Surveillance (Face Recognition)

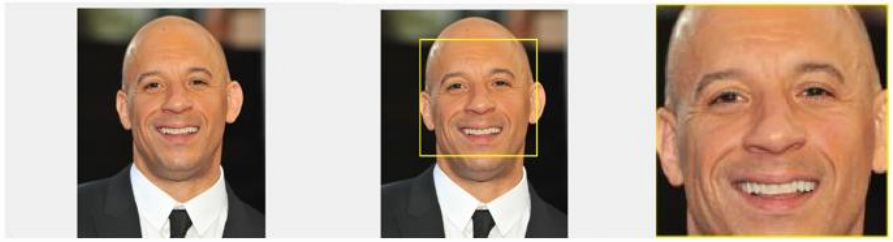
BY: VIBHOR MISHRA AND SHUBHAM MUDGAL

Algorithm

There were five main sections in which the project was divided.

1) Collect multiple images of a set of people to form training dataset which will be labeled and used for classification.
2) Detect the faces in those images
3) Extract features
4) Train classifier using the feature set
5) Test the system using test data

Face Recognition Workflow

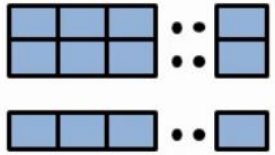


Classifier / Model → vindiesel

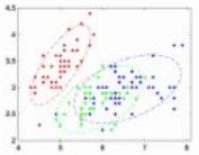
Input Frame → Face Detection → Face Registration → Feature Extraction → Classification

Face Database / Gallery

→ Feature Extraction



→ Learning or Modelling



→ Classification

Recognition of Multiple Faces in one frame

To detect and recognize multiple faces in a video feed, the following steps were used:

- Process first frame from the video which has a face in it and henceforth, process one frame at a time till the end of the video. The detected faces are then cropped and the features are extracted from them.
- For each face, corner points are found which are used to uniquely identify faces. These points and the bounding box IDs are stored to keep track of each face or box in the video frames.
- To check if the detected faces in the frame are new or from the last frame, the area of overlapping box region of already existing and new boxes (faces) is compared to a threshold value.

Continued….

- If the threshold is not satisfied:
  - it's a new face. Compute corner points and store bounding box information enclosing this new face.
  - Extract features and run classification for this new face and store the classified label.
- If the threshold is satisfied:
  - It's an already existing one. Update the points and box information.
  - Don't perform classification for that particular box enclosing face and use the previous prediction.

Continued….

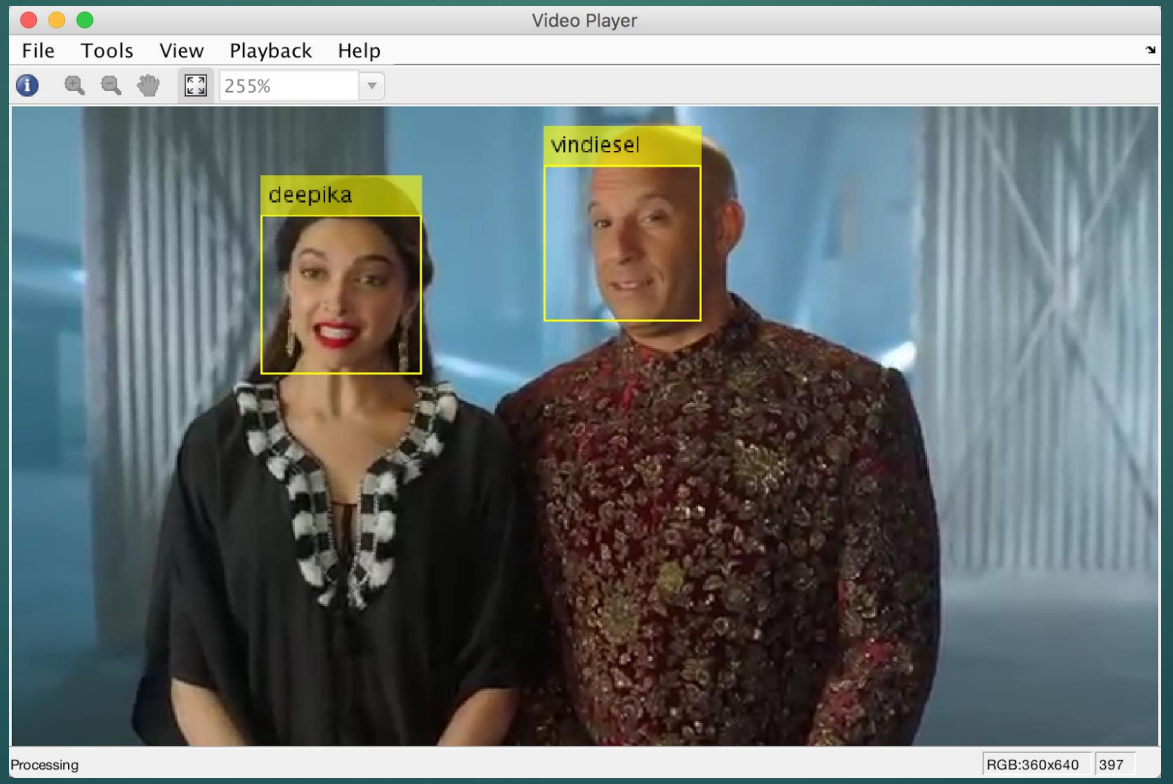
- In many cases, the face movement was very abrupt and the boxes were not matched and deleted from the object. A new box was then assigned to the newly detected old face and process was continued. This approach was used to detect and keep track of multiple faces detected in a video feed.

- One problem with this approach was that once a face is detected and recognised to be of class A, it would then track that face unless the person moves out of the video. If the initial prediction was however not correct, it would propagate the wrong prediction for as long as the person is in the video. Therefore to prevent this flaw, detection and recognition on the current boxes(faces enclosed in boxes) was done every 10 frames. This way the faces detected initially would not propagate to the whole video feed.

Results

- As seen below, for the dataset with 25 images of each class, we get the accuracy of 92% with HOG feature extractor and using SVM classifier.

- On increasing the dataset to huge number for each classes, the accuracy we get in videos is around 82%. With videos, the accuracy is computed by calculating the [time for which the classifier predicted labeled trained class correctly] dividing by [total video time] *100:



Demo

THANK YOU