Video Processing Pipeline

Stage 0 – Input

Output : Video files

Stage 1 - Shot Segmentation

Input : Video files

Output : CSV file with row format:
[Shot start (int) – Start frame number for shot,
Shot end(int) - End frame number for shot]

Stage 2 - Face Detection

Input : CSV file with row format:
[Shot start (int),
Shot end(int)]

Output : CSV file with row format:
[Frame number (int),
Shot number (int),
Shot end frame (int),
Bounding boxes (x1_y1_x2_y2) – where (x1,y1) is top left and (x2,y2) is bottom right point of face bounding box]

Stage 3 - Face Tracking

Input : CSV file with row format:
[Frame number (int), Shot number (int), Shot end frame (int), Bounding boxes (x1_y1_x2_y2)]

Output : CSV file with row format:
[Frame number (int),
Shot number (int),
Face id (int) – unique integer id for face,

x_min (int) – x coordinate for top left point,
y_min (int) - y coordinate for top left point,
x_max (int) - x coordinate for bottom right point,
y_max (int) - y coordinate for bottom right point]

Stage 4 - Face Cropping

Input : CSV file with row format:
[Frame number (int), Shot number (int), Face id (int), x_min (int), y_min (int), x_max (int), y_max (int)]

Output :
1. Video with cropped face

2. CSV file with row format:

[Frame no (int),
Shot no (int),
Face id (int),
Crop_x_min, Crop_y_min, Crop_x_max, Crop_y_max - coordinates of cropped bounding box in original video]

Stage 5 - Face Alignment

Input : **Video with cropped face**

Output : **Video with aligned face**

Stage 6 - 'Feature Extraction

Input : **Video with cropped face (** output of Stage 4 - Face Cropping **)**

Output -
1. Face video with features visualized
2. CSV file with row format:
[frame (int) – frame number in cropped video,
face_id (int) – unique integer face id,
timestamp (float)– frame time in cropped video,
confidence (float)– confidence score of successful feature extraction,

success (int) - binary feature extraction success indicator,
Other extracted features – see https://github.com/TadasBaltrusaitis/OpenFace/wiki/Output-Format for more details **]**

Stage 7 - Face Clustering

Input : **Cropped and aligned face videos (** output of Stage 5 - Face Alignment**)**

Output - CSV file with row format:
[Shot Number (int),
Face id (int),
Identity (int) – unique number denoting identity of person]

Stage 8 – Process Output

Input : **Output of Stage 0, 3, 4, 6**

Output –

1. Video with eye gaze visualization
2. Video with all extracted features visualized. Eg. Face landmarks, eye gaze, head pose, speaker state, identity and selected action units

3. CSV file with row format:
[frame_number (int) – frame number in original video,
shot_number (int) – shot number in video
face_id (int) – unique integer face id,
identity (int) – unique number denoting identity of person,
face_xmin, face_ymin, face_xmax, face_ymax - coordinates of detected face bounding box video,
confidence (float)– confidence score of successful feature extraction,
success (int) - binary feature extraction success indicator,
Other extracted features – see https://github.com/TadasBaltrusaitis/OpenFace/wiki/Output-Format for more details,
isSpeaking (int) – 1 if a person is speaking based on lip movements, 0 otherwise.**]**