Approach: Traditional Feature-Based Image Similarity Analysis
For Identifying Overlapping Scenes and Noise-Augmented Images

1. Problem Statement

A dataset contains images with:

- Overlapping scenes (different but partially overlapping views)
- Same scenes with different noise augmentations

We need to identify similar pairs and assign a similarity score based on that (between 0 to 1).

2. Traditional Approach Overview

Our solution uses feature engineering and descriptor matching with three key phases:

Phase	Components	Purpose
Preprocessing	Grayscale conversion, Noise reduction (Gaussian blur)	Standardize input, reduce noise interference
Feature Extraction	ORB (Oriented FAST and Rotated BRIEF)	Detect Keypoints and describe local patterns
Similarity Calculation	Brute-force matching, Hamming distance, Score normalization	Quantify similarity between images

3. Methodology Breakdown

3.1 Feature Extraction with ORB. Why ORB?

- **Computational Efficiency**: Optimized for real-time use (FAST detector + BRIEF descriptor)
- **Rotation Invariance**: Adds orientation to BRIEF via intensity centroid
- **Scale Awareness**: Built-in pyramid scaling

Key Parameters: nfeatures=2000, scaleFactor=1.2 , nlevels=8 , edgeThreshold=31

3.2 Feature Matching Process

- Brute-Force Matcher: Uses Hamming distance (bitwise XOR count) for binary descriptors

3.3 Similarity Score Calculation

- Normalization: Accounts for variable feature counts between images
- Range: 0 (no similarity) to 1 (perfect match)

4. Preprocessing for Noise Robustness: Gaussian Blur (blur_kernel=(13,13), blur_sigma=7)

- Purpose: Smooth high-frequency noise while preserving edges
- Trade-off: Larger kernels reduce noise but blur details

Result: Score between 0 to 1

Similarity between image 1 and image 2: 0.27

Similarity between image 1 and image 4: 0.9

Similarity between image 2 and image 3: 0.41

Similarity between image 1 and image 5: 0.35

Similarity between image 1 and image 6: 0.35