Team 4

Face Detection using Viola Jones (VJ) Algorithm

Members

- Dilshod Obidov (12225243)
- Mironshoh Muhammedov (1225252)
- Nilufar Kurbonova (12225241)

Github repository link:
https://github.com/dilshod-obidov/vj-face-detection

Face detection

Cascade Classifier Windows GUI

We installed Cascade Classifier Windows GUI and trained cascade with it on a Windows 11 laptop.

Public dataset

- Positive samples (~500) from **Face-Detection-Dataset**: https://www.kaggle.com/datasets/fareselmenshawii/face-detection-dataset/data
- Negative samples (425) from **haarcascade-negatives:** https://github.com/JoakimSoderberg/haarcascade-negatives/

We took around 500 positive images from the home-face-detection dataset and 425 negative images (background) from the haarcascade-negatives dataset for training.

Private dataset

- Dataset of team members consist of around 400 cropped faces from images in different scenarios

Analyzing the accuracy of the VJ detector in the presence of the

⚫ **Viewpoint Variation:** VJ's accuracy drops significantly when detecting non-frontal or rotated faces, as it is optimized for faces directly facing the camera.

⚫ **Deformation**: The detector relies on rigid features, which reduces accuracy with expressions or deformations, limiting its flexibility in capturing varied facial structures.

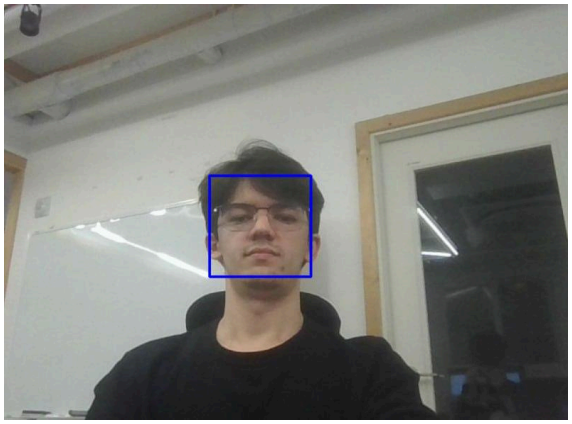
⚫ **Occlusion**: VJ struggles with partially occluded faces, as its algorithm depends on the visibility of specific facial features. Even minor occlusions, like sunglasses, can hinder detection.

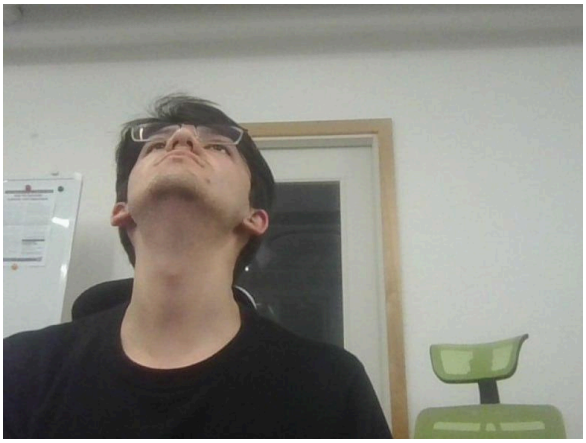
⚫ **Illumination Conditions**: The algorithm is highly sensitive to lighting; it performs best under controlled, even lighting and struggles in dim or overly bright conditions.

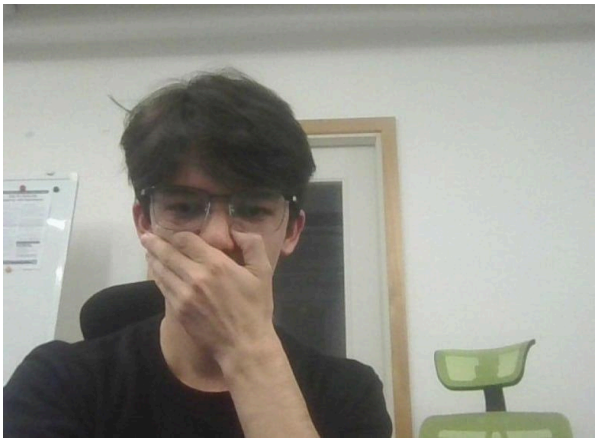
⚫ **Cluttered or Textured Background**: Detection accuracy drops in complex backgrounds, as the VJ detector may confuse background textures with facial features.

⚫ **Intra-class Variation**: VJ's limited training on diverse facial variations means it lacks adaptability to a wide range of appearances, reducing accuracy across different ages, ethnicities, and facial features.

Our program is weaker at detecting faces under challenging conditions like varied viewpoints, deformations, occlusions, lighting changes, complex backgrounds, and diverse face types.







Discussions

❖ About the Team-observations on the evaluations and comment on the performance of our detector.

Team Observations on Evaluations

1. **Detection Speed**:
    - Viola-Jones operates swiftly, especially for face detection, due to its use of an integral image and cascade structure, making it suitable for real-time applications. This speed was consistent across varied resolutions, though higher resolutions slowed down slightly.
2. **False Positives/Negatives**:
    - There were noticeable false positives, especially in complex scenes or cluttered backgrounds. Certain objects (like faces in profile or partially obscured faces) often cause not detecting. Adjusting the threshold, or minimum size, helped reduce this, but also slightly impacted true positive rates. In summary, it mainly detected images that faced directly to the camera, but if it faces any images with Viewpoint variation, Deformation, Occlusion, Illumination conditions, Cluttered or textured Background, Intra-class variation
3. **Performance under Lighting Variations**:
    - Poor lighting significantly impacted the accuracy. Bright lighting conditions produced satisfactory results, while low-light conditions increased false negatives due to poor feature extraction.
4. **Scale Invariance**:
    - The detector struggled with detecting objects at different scales, particularly smaller objects. The performance was better when objects were within a moderate range of sizes.

Performance Comments

- **Strengths**: The Viola-Jones Cascade is efficient for frontal detection tasks in controlled environments, where objects are mostly static and well-lit. This makes it viable for basic applications that prioritize speed over accuracy.
- **Limitations**: The model's reliance on Haar-like features and cascade training limits its versatility in more complex scenes. It's sensitive to lighting, pose, and scale, which restricts its use in uncontrolled environments or applications requiring high accuracy.