CSC420 Project Proposal

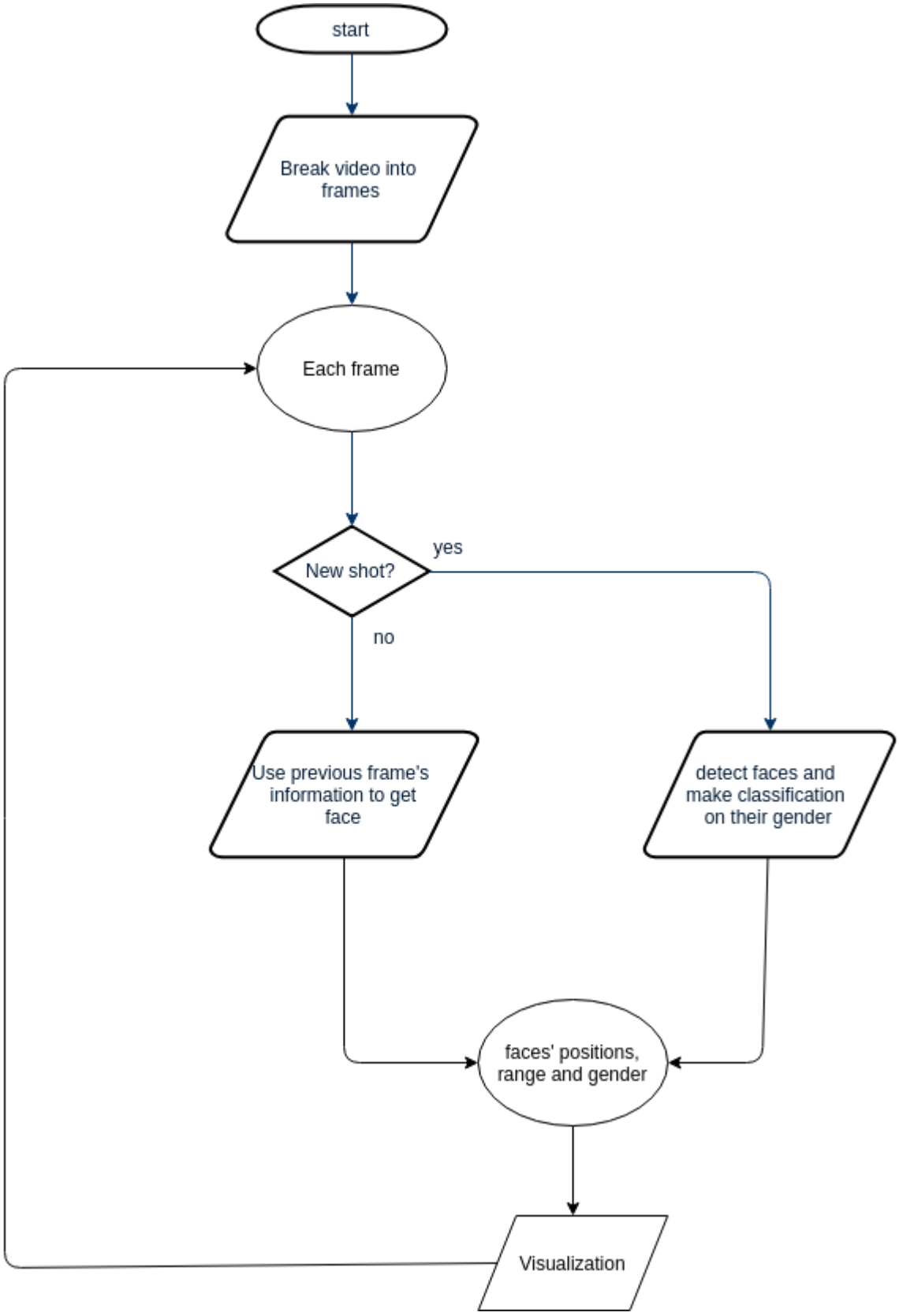
Member: Yige Tao (1000741094), Zhibo Zhou (1001608757)

Overview:

Our group chooses to work on project 1, analysis of news broadcast video clips. In order to approach our target, we defined the following problems:

1. Detect shots in the video. In general, for every frame in the video, we need to compare it with its next frame to see if they belong to the same group.
2. Detect logo.
    - If the logo is known, we need to find where it is in the video (template matching problem)
    - If we are not given how the logo looks like, we need to build a model to determine if a pixel belongs to logo class or not (object detection or semantic segmentation)
3. Detect face.
    - need a sliding window to detect the valid position where a face appears.
    - since the face in an image can be in a different size, need to use multiple sizes of windows to scan on a frame
    - for every pixel in a window, we put all of them as inputs into a model to give us a result of how possible the window contains a human face.
4. Face tracing.
    - In a scenario that two frames are in a same shot, we need to track the faces from the previous frame in the following frame.
5. Gender detection.
    - Need a classifier to predict if an image is a male face or female face. Training set and face picture cropped from video should have the same size.
6. Visualization

Flow chart


start
Break video into frames
Each frame
New shot?
yes
no
Use previous frame's information to get face
detect faces and make classification on their gender
faces' positions, range and gender
Visualization

Plan:

1. Detect shots
    - We might start with reading TREC Video Retrieval Evaluation (TRECVID) 2005's report first. One of TRECVID 2005's task is about shot boundary determination for news videos. (Over, P., Ianeva, T., Kraaij, W., & Smeaton, A. F.)
    - will use python OpenCV to solve this problem
2. Detect logo
    - If we are given the template of the video's logo, do simple template matching.
    - Otherwise, we need to generate our own template. With some analysis (Zhang, L., Xia, T., Zhang, Y., & Li, J. (2011, September)), what we need is to focus on either top left or top right of a video and do some corner detection to compare feature points. Find the set of feature points that are not changed in most frames, we can reference these points as logo features and crop them out from the video.
    - Using OpenCV should also sufficient here.
3. Detect face.
    - We plan to work on Viola-Jones face detector (Viola, P., & Jones, M. (2001, December)).
    - If possible, we might want to try using convolutional neural network in our sub-classifier. Which is an implementation of the report by Li, H., Lin, Z., Shen, X., Brandt, J., & Hua, G. (2015).
4. Tracking face.
    - No much idea. Seems tracking feature points of a face between different frames is easy to do. Lucas-Kanade Optical Flow in OpenCV looks good to start with.
5. Gender detection.
    - This is a classic classification problem. First, we need to scale our face images into one size. In order to do that, we need seam carving.
    - When we have a valid test case, we need to train a model. The plan is to do a simple linear regression model first using pytorch. If we have time we can improve it.
6. Visualization
    - Will try to follow this tutorial. (https://towardsdatascience.com/object-detection-and-tracking-in-pytorch-b3cf1a696a98)

Reference

Li, H., Lin, Z., Shen, X., Brandt, J., & Hua, G. (2015). A convolutional neural network cascade for face detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5325-5334).

Over, P., Ianeva, T., Kraaij, W., & Smeaton, A. F. (2005). TRECVID 2005-an overview.

Viola, P., & Jones, M. (2001, December). Rapid object detection using a boosted cascade of simple features. In null (p. 511). IEEE.

Zhang, L., Xia, T., Zhang, Y., & Li, J. (2011, September). Hollow TV logo detection. In 2011 18th IEEE International Conference on Image Processing (pp. 3581-3584). IEEE.