REPORT

Group Member: Bin Liu BingZhao Shan

Utorid: binliu shanbin1

Introduction

The problem we choose is video analysis. It can be further divide into four sub-problems:

- • Shot detection: Identity different shot where shot represent a set of consecutive frames with a smooth camera motion. (scene change)
- • Object detection: Detect and mark object in a video frame, object includes faces, company logo and clowns.
- • Object tracking: After detected objects in current frame, correctly predict the location of objects in the next frame without detect twice, and mark the objects.
- • Gender classification: After getting a cropped image of face, correctly predict the gender of it.

We picked it because it has been used in a lot of area we are interested in (video surveillance, Photography etc.).

Previous work we found and reference to Lecture Slides:

- • Shot detection: We have read “SSD: Single Shot MultiBox Detector” and Histogram Difference Approach, we implemented both of these two methods and made comparisons between them.
- • Object detection: We read the procedure of Sliding Window with HoG detector on the lecture slides as well as some paper related to training SVM during image processing. We made our own HoG Face Detector with linear SVM in this assignment.
- • Object tracking: Assignment3 has similar procedure of extracting features and matching similar points using RANSAC. We also read “NumPyramidLevels.” which uses pointTracker for tracking.
- • Gender classification: We read papers that used HoG and SVM with different kernel for gender classification and we tried three kernel function while training the SVM.

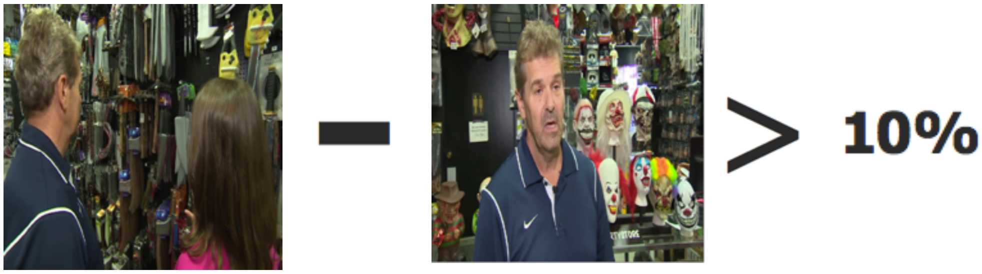
Methods

1. Shot Detection (Two Approaches)

Approach1: Sum of Absolute Difference (SAD) with NMS.

Shot changes are always accompanied by huge change in pixels at all positions of two adjacent frames F1 and F2. We converted the two Frames into gray images and used the method of Normalized Correlation to analyze the similarity between them. Then we minus the similarity from 1 and change it to percent to get the difference (or change) between F1 and F2:

$$Similarity = \frac{\sum_x \sum_y [F_1(x,y) \cdot F_2(x,y)\ ]}{||F_1|| ||F_2||} \qquad changingRate = (1 - Similarity)\times 100$$

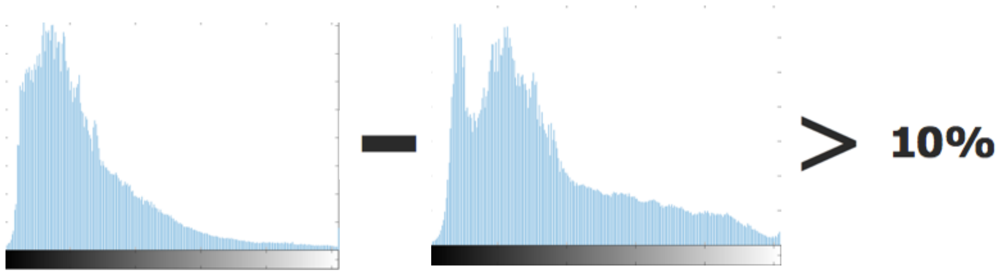


We use threshold to analyze the changingRate between adjacent frames, if the changeRate is greater than the threshold we set, then we collect them as candidate pair of shot Detection. We used non-maximal-suppression to make sure that the shot is not too close to each other.

Approach2: Histogram Difference of adjacent frames (HIST/HD) with NMS.

HD is also a sensitive property for Shot Detection. When the shot changes, There is significant difference between Histograms (h1 and h2) of adjacent frames F1 and F2. We collect the Histogram Data of F1 and F2 using imhist (MATLAB built-in function) and analyze the similarity between them using Normalized Correlation and again change them into changing Rate:

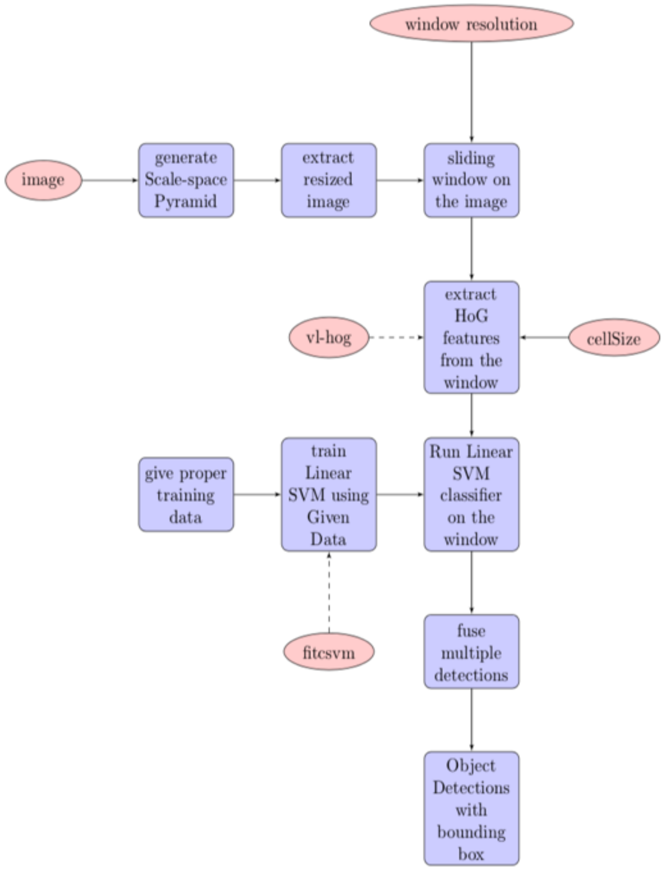
$$Similarity = \frac{\sum_x [h_1(x) \cdot h_2(x)\ ]}{||h_1|| ||h_2||} \qquad changingRate = (1 - Similarity)\times 100$$



Again, we use threshold to analyze the changingRate between histogram of adjacent frames, if the changingRate is greater than the threshold we set, then we collect them as candidate pair of shot Detection. We used non-maximal-suppression to make sure that the shot is not too close to each other.

2. Logo Detection and Face Detection (HOG Detector)

We detect the news company logos and faces in the images using Sliding Windows with HoG Detector. The algorithms are the same for both logo and face Detection, we choose the face Detection Algorithm to Demo how we code and what source we used.


window resolution
image
generate Scale-space Pyramid
extract resized image
sliding window on the image
vl-hog
extract HoG features from the window
cellSize
give proper training data
train Linear SVM using Given Data
Run Linear SVM classifier on the window
fitcsvm
fuse multiple detections
Object Detections with bounding box

Note:

(1) We used vl_hog in the Vl_feat package to extract HoG features of the current window.

(2) We cropped the face images so that the the detector has better performance.

(3) We trained MATLAB function fitcsvm with linear kernel function as classifier.

(4) We trained the SVM with face and 'non-face' images. We cropped the 'non-face' images from Movie **'Teddy Bear'.**

(5) We used another kind of NMS method in the 'Fuse' Step:

Use Greedy Algorithm to find box1.

If $\frac{box1 \ \cap box2}{\min\left(box1, box2\right)} > 0.5$, then remove box2

3. Face Tracking: Affine and Similarity with RANSAC

We tried both Affine transformation and Similarity Transformation using RANSAC to track the faces.

First we get the face bounding box BOX1, then we extract features from BOX1 and match these features with the following frame F2, find all the matched points that passed some threshold and compute the transformation using RANSAC method with 100 trial.

4. Gender Classification Using SVM and HoG

We used the SVM classifier and around 1400 male vs. female Training Data to train our classifier. We tried to train the classifier in using three kernel function: linear, Gaussian (RBF) and polynomial (degree=2). The procedure chart is as follows:

Note: (1) we used MATLAB built-in function to extract HoG features and we used fitcsvm. (2) we tried to classify the clowns as 3rd class using fitccoc, but the result is disappointing because the training data of clowns has very limited size.

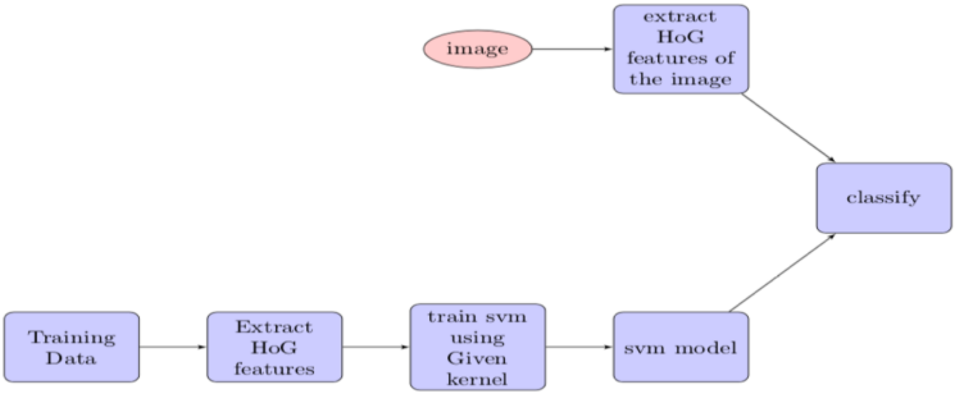


image
extract HoG features of the image
classify
Training Data
Extract HoG features
train svm using Given kernel
svm model

Main Challenges

1. We tried to find proper threshold before we do Normalized Correlation to adjacent frames but the threshold is very large and hard to adjust for better performance. We applied the NC method to for difference comparing from the slide 13. The threshold then become constrained to 0~1. Easier to adjust.

2. For soft Shot Change, we detected several shots during the transformation time between two frames that forms a shot. This result in our bad precision at first, but after we use NMS, the precision improves a lot.

3. We used the NMS in the HoG detector as well, but some outliers are still hard to eliminate. For example, a small outlier totally inside a correctly detected face. Since the area of the outlier is small, when we analyze $\frac{area(box1 \cap box2)}{area(box1 \cup box2)}$, we get area(box2) if the outlier box2 is inside correct detection box1. To fix this, we used $\frac{box1 \cap box2}{\min{(box1, box2)}}$ to decide whether box2 should be removed from the detection.

4. Face Detection and Gender Classification have very low accuracy at the very beginning. We cropped the face Data so that the background is removed from the training images. The performance is much better.

5. Time cost. Our HoG works very well but slow compared to MATLAB built-in face-Detector. We are improving all the time but still need around 40s per image to get fairly well performance.

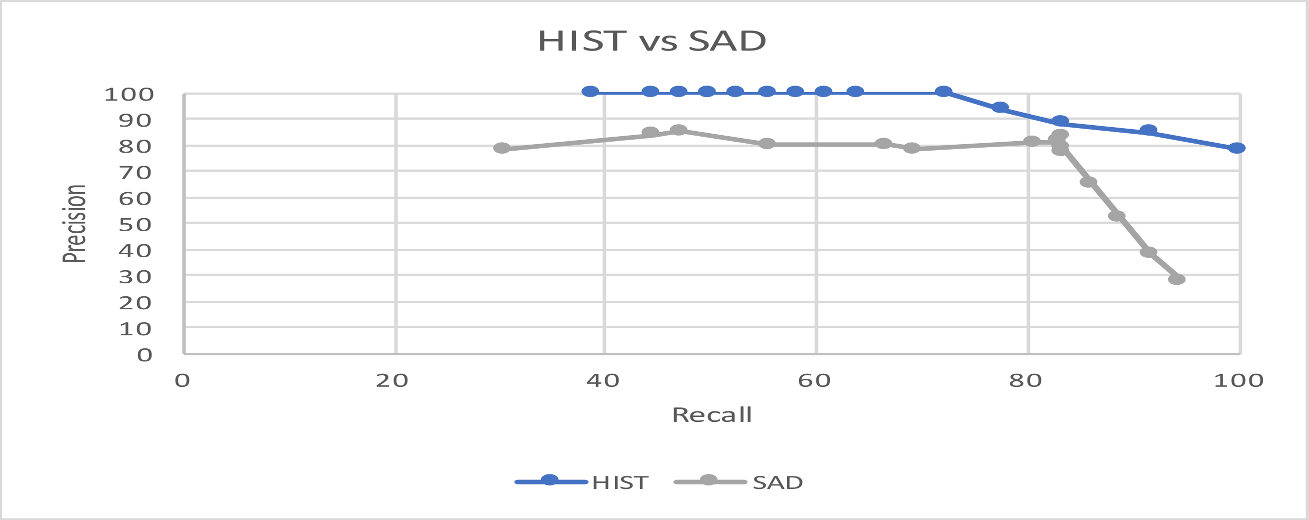
Result

1. Shot Detection:

(1) Precision and Recall:

	Parliament.mp4		Clowns.mp4		Lego.mp4	
	Precision %	Recall %	Precision %	Recall %	Precision %	Recall %
Optimal SAD	100	100	83.3333	83.3333	41.3043	63.3333
Optimal HIST	100	100	84.6154	91.6667	45.0000	90.0000

We collect the Precision and Recall curve of HIST and SAD using Clowns.mp4 :



(2) Comparison and Limitation:

FACEBOOK.COM/PROJECTOTVPHOTOGRAPHY
FACEBOOK.COM/PROJECTOTVPHOTOGRAPHY

Shaking scene:

SAD: This is not shot (right)

HIST: This is shot (wrong)

Dissolve Shot:

SAD: This is not shot (wrong)

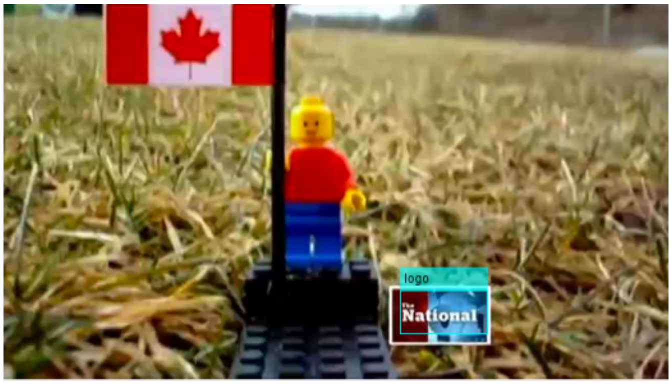
HIST: This is shot (right)

SAD: SAD Approach is good at detecting abrupt Shot and more robust to shaking scene but almost unable to detect soft shot like wipe, dissolve.

HIST: HIST Approach is more likely to lose precision when the scene is shaking heavily but it is sensitive to soft Shot changes.

2. Logo Detection:

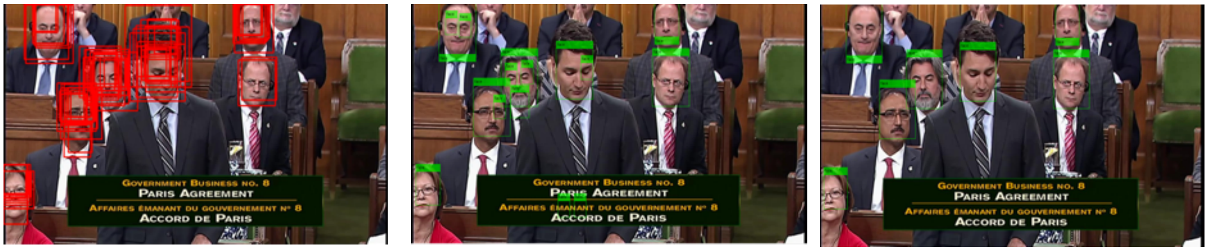
```
Scale-Space Pgramid = [0.0100 0.1100 0.2100 0.4500 0.6750 0.9000 1.1250 1.3500]
stride = 10;
thresh = 0.6;
cellSize = 20;
model_size = [48 95];
```



The performance of logo Detector is quite good. This is because the news company logo is a rectangular and it is easy to be detected by sliding window with similarly window size.

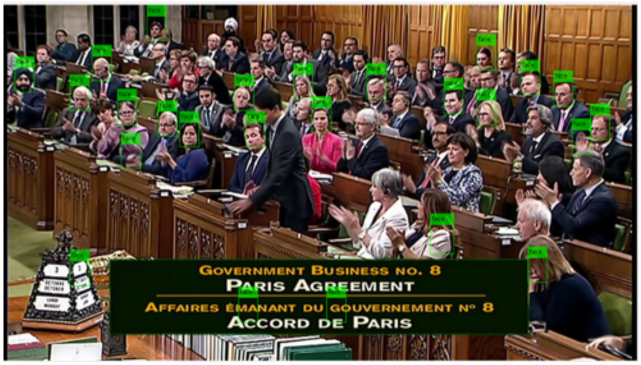
3. Face Detection:

(1) Different NMS approach


Government Business no. 8
Paris Agreement
Affaires émanant du gouvernement nº 8
Accord de Paris

(a) without NMS (b) normal NMS (c) NMS with min box

(2) Multiple Small Faces Performance

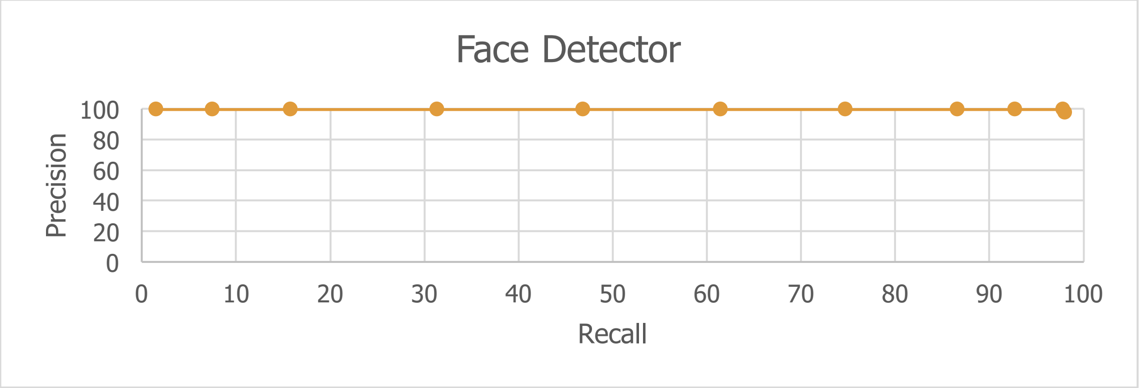

Government Business no. 8
Paris Agreement
Affaires émanant du gouvernement nº 8
Accord de Paris

(3) Performance and Limitation

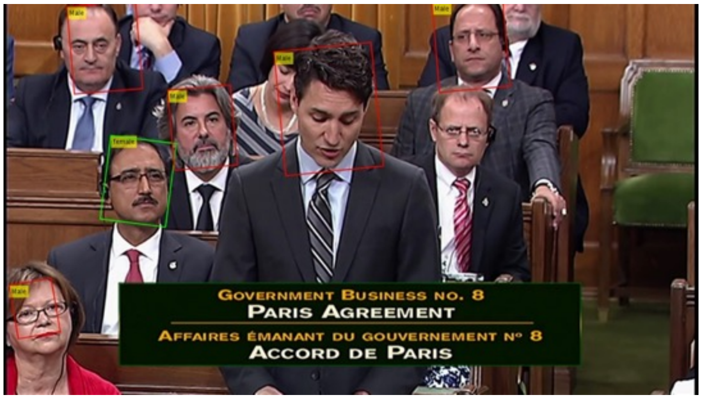
The accuracy is quiet good when detecting Face in normal size (Shown in picture c). But there are two main limitations: a. Not very accurate when analyzing small faces. b. Time complexity is very high. Our algorithm need about 40s per image to detect faces with satisfying detection accuracy.

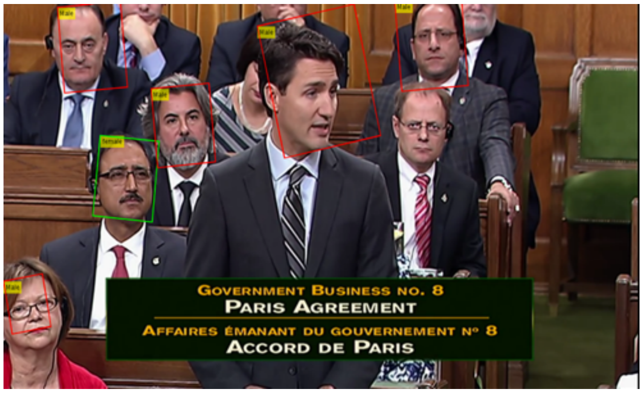
(4)Precision vs recall

Precision and recall on testing folder



4. Face Tracking:

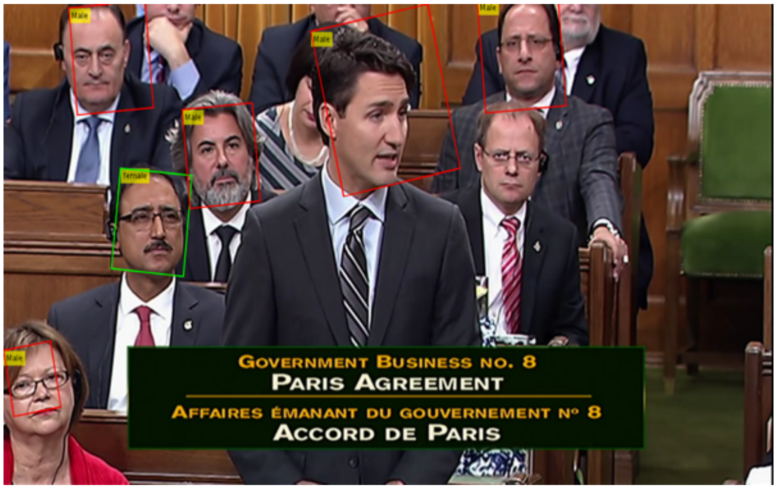

Male
Male
Male
Male
Female
Male
Government Business no. 8
Paris Agreement
Affaires émanant du gouvernement nº 8
Accord de Paris


Male
Male
Male
Male
Female
Male
Government Business no. 8
Paris Agreement
Affaires émanant du gouvernement nº 8
Accord de Paris

The main problem in our approach is that when the number of interest points inside the face box is less than two, we cannot compute the transformation, the face loses tracking then. Once the face is tracked, the performance is good.

5. Gender Classification:

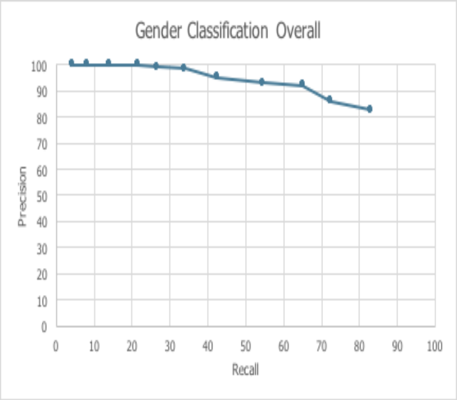
Kernel	Female accuracy %	Male accuracy %	Average accuracy %
linear	93.5443%	95.1220%	94.3478%
RBF	97.8481%	98.9024%	98.3851%
Polynomial (deg=2)	99.4937%	99.8780%	99.6894%


Male
Male
Male
Male
Female
Male
Government Business no. 8
Paris Agreement
Affaires émanant du gouvernement nº 8
Accord de Paris

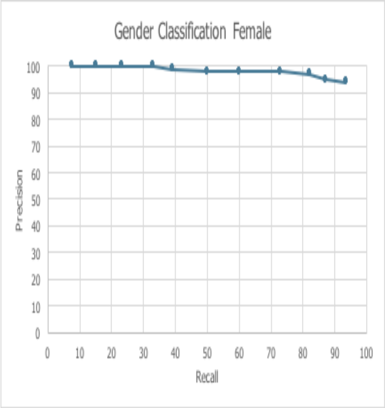
Performance and limitation:

We used different kernel to train the SVM and get very high accuracy while testing. But when it is applied into real use, the accuracy is much lower. This is result from two main reasons: a. The training data has low quality and resolution. b. The face detector uses low resolution window (70x55), which is not large enough for gender classification.

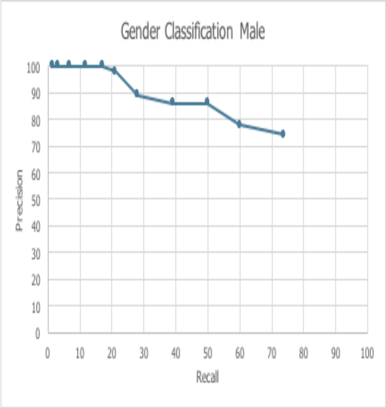
Precision vs recall:



(Overall)



(Female)



(Male)

Conclusion and Future Work

1. SAD and HIST have very limited performance on shot detection. Machine learning can be applied to improve the performance.
2. Sliding Windows with HoG detector is good when detect objects but the speed is slow and there are too many parameters to adjust. Better feature Detector can be applied when detecting objects.
3. Training data is very important for classifier. Our gender classifier has very high accuracy on the testing data but lower performance when applied into use. Training data with higher quality is needed.

Reference

Data used:

“IMDB-WIKI – 500k face images with age and gender labels.” Deep expectation of real and apparent age from a single image without facial landmark, Rasmus Rothe and Radu Timofte and Luc Van Gool, June 2106, data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/.

“Collection of Facial ImagesSpacek, Libor. Face Recognition Data, cswww.essex.ac.uk/mv/allfaces/index.html.

Function used:

“"extractHOGFeatures".” Extract histogram of oriented gradients (HOG) features - MATLAB extractHOGFeatures, www.mathworks.com/help/vision/ref/extracthogfeatures.html.

“Fitcsvm.” Train binary support vector machine classifier - MATLAB fitcsvm, www.mathworks.com/help/stats/fitcsvm.html?searchHighlight=fitcsvm&s_tid=doc_srchtitle.

“NumPyramidLevels.” Track points in video using Kanade-Lucas-Tomasi (KLT) algorithm - MATLAB, www.mathworks.com/help/vision/ref/vision.pointtracker-system-object.html?searchHighlight=pointtracker&s_tid=doc_srchtitle.

Paper read:

ES. Azzakhnini, L. Ballihi and D. Aboutajdine, "A learned feature descriptor for efficient gender recognition using an RGB-D sensor," 2016 International Symposium on Signal, Image, Video and Communications (ISIVC), Tunis, 2016, pp. 29-34.

Rui, Yong, et al. Exploring video structure beyond the shots - IEEE Conference Publication, 2 Aug. 2002, ieeexplore.ieee.org/document/693648/#full-text-section.

Liu, Wei, et al. “SSD: Single Shot MultiBox Detector.” SpringerLink, Springer, Cham, 8 Oct. 2016, link.springer.com/chapter/10.1007%2F978-3-319-46448-0_2.

“Digit Classification .” Digit Classification Using HOG Features - MATLAB & Simulink, www.mathworks.com/help/vision/examples/digit-classification-using-hog-features.html.

“Gender Classification with support vector machines” http://ieeexplore.ieee.org/document/840651/