambers, S. Chernyak, R. J. Fernandez-Moctezuma, R. Lax, S. McVeety, D. Mills, F. Perry, E. Schmidt, and S. Whittle, "The dataflow model: A practical approach to balancing correctness, latency, and cost in massive-scale, unbounded, out-of-order

data processing," *Proceedings of the VLDB Endowment*, vol. 8, no. 12, pp. 1792–1803, 2015.

[9] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics*, vol. 42, no. 4, 2023.

[10] Z. Yu, A. Chen, B. Huang, T. Sattler, and A. Geiger, "Mip-splatting: Alias-free 3d gaussian splatting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2024, pp. 19 447–19 456.

[11] M. Zwicker, H. Pfister, J. van Baar, and M. Gross, "Surface splatting," in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 2001, pp. 371–378.

[12] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, and P. P. Srinivasan, "Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 5855–5864.

[13] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Proceedings of the IEEE Symposium on Visual Languages*. IEEE, 1996, pp. 336–343.

[14] J. J. van Wijk, "The value of visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 11, no. 1, pp. 79–86, 2005.

[15] N. Cao, C. Lin, Q. Zhu, Y.-R. Lin, X. Teng, and X. Wen, "Voila: Visual anomaly detection and monitoring with streaming spatiotemporal data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 23–33, 2018.

[16] W. Aigner, S. Miksch, H. Schumann, and C. Tominski, *Visualization of Time-Oriented Data*. Springer, 2011.

[17] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529–541, 1981.

[18] X. Hou and L. Zhang, "Saliency detection via spectral residual," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2007, pp. 1–8.

[19] F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, 1978.

[20] P. D. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[21] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky, "Guidelines for using multiple views in information visualization," in *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '00)*. ACM, 2000, pp. 110–119.

[22] C. North and B. Shneiderman, "Snap-together visualization: A user interface for coordinating visualizations via relational schemata," in *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '00)*. ACM, 2000, pp. 128–135.

[23] T. Hachisuka, W. Jarosz, R. P. Weistroffer, K. Dale, G. Humphreys, M. Zwicker, and H. W. Jensen, "Multidimensional adaptive sampling and reconstruction for ray tracing," *ACM Transactions on Graphics*, vol. 27, no. 3, 2008.

[24] M. Sedlmair, M. Meyer, and T. Munzner, "Design study methodology: Reflections from the trenches and the stacks," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2431–2440, 2012.

[25] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale, "Empirical studies in information visualization: Seven scenarios," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 9, pp. 1520–1536, 2012.

[26] J. C. Roberts, "State of the art: Coordinated and multiple views in exploratory visualization," in *Proceedings of the International Conference on Coordinated and Multiple Views in Exploratory Visualization (CMV)*. IEEE, 2007, pp. 61–71.

[27] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," in *European Conference on Computer Vision (ECCV)*, 2020, pp. 405–421.

[28] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Transactions on Graphics*, vol. 41, no. 4, 2022.

[29] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," in *Proceedings of the IEEE International Conference on Cybernetics and Society*, 1975.

[30] D. Gabor, "Theory of communication," *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, vol. 93, no. 26, pp. 429–457, 1946.

[31] J. B. Allen and L. R. Rabiner, "A unified approach to short-time fourier analysis and synthesis," *Proceedings of the IEEE*, vol. 65, no. 11, pp. 1558–1564, 1977.

[32] R. G. Brown, *Statistical Forecasting for Inventory Control*. McGraw-Hill, 1959.

[33] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, 1993.

[34] P. J. Rousseeuw, M. Hubert, and H. Seltman, "Anomaly detection by robust statistics," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2017.

[35] D. A. Keim, J. Kohlhammer, G. Ellis, and F. Mansmann, Eds., *Mastering the Information Age: Solving Problems with Visual Analytics*. Eurographics Association, 2010.

[36] T. Munzner, "A nested model for visualization design and validation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 921–928, 2009.

[37] S. K. Card, J. D. Mackinlay, and B. Shneiderman, *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, 1999.

[38] H. Hochheiser and B. Shneiderman, "Dynamic query tools for time series data sets: Timebox widgets for interactive exploration," *Information Visualization*, vol. 3, no. 1, pp. 1–18, 2004.

[39] M. Monroe, R. Lan, J. Morales del Olmo, B. Shneiderman, and C. Plaisant, "The challenges of specifying intervals and absences in temporal queries: A graphical language for event sequences," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2689–2698, 2013.

[40] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, "A benchmark for the evaluation of RGB-D SLAM systems," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012, pp. 573–580. [Online]. Available: https://vision.in.tum.de/data/datasets/rgbd-dataset

[41] M. Tancik, E. Weber, E. Ng, R. Li, B. Yi, T. Wang, A. Kristofferson, J. Austin, K. Salahi, A. Ahuja, D. McAllister, J. Kerr, and A. Kanazawa, "Nerfstudio: A modular framework for neural radiance field development," in *ACM SIGGRAPH 2023 Conference Proceedings*, 2023. [Online]. Available: https://docs.nerf.studio/

[42] C. Tominski, *Interaction for Visualization*, ser. Synthesis Lectures on Visualization. Morgan & Claypool Publishers, 2015.

[43] J. C. Roberts, "Coordinated and multiple views in exploratory visualization," *Information Visualization*, vol. 6, no. 1, pp. 61–71, 2007.