Clarity Without Reconstruction: Trust-Guided Temporal Fusion for Perceptual Enhancement

Andrew Chen
ajychen@stanford.edu
Stanford University
Stanford, CA, USA

KEYWORDS

video processing, burst photography, temporal fusion, perceptual enhancement

ACM Reference Format:
Andrew Chen. 2026. Clarity Without Reconstruction: Trust-Guided Temporal Fusion for Perceptual Enhancement. In *Proceedings of EE367 Winter 26 Project Proposal (EE367 Winter 2026).* ACM, New York, NY, USA, 2 pages. https://doi.org/XXXXXXX.XXXXXXX

1 MOTIVATION

In video conferencing scenarios with distant capture, faces often occupy only a small region of the frame (e.g., ~200×150 pixels), making facial details difficult to perceive even when the overall stream is high resolution. Importantly, perceived clarity is not determined solely by spatial resolution: Human visual perception relies strongly on temporal stability and edge consistency. These time-related cues should be able to be exploited for video imaging systems: structures that are stable across time and exhibit consistent contrast are more easily perceived as sharp and interpretable, even when fine spatial detail is limited.

Modern camera pipelines exploit temporal redundancy via multi-frame fusion, often in the form of *burst photography*, where a short fixed window of frames is aligned and aggregated to improve image quality without synthesizing detail. In contrast, *temporal reconstruction* methods for video operate online by reusing a reprojected history buffer, but must carefully validate history to avoid ghosting and flicker. Extending this idea to continuous video is challenging because motion, exposure variation, and occlusions make temporal evidence inconsistent. Naively aggregating frames can introduce artifacts such as ghosting, flicker, and unstable sharpening (commonly observed in auto-exposure and HDR pipelines). Therefore, improving perceived clarity in video requires not only leveraging temporal information, but also determining *when* temporal evidence is reliable enough to reuse.

In this project, we aim to investigate whether perceived clarity can be improved by explicitly modeling temporal reliability and using it to guide perceptual enhancement. The aim is to propose a temporal enhancement pipeline that aligns consecutive frames and computes a per-pixel trust map based on photometric agreement, motion magnitude, and edge evidence. This trust map governs temporal accumulation and selectively gates sharpening and contrast enhancement, allowing stable structures to be reinforced while suppressing unstable regions. By extending the principles of multi-frame fusion to continuous video and explicitly modeling temporal reliability, we aim to improve perceived sharpness without reconstructing artificial detail.

EE367 Winter 2026, February 20, 12026, Stanford, CA, USA

2 RELATED WORK

Our work situates between prior research in multi-frame fusion, temporal reconstruction, and human perception of sharpness, all of which exploit temporal and structural cues to improve image quality.

Multi-frame fusion techniques improves image quality by aggregating signal across multiple aligned frames. Hasinoff et al. [3] demonstrated how burst imaging enables noise reduction and improved image quality in handheld cameras, while Wronski et al. [5] extended this concept to recover additional spatial detail through alignment and fusion. These methods rely on confidence-weighted aggregation to improve reconstruction fidelity. However, burst photography typically operates on short, fixed sequences (a finite window) and focuses on reconstruction fidelity, whereas our setting requires online accumulation in continuous video with explicit history validation.

Temporal reconstruction techniques extend multi-frame fusion to continuous image sequences by reusing information from previous frames. In graphics and video processing, temporal anti-aliasing pipelines accumulate reprojected history while suppressing artifacts through reliability estimation based on photometric agreement and motion consistency [6]. Similarly, motion-compensated video enhancement methods align and aggregate frames to improve signal quality while minimizing noise and temporal artifacts [2]. These approaches exploit temporal redundancy but primarily optimize reconstruction accuracy or noise reduction, with reliability treated as an internal mechanism rather than a cue for perceptual enhancement.

Furthermore, human vision research shows that perceived sharpness depends strongly on contrast structure and temporal stability rather than spatial resolution alone. Studies on contrast sensitivity and perceptual image quality demonstrate that temporally stable, high-contrast structures are more likely to be perceived as sharp and interpretable even when spatial detail is limited [1, 4].

3 PROJECT OVERVIEW

Building on ideas from multi-frame fusion and temporal reconstruction, we propose a temporal enhancement pipeline that improves perceived clarity in low-resolution video by explicitly modeling temporal reliability as a per-pixel trust map and using it to control both temporal fusion and perceptual enhancement. While existing

methods typically use reliability internally to improve reconstruction fidelity, our approach treats temporal reliability as a first-class signal for perception-driven enhancement in continuous video. Because video contains dynamic motion and photometric variation, naive temporal averaging can produce ghosting, flicker, and unstable sharpening. Instead, we align temporal evidence to the current frame, perform trust-driven temporal accumulation, and selectively gate sharpening and contrast enhancement based on trust. This reinforces temporally stable structures while suppressing unreliable regions, improving perceived sharpness without aiming to reconstruct ground-truth high-frequency detail.

Given an input video stream, we estimate inter-frame motion and warp a *history buffer* (the previous accumulated output) into the current frame's coordinate system. We use an affine alignment method (with optical flow as an optional extension) and compute a validity mask to identify pixels where the warp is reliable. We then compute a per-pixel *trust map* that measures temporal reliability at each location by combining three cues: (1) **photometric agreement** between the current frame and warped history, (2) **motion magnitude** as a proxy for temporal uncertainty, and (3) **edge evidence** to favor stable spatial structures. Pixels that are well-aligned and temporally consistent receive higher trust, while pixels affected by occlusion, motion boundaries, or lighting changes receive lower trust (or are masked out).

The trust map drives two downstream operations. First, we perform **trust-weighted temporal accumulation**, blending warped history and the current frame proportionally to trust. This improves signal-to-noise and stabilizes reliable structures while reducing history leakage in low-trust regions. Second, we apply **trust-gated perceptual enhancement** (e.g., sharpening and local contrast boost) primarily in high-trust regions, reinforcing edges and stable contrast cues without amplifying unreliable regions that would otherwise produce flicker or halos. We will implement the full pipeline in Python using OpenCV and NumPy and evaluate it on captured video of distant faces. We will also generate diagnostic visualizations (trust maps, motion/difference maps, and residuals) to analyze when and why temporal enhancement succeeds or fails. Our goal is to demonstrate that reliability-guided temporal fusion, coupled with selective enhancement, can improve perceived clarity while maintaining temporal stability in low-resolution video conferencing settings (if our evaluation says so).

4 MILESTONES AND TIMELINE

Week 0: Prototype and Baseline Pipeline

Goal: Establish a working temporal processing prototype.

- Implement video loading, frame extraction, and preprocessing
- Implement inter-frame motion estimation and frame warping
- Implement baseline temporal averaging and per-frame sharpening
- Generate diagnostic outputs such as difference maps and motion maps
- Obtain initial test video sequences

Deliverables:

- Working prototype with temporal alignment and baseline processing
- Diagnostic visualizations (difference maps, motion maps)
- Baseline outputs for comparison

Week 1: Trust-Guided Temporal Fusion

Goal: Develop trust-based fusion and perceptual enhancement.

- Implement per-pixel trust map computation
- Integrate trust-guided temporal accumulation
- Implement trust-gated sharpening and contrast enhancement
- Generate trust map and enhancement visualizations

Deliverables:

- Functional trust-guided temporal enhancement pipeline
- Visualization of trust maps and enhancement behavior
- Initial qualitative comparison against baselines

Week 2: Evaluation, Analysis, and Final Report

Goal: Evaluate performance and finalize project deliverables.

- Evaluate perceptual clarity improvements on captured video
- Analyze trust behavior, stability, and failure cases
- Generate final visual outputs and comparison results
- Prepare final report and poster presentation

Deliverables:

- Final enhanced video results
- Analysis and visualization of trust-guided enhancement
- Completed project report and presentation materials

REFERENCES

[1] F. W. Campbell and J. G. Robson. 1968. Application of fourier analysis to the visibility of gratings. *The Journal of Physiology* 197, 3 (1968), 551–566. https://doi.org/10.1113/jphysiol.1968.sp008574

[2] Kostadin Dabov, Alessandro Foi, and Karen Egiazarian. 2007. Video denoising by sparse 3D transform-domain collaborative filtering. In *2007 15th European Signal Processing Conference*. 145–149.

[3] Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. 2016. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Trans. Graph.* 35, 6, Article 192, 12 pages. https://doi.org/10.1145/2980179.2980254

[4] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13, 4, 600–612. https://doi.org/10.1109/TIP.2003.819861

[5] Bartlomiej Wronski, Ignacio Garcia-Dorado, Manfred Ernst, Damien Kelly, Michael Krainin, Chia-Kai Liang, Marc Levoy, and Peyman Milanfar. 2019. Handheld multi-frame super-resolution. *ACM Trans. Graph.* 38, 4, Article 28 (July 2019), 18 pages. https://doi.org/10.1145/3306346.3323024

[6] Lei Yang, Shiqiu Liu, and Marco Salvi. 2020. A Survey of Temporal Antialiasing Techniques. *Computer Graphics Forum* 39, 2 (2020), 607–621. https://doi.org/10.1111/cgf.14018