A Robust Two-Stage Method for Slide Change Detection using Global Structural Similarity and Temporal Stability Verification

bit-admin

October 4, 2025

Abstract

This report details a robust, two-stage method for detecting meaningful content changes in video streams of presentation slides. The primary challenge is to differentiate significant content updates from transient on-screen events such as mouse movements, animations, or minor video artifacts. We propose a system that combines a high-efficiency change trigger with a temporal stability check. The first stage employs a Global Structural Similarity Index Measure (G-SSIM) to detect potential changes. G-SSIM's inherent insensitivity to small, localized noise makes it an effective first-pass filter. The second stage, triggered by a G-SSIM score drop, initiates a verification phase. This phase repeatedly checks the new slide content for stability over a defined period, ensuring that only final, static slides are captured. This two-stage approach significantly enhances detection accuracy by filtering both localized spatial noise and transient temporal events. We provide the mathematical formulation, a complete pseudocode implementation, and a justification for the method's effectiveness.

1 Introduction

1.1 Problem Statement

The objective is to develop a reliable and computationally efficient algorithm to automatically capture static slides from a dynamic video feed of a presentation. A successful system must distinguish semantically meaningful slide transitions from a variety of transient or insignificant on-screen events, including:

- Localized spatial noise: Mouse cursor movements, blinking text cursors.
- Transient temporal noise: Slide transition animations, slow-loading content, brief pop-ups.
- Minor acquisition noise: Uniformly distributed video compression artifacts.

1.2 Proposed Two-Stage Approach

Standard image similarity metrics often fail to meet these requirements. The sliding-window SSIM is computationally intensive and, by design, highly sensitive to local changes, making it susceptible to false positives from mouse movements. Simple frame differencing is similarly plagued by this issue.

We propose an alternative approach: applying the Structural Similarity Index Measure globally. This method, which we term Global SSIM (G-SSIM), calculates the statistical properties (mean, variance, covariance) across the entire pixel set of two frames. The central hypothesis is that G-SSIM's primary weakness in general image quality assessment—its blindness to local detail—becomes its principal strength for this specific task.

1. **Stage 1: G-SSIM Change Detection.** We use the Global Structural Similarity Index Measure (G-SSIM) as a high-speed, low-complexity trigger. By applying the SSIM calculation to the entire frame, the algorithm effectively ignores localized noise like mouse movements, which have a negligible impact on global image statistics. A significant drop in the G-SSIM score signals a potential slide change and triggers the next stage.

2. **Stage 2: Temporal Stability Verification.** Upon detecting a potential change, the system enters a verification state. It captures the candidate new frame and then, over a short period, repeatedly compares subsequent frames against this candidate. The slide is only confirmed and saved if it remains static for a predefined number of checks. This stage effectively filters out animations and other transient states.

2 Methodology

2.1 Stage 1: Change Detection using G-SSIM

The first stage compares two full image frames, X and Y, by evaluating their global statistical properties.

2.1.1 Mathematical Formulation

The G-SSIM value is a product of three distinct components: luminance (l), contrast (c), and structure (s).

$$\text{G-SSIM}(X, Y) = l(X, Y) \cdot c(X, Y) \cdot s(X, Y)$$

These components are calculated using global statistics of the entire frame. Let N be the total number of pixels in a frame.

1. Global Mean Luminance (μ): The average pixel value across the entire frame.

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{and} \quad \mu_y = \frac{1}{N}\sum_{i=1}^{N} y_i$$

2. Global Variance (σ^2): The variance of pixel values across the entire frame, representing global contrast.

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)^2 \quad \text{and} \quad \sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu_y)^2$$

3. Global Covariance (σ_{xy}): A measure of the joint variation of pixel values between the two frames.

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)$$

Using these global statistics, the final G-SSIM formula is expressed as:

$$\text{G-SSIM}(X, Y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{1}$$

where C_1 and C_2 are stabilization constants to avoid division by zero. They are defined as $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$, where L is the dynamic range of the pixel values (typically 255 for 8-bit grayscale images). Standard values are $K_1 = 0.01$ and $K_2 = 0.03$, yielding:

- $C_1 = (0.01 \times 255)^2 = 2.55^2 \approx 6.5025$
- $C_2 = (0.03 \times 255)^2 = 7.65^2 \approx 58.5225$

2.1.2 Triggering Mechanism

The implementation directly follows the formulation in Equation 1. A potential change is flagged when the G-SSIM score between the last known stable frame (F_{stable}) and the current frame ($F_{current}$) drops below a high-confidence threshold, τ.

$$\text{Potential Change if} \quad \text{G-SSIM}(F_{stable}, F_{current}) < \tau$$

A typical value for τ is 0.999, ensuring that only statistically significant deviations trigger the verification stage.

2.2 Stage 2: Temporal Stability Verification

Once a potential change is detected, the system transitions from a 'DETECTING' state to a 'VERIFYING' state. The logic is as follows:

1. The frame that triggered the change ($F_{current}$) is stored as a candidate, 'potentialNewFrame'.
2. The system then waits for a set interval (e.g., 2 seconds) and captures the next frame, 'verificationFrame'.
3. It calculates the G-SSIM between 'potentialNewFrame' and 'verificationFrame'.
4. **If the score is high (above τ)**, the slide is considered stable for this check. The verification counter increments. This process repeats for a user-defined number of times ('verificationCount').
5. **If the score is low (below τ)**, it implies the slide is still in a transient state (e.g., an animation is still running). The verification fails, and the system resets to the 'DETECTING' state, discarding the candidate.
6. Only after 'verificationCount' successful and consecutive checks is the 'potentialNewFrame' confirmed as a new, stable slide, saved, and set as the new 'lastStableFrame'.

3 Pseudocode Implementation

The following pseudocode outlines the complete logic of the two-stage detection and verification process.

Algorithm 1 Two-Stage Slide Detection Algorithm

```
Initialize:
DETECTION_INTERVAL ← 2.0                                              ▷ seconds
VERIFICATION_COUNT ← 3                                                ▷ number of checks
GSSIM_THRESHOLD ← 0.999

state ← "DETECTING"
last_stable_frame ← CaptureInitialFrame()
potential_new_frame ← null
verification_counter ← 0

procedure MAINLOOP
    while video stream is active do
        Wait(DETECTION_INTERVAL)
        current_frame ← CaptureCurrentFrame()
        if state = "DETECTING" then
            HandleDetection(current_frame)
        else if state = "VERIFYING" then
            HandleVerification(current_frame)
        end if
    end while
end procedure

procedure HANDLEDETECTION(current_frame)
    score ← CalculateGSSIM(last_stable_frame, current_frame)
    if score < GSSIM_THRESHOLD then
        UpdateStatus("Potential change detected. Verifying...")
        state ← "VERIFYING"
        potential_new_frame ← current_frame
        verification_counter ← 1                    ▷ First successful capture is check 1
    end if
end procedure

procedure HANDLEVERIFICATION(current_frame)
    score ← CalculateGSSIM(potential_new_frame, current_frame)
    if score ≥ GSSIM_THRESHOLD then                                   ▷ Slide is stable
        verification_counter ← verification_counter + 1
        UpdateStatus("Verifying... (" & verification_counter & "/" & VERIFICATION_COUNT
& ")")
        if verification_counter ≥ VERIFICATION_COUNT then
            UpdateStatus("Verification successful. Saving slide.")
            SaveSlide(potential_new_frame)
            last_stable_frame ← potential_new_frame
            ResetState()
        end if
    else                                          ▷ Slide is unstable, verification failed
        UpdateStatus("Verification failed. Re-detecting...")
        ResetState()
    end if
end procedure

procedure RESETSTATE
    state ← "DETECTING"
    potential_new_frame ← null
    verification_counter ← 0
end procedure
```

4 Justification and Analysis

4.1 Inherent Insensitivity to Localized Noise

The core strength of this approach lies in its mathematical foundation. Consider a mouse cursor appearing on a 1920x1080 (Full HD) frame. The cursor may alter approximately $32 \times 32 = 1024$ pixels out of a total of over 2 million.

The change in the global mean ($\Delta\mu$) will be infinitesimally small, as the sum of pixel values is dominated by the 99.95% of unchanged pixels. Similarly, the impact on global variance and covariance will be negligible. As a result, the G-SSIM score will remain extremely high (very close to 1.0), effectively and automatically filtering out this noise without requiring complex pre-processing steps like cursor detection and removal.

4.2 Sensitivity to Meaningful Content Changes

Conversely, a meaningful change, such as the addition of a new line of text, alters the frame's statistical profile more significantly. While the area of change may still be small (e.g., 1-2% of the screen), it is structurally correlated. The new text introduces a cluster of pixels with different luminance values compared to the background, systematically shifting the global mean. More importantly, it alters the distribution of pixel intensities, causing a detectable change in the global variance (σ^2) and its relationship with the previous frame (covariance σ_{xy}).

Because the detection threshold τ is set so high (e.g., 0.999), even a subtle but systematic shift in these global statistics is sufficient to push the G-SSIM score below τ, successfully flagging a meaningful event.

4.3 Implemention of two-stage approach

The two-stage approach provides a synergistic defense against false positives.

- **G-SSIM for Spatial Noise:** The G-SSIM trigger acts as a computationally cheap and effective filter for high-frequency spatial noise like mouse movements. Its global nature ensures that such minor, localized pixel changes do not meet the threshold for a potential slide transition.
- **Verification for Temporal Noise:** The stability check addresses the challenge of transient content. Slide animations, fades, and builds are correctly identified as unstable states. The system patiently waits until the slide content "settles" before making a final decision, ensuring that only the intended final version of the slide is captured. This enhances the quality and relevance of the output significantly.

5 Limitations

The robustness of the system relies on the correct configuration of its parameters.

- **Parameter Tuning:** The 'DETECTION_INTERVAL' and 'VERIFICATION_COUNT' must be chosen carefully. A verification period that is too short may fail to outlast slow animations, while a period that is too long could miss a rapid succession of valid slides. Also, the G-SSIM performance is highly dependent on the threshold τ. This value may need empirical tuning based on video resolution, compression quality, and content type.
- **Handling of Large Overlays:** The G-SSIM trigger is still susceptible to large-area, non-content changes (e.g., a full-screen notification), which would be correctly identified as

a stable new "slide" after passing verification. Contextual analysis would be needed to filter such cases.

6 Conclusion

The proposed two-stage method, combining a G-SSIM trigger with a temporal stability verification phase, offers a robust and efficient solution for slide detection in screen recordings. It effectively isolates meaningful content changes from both localized spatial noise and transient temporal events. By leveraging G-SSIM for its computational speed and noise-filtering properties, and complementing it with a logical verification layer, the system achieves a high degree of accuracy and reliability, making it well-suited for automated presentation analysis and archival applications.