MMAI - shot detection & video summarization
姓名：江東峻
學號：B01902032 資工四

a. shot detection

#	Genre	Transition Type	Frame Counts	Average Shot Length (#Frames \ (#Boundaries+1))
1	News	Cut	829	829/(8+1) = 92.11
2	Trailer	Fade in + Fade out + Dissolve	751	751/(12+1) = 57.77
3	Ad	Cut + Fade in + Dissolve	1480	1480/(38+1) = 37.95
4	Anime	Cut + Fade in + Fade out	776	776/(16+1) = 45.65
5	Opening Credits	Cut + Dissolve	1230	1230/(38+1) = 31.54

We implement two methods of shot detection: color-histogram and region-histogram. Color-histogram is mentioned in class and region-histogram is mentioned in *Comparison of video shot boundary detection techniques*[1]. Since region-histogram is not useful to deal with fade in and fade out problems and its large computation, we finally choose color-histogram to be our shot detection method.

(1) We choose RGB color space rather than HSV color space because the cosine similarity treats every dimension in the same weight. So, if the values of some dimensions are apparently larger than others, the contribution from other dimensions will be ignored. In RGB color space, when fading out, the R, G and B values decrease uniformly, so we think RGB color space is better for cosine similarity.

(2) We quantize each frame into color histograms (64 bins for each R, G, B value). so each frame is quantized into 64 * 3 = 192-dim vector. (In region-histogram method, before quantization, each frame will be split into 4 * 4 squares, and do quantization for each square. So the computation is much larger than color-histogram method.)

(3) Then, we detect a shot boundary if two vectors are very different. We use cosine distance and a threshold to tell a vector pair is ‘very different’.

[1] J. S. Boreczky and L. A. Rowe. Comparison of video shot boundary detection techniques. pages 170–179, 1996.

After above steps, we get a sequence of shot boundary from the video frames. This method can deal with ‘cut’ very well, but suffers from ‘fade in’ and ‘fade out’. A ‘fade in’ or ‘fade out’ results to a continuous shot boundary. This problem also appears in ‘dissolve’ case. To solve this problem, we can use hierarchical clustering method to group those continuous shots by a threshold and choose a representation from them. And the clustering result can be used in video summarization.

b. video summarization

In this part, we use the shots detected in above part and the clustering result to do video summarization.

(1) First, we describe the hierarchical clustering more clearly. Given shots and a threshold, we can build a tree structure by cosine similarities between shots (we merge two close key frames), and the threshold decides the ending point of merge.

(2) We compute importance mentioned in *SUMMARIZING VIDEO USING A SHOT IMPORTANCEMEASURE AND A FRAME-PACKING ALGORITHM* [2] and select the shot with the highest importance in each cluster to be the representation of the cluster and output the mean frame in this shot as key frame.

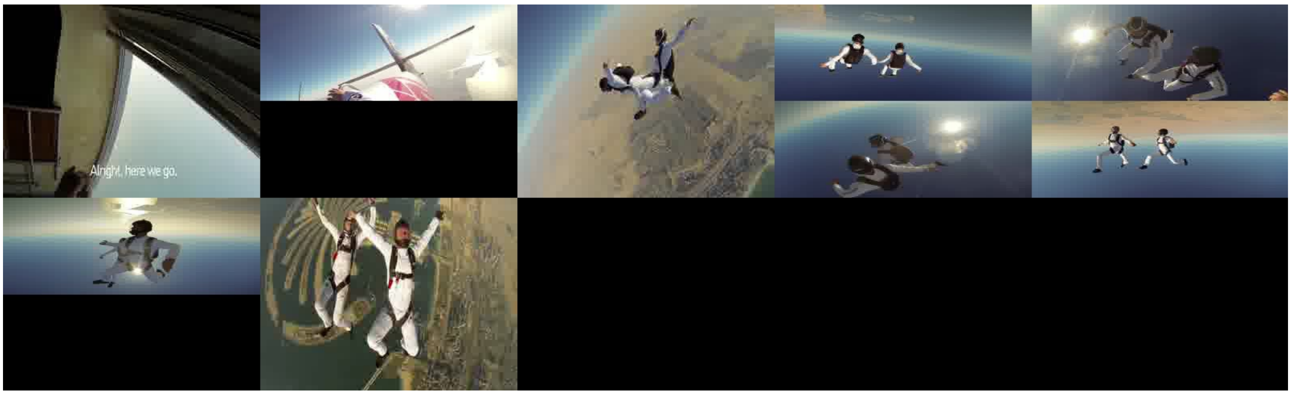
(3) We reshape the key frames by their importance. We choose two sizes: 2 * 2 and 1 * 2 for reshaping. If the importance of a frame is less than 0.5 * max importance of key frames, the frame will be reshaped into 1 * 2 size. And we collage the reshaped key frames into our summary image. The highest priority of order is time order, and if we get two 1 * 2 frames, they may be placed in the same column. Finally, we output the summary of the video.

The summaries are shown below. The black parts do not present black frame. They are just empty because we do not use layout optimal. And if the frames width exceeds the width summary image, the next frame will begin in the next row below.

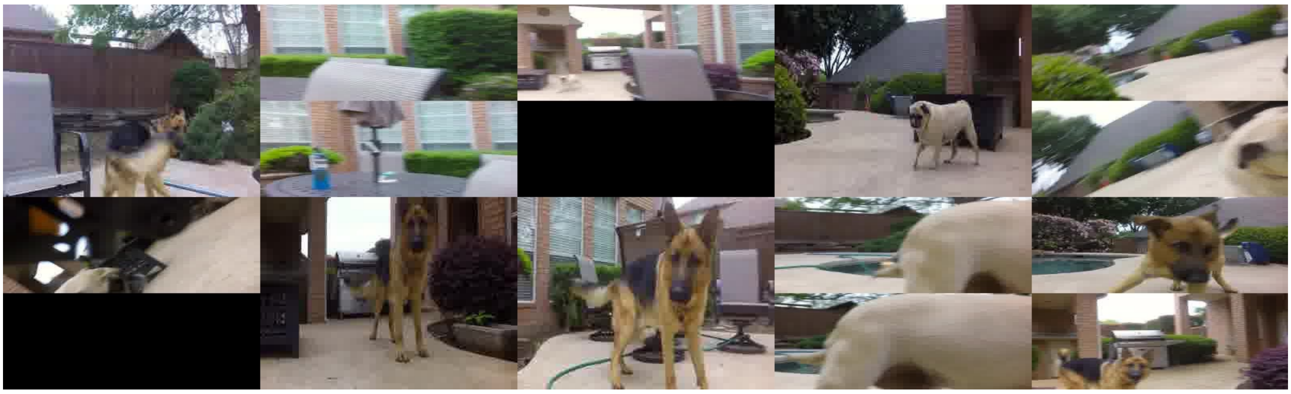
Since we do not use HOG or other edge features, we may group two very different frames only because their cosine similarity of the whole frame(color-histogram) is close. So we may lose some key frames or can not tell the two similar frames just because of their difference in lightness.

[2] Uchihashi, S.; Foote, J., "Summarizing video using a shot importance measure and a frame-packing algorithm," in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on* , vol.6, no., pp.3041-3044 vol.6, 15-19 Mar 1999

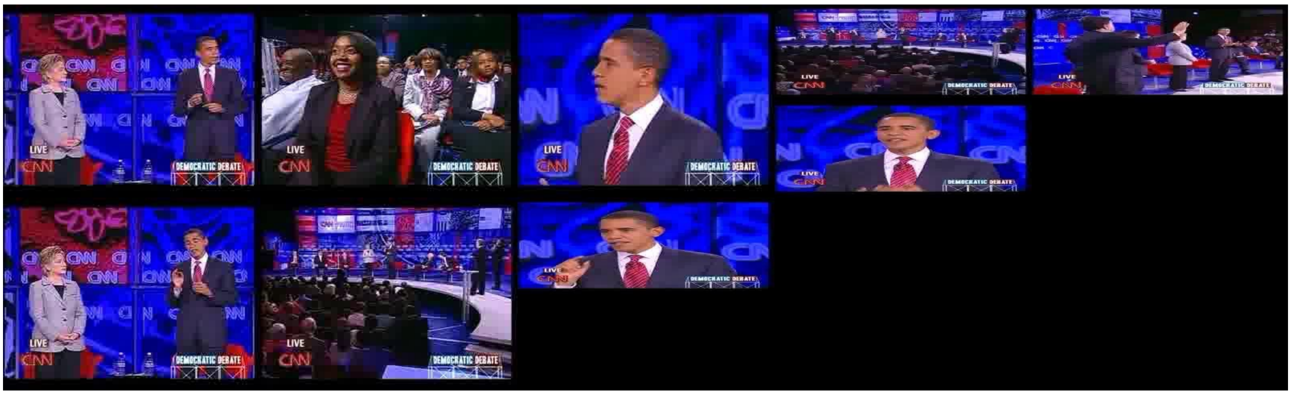
the summary of video #6:



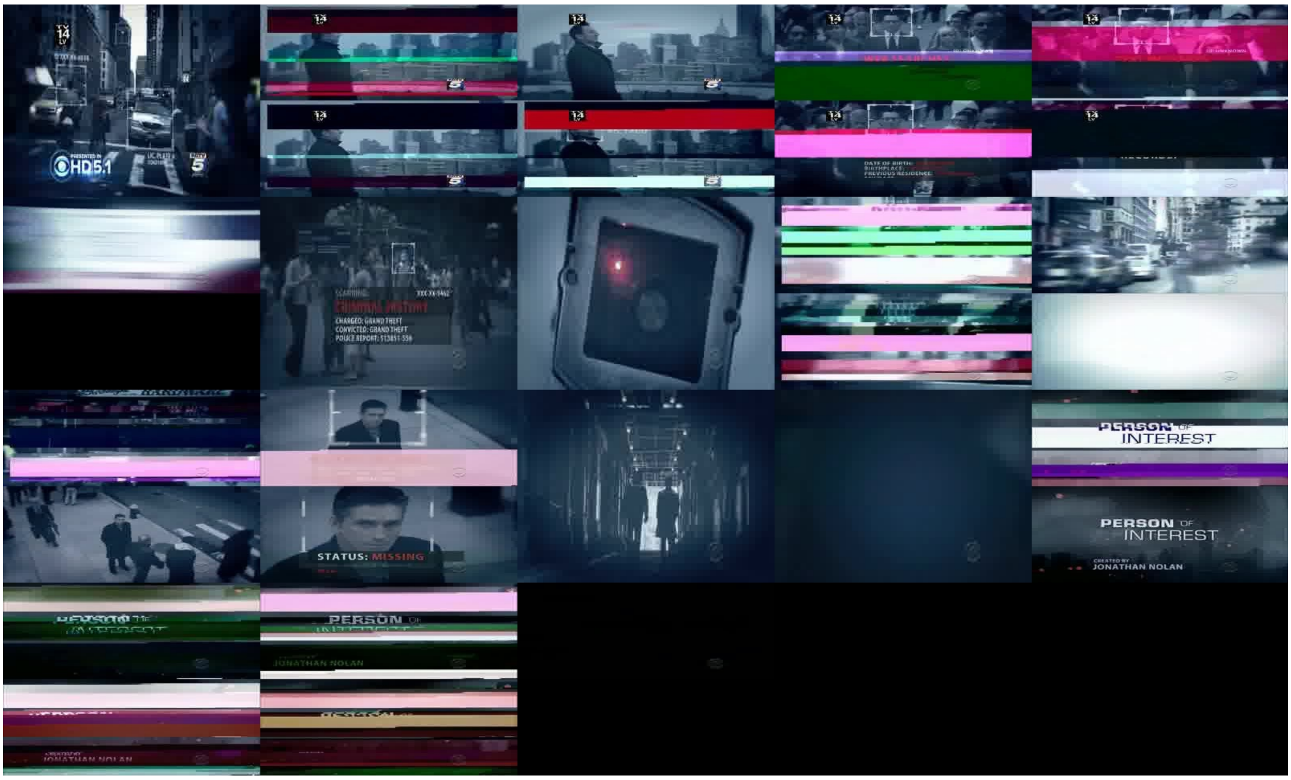
the summary of video #7:



the summaries of video #1 - #5:



Time for Taiwan


HD 5.1
CRIMINAL HISTORY
STATUS: MISSING
PERSON OF INTEREST
CREATED BY JONATHAN NOLAN