

多媒體系統與分析 hw2

F74082141 資訊 112 王浩

1、介紹環境

使用 python 3.11.0, opencv 4.7.0

2、使用到的 feature

2.1. HSV color histogram

對輸入的圖片 RGB 形式先轉換成 HSV 形式，為了比免過大的顆粒度，我們使用 hue、Saturation、Lightness 的 Bin 比例為 8:4:4 共 $8*4*4=128$ 個 Bin 的 Histogram，且我們將採計的範圍限縮在 hue 0~180, saturation 0~256, brightness 0~256 之間，如此一來，我們就可以利用顏色統計來得知每張圖片的特徵。

2.2 SIFT (Scale-invariant feature transform)

SIFT 的原理要先從 Gaussian Distribution 談起，因為我們得到一個輸入圖片，可以將其視為一個 signal 輸入，而為了能夠找到特徵點，我們需要針對圖片上的「點」找出能夠代表他的 scale 大小，所以此時我們就可以利用 Second Derivative of Gaussian 與 signal 相乘後，再多乘上待預估的 delta 以求出能夠代表其大小的 scale 數值，預估方式就是利用 local extrema，此時的波形極值最為明顯，這時候就是找到 point of interested scale size。

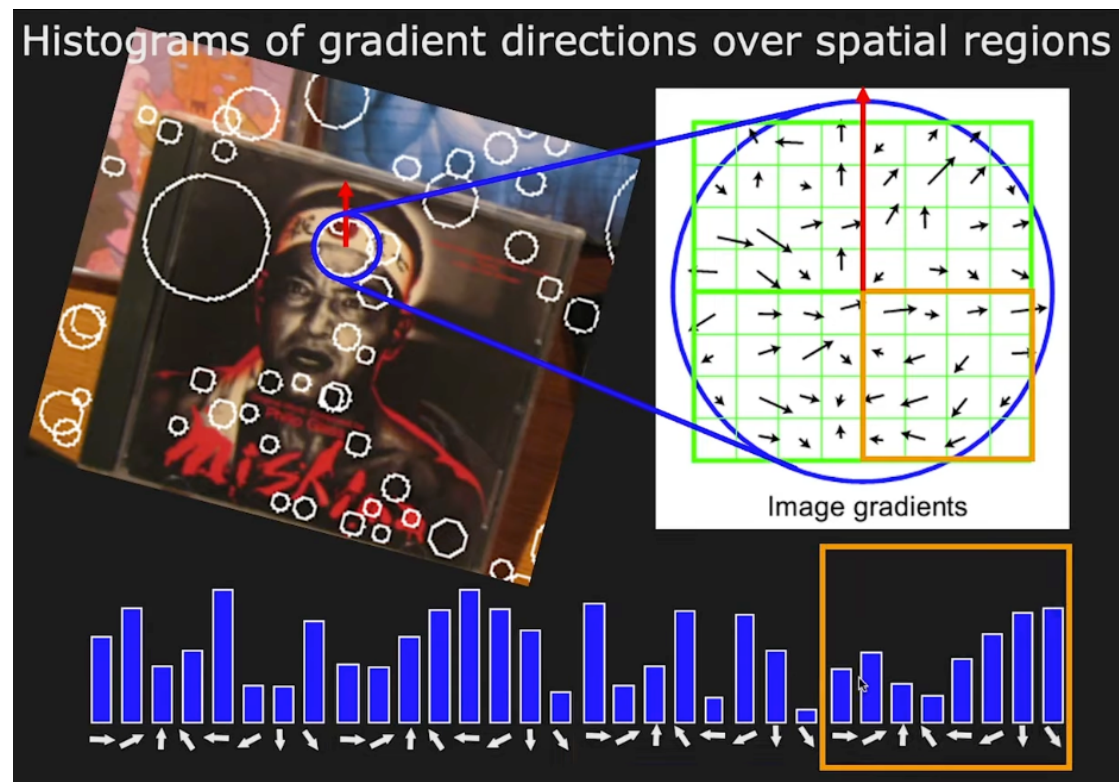
再進入 2 維後，我們就稱為 Normalized Laplacian of Gaussian (NLoG)



而我們也可以利用 Difference of Gaussian (DoG) approximate NLoG，所以在得知每一個特徵點的大小之後，我們就可以利用兩者的大小比例來調整大小。此外我們可以利用圖片本身每一點已經找到的 scale，並切開每一個點、利用圖片算出的

gradient direction θ ，並且統計出最大者，代表這一個 IOR 的方向。

最後 SIFT Descriptor，為了避免光線強弱影響，在考慮 orientation 時，要避免採用 magnitude，而這也是跟第一點 color histogram 最大的差別，實際判斷 POI 的作法為切點為四個方向，並且統計出 direction 的 histogram，所以在配對 match 時，我們則可以挑選 L2 Distance、Normalized Correlation、Intersection 來做兩個 orientation histogram 的比較。



3、shot change detection 演算法

3.1 HSV color histogram

當我們利用剛剛的 feature，可以得到依序每組的 color histogram difference，我們再根據每一組的 difference 以 threshold 調整，找出最佳的演算法。

3.2 SIFT (Scale-invariant feature transform)

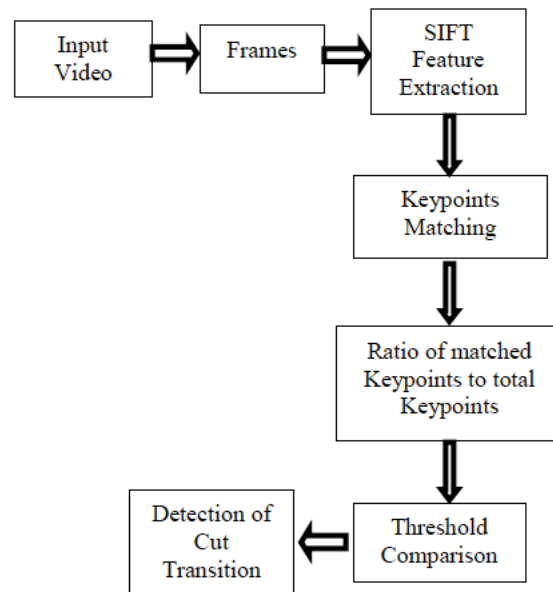


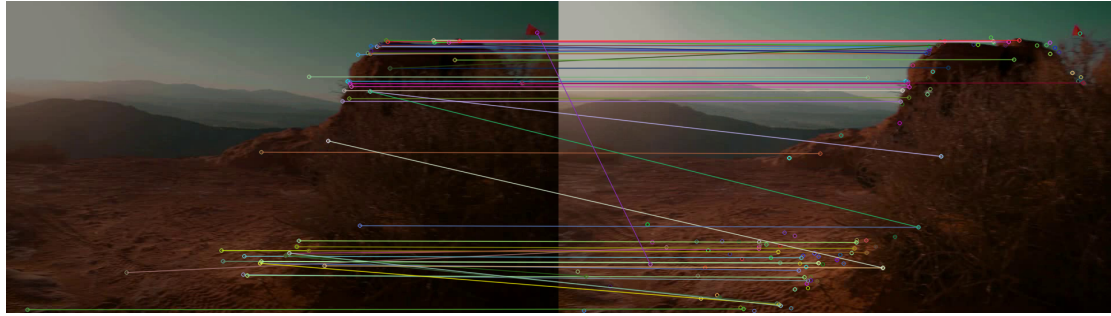
Fig. 2. The proposed system for Video Shot detection

[1]

使用到前述的 SIFT feature，將每一個 frame 都使用 SIFT feature extraction，並且與後方的 feature 互相進行配對，配對方式為 OpenCV 套件中的 knnMatch，與每一個點都做相似度比擬，將最接近的 Match point 作為參選點，再利用絕對距離來判斷出是否為合理的對應點。

以下為我們使用本次 climate_out 的資料集的測試結果，上左為先將影像轉為 gray 之後的原圖，上右為進行單一——張圖片的 SIFT，下方則為利用兩張連續的 SIFT 並且進行 matching 配對，若有符合我們自行手動設定的 threshold，我們就稱其為成功 matching。





接著我們就可以利用 threshold ratio，來判斷出每一次間格的 frame 彼此之間有多少的點是互相對應，由相交點數量除以較大者，來作為 ratio 作為比較，而我們就可以根據不同情況的 threshold 來做出不一樣的情況的圖，進而判斷哪一種的參數會有最好的情況。此外我們也會根據 shot change 的連續情況來刪減，以免有多餘的 shot change 造成誤判。

此外，因為觀察到不少轉場會有較低的 key point 數量。我們也多利用 key point 點的數量來試圖解決轉場的問題，如：

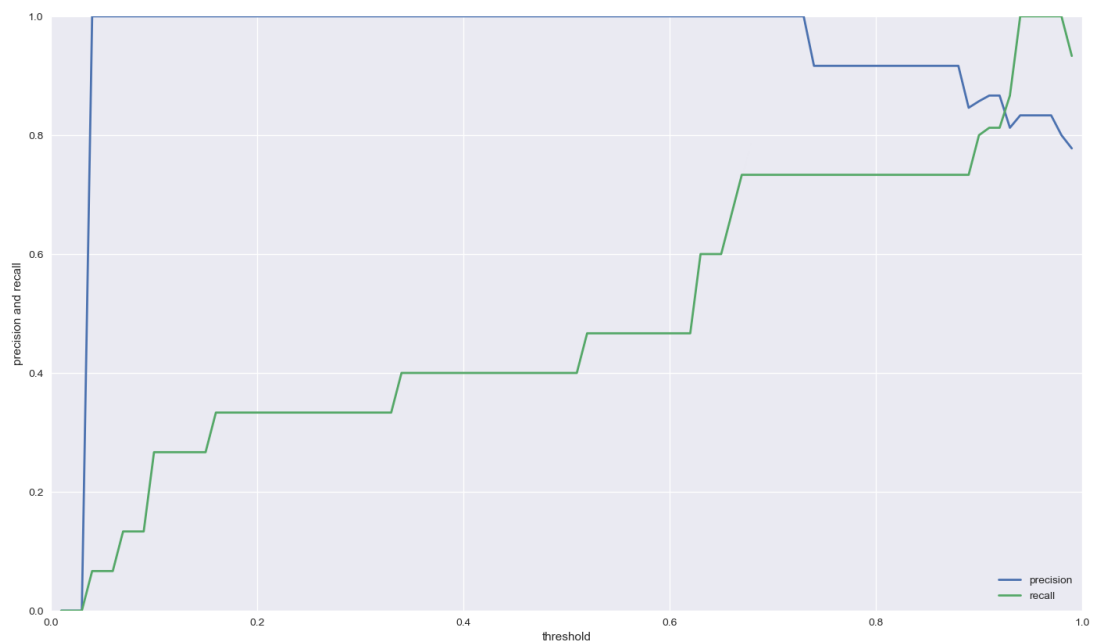


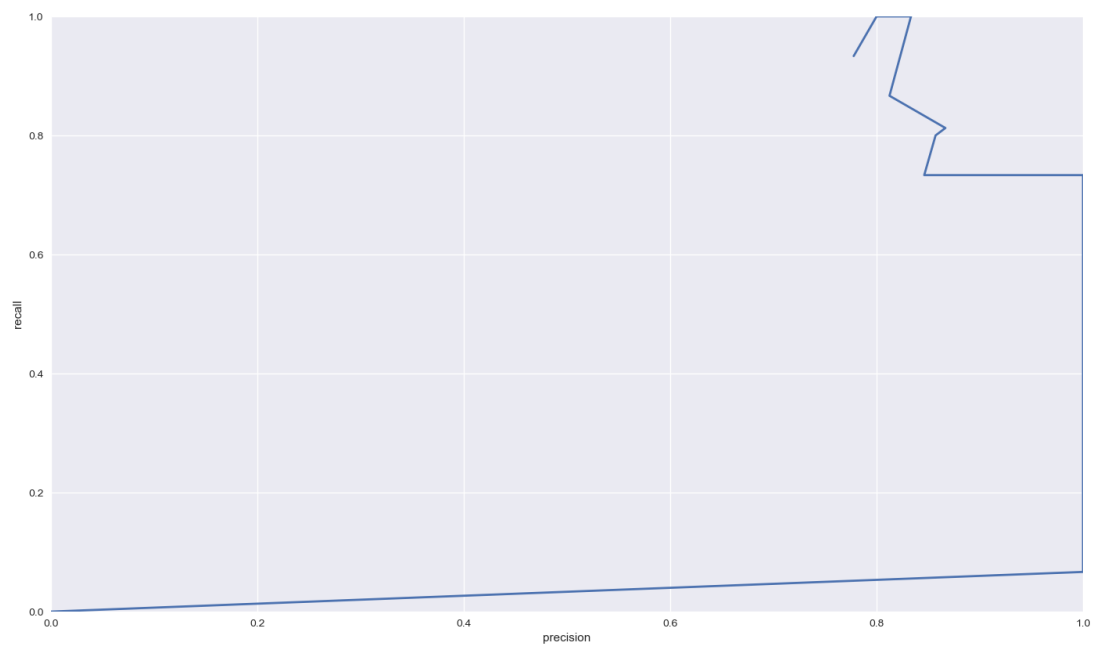
4、偵測效能

4.1 Color

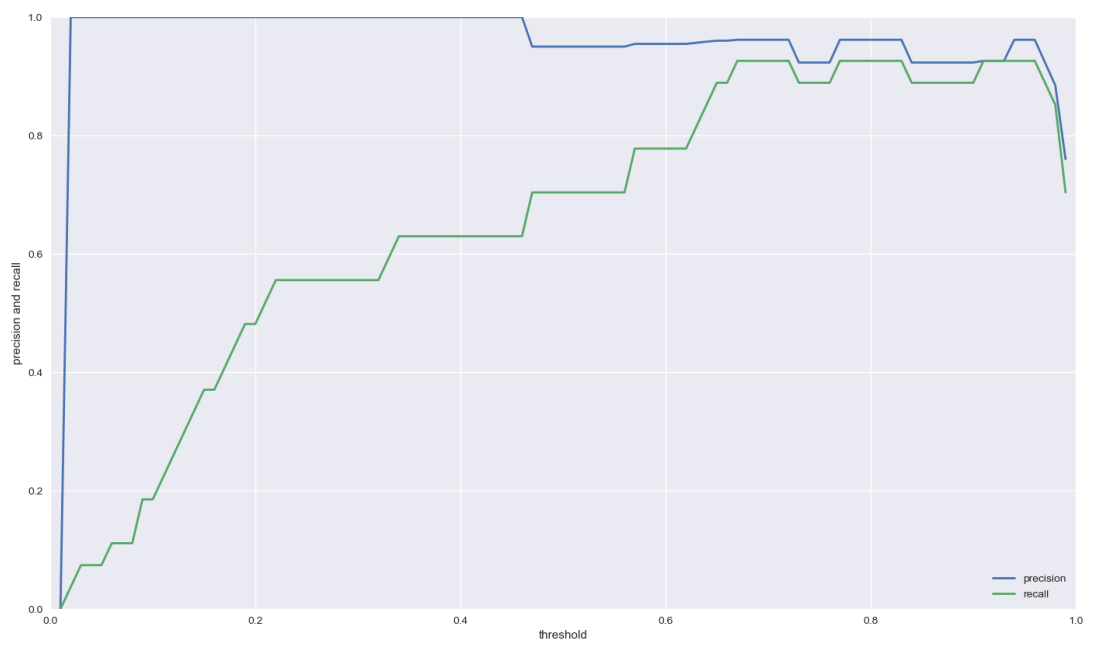
	第一張圖	第二張圖
X 軸	threshold 介於 0~1	precision
Y 軸	precision (blue line) and recall (green line)	recall

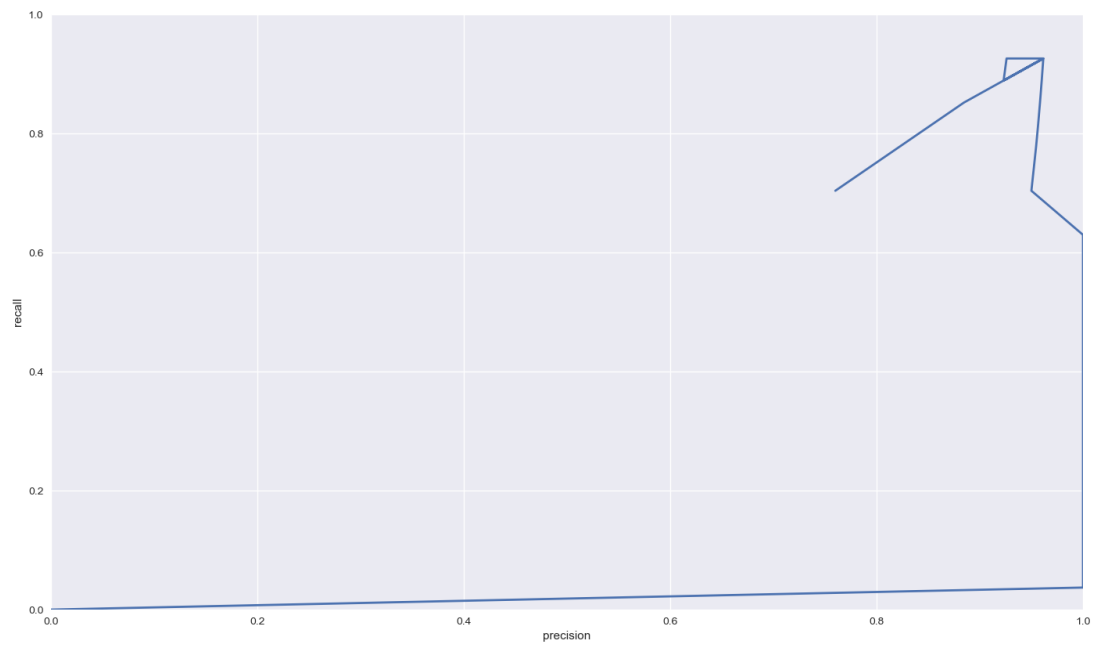
4.1.1 Climate out



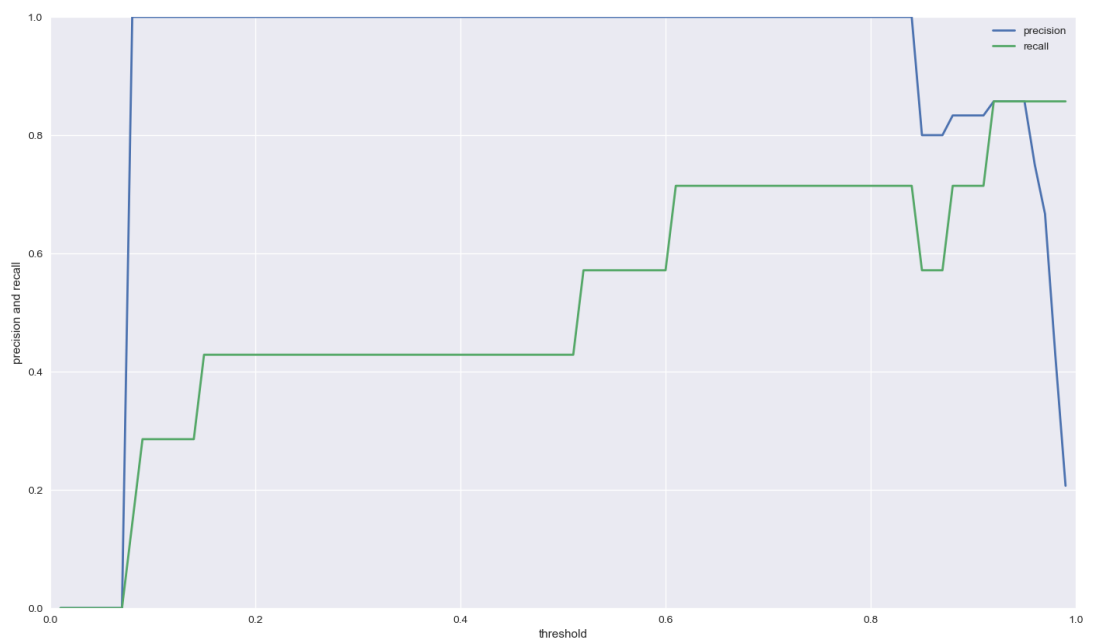


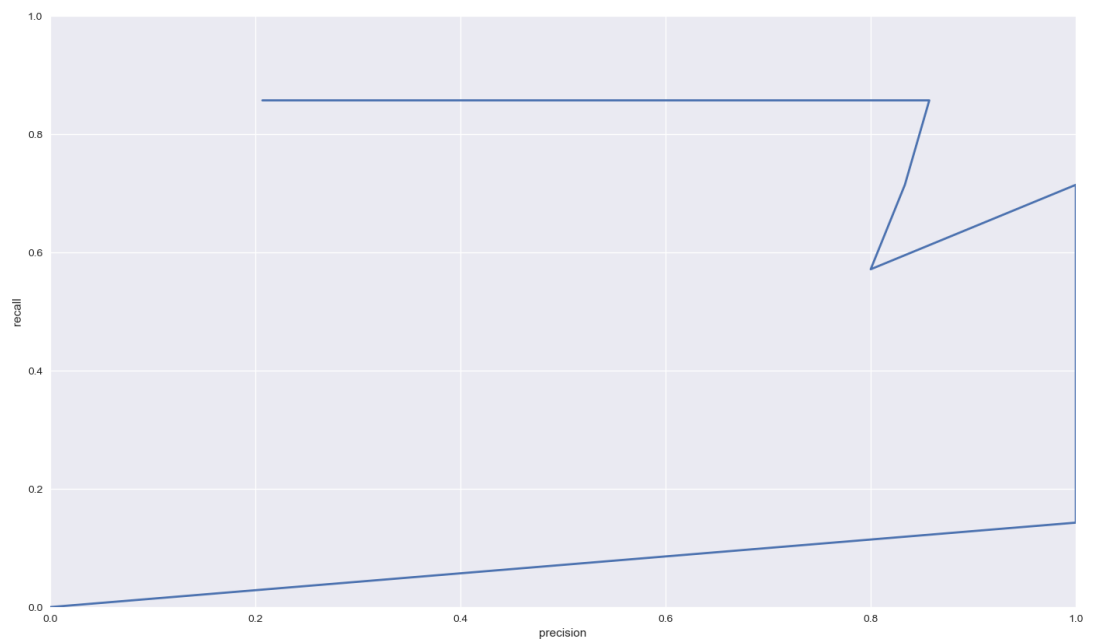
4.1.2 ftfm out





4.1.3 news out

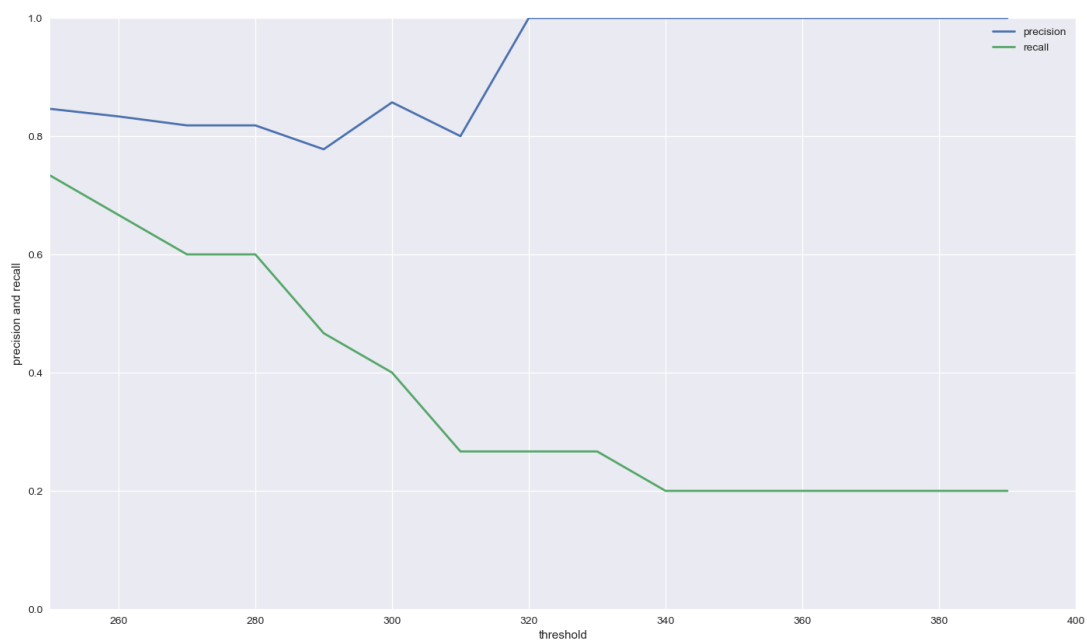




4.2 SIFT (Scale-invariant feature transform)

	第一張圖	第二張圖
X 軸	threshold 介於 250~400	precision
Y 軸	precision (blue line) and recall (green line)	recall

4.2.1 Climate out



最佳狀態

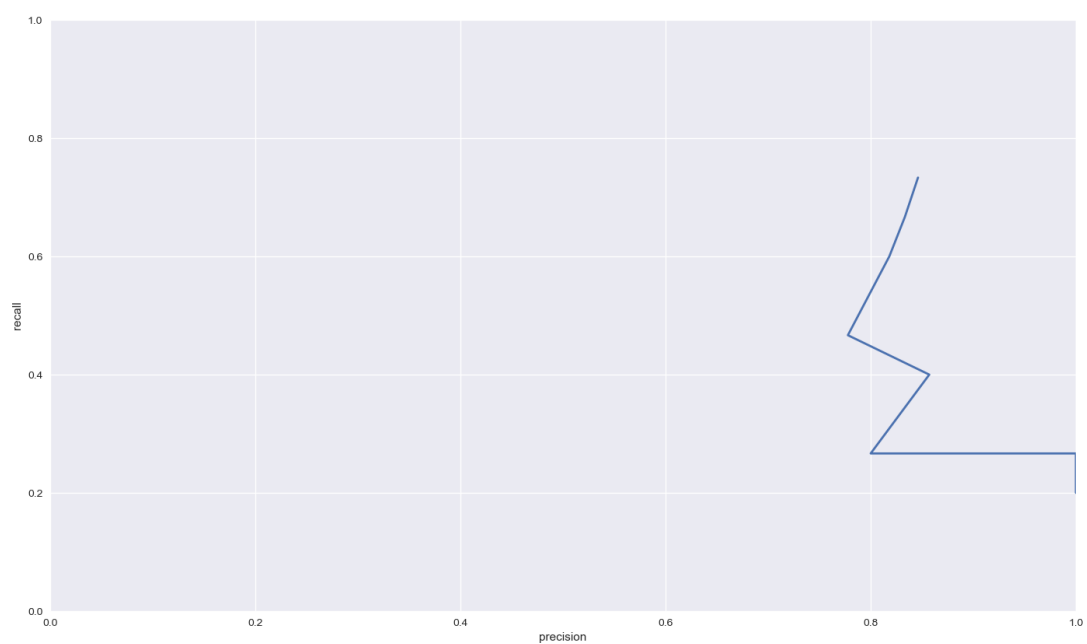
當 Threshold 訂在 250

(表當 match 點距離超過 250pixels 時，就視為斷點)

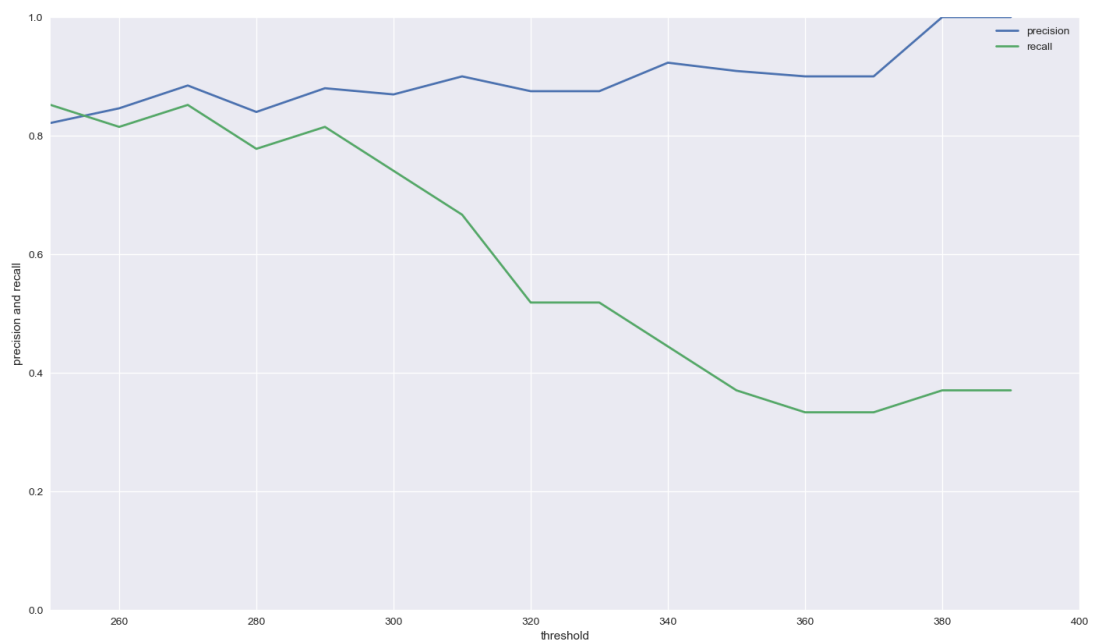
Ratio threshold 定為 0.6

Precision 0.8462

Recall 0.7333



4.2.2 ftfm out



最佳狀態

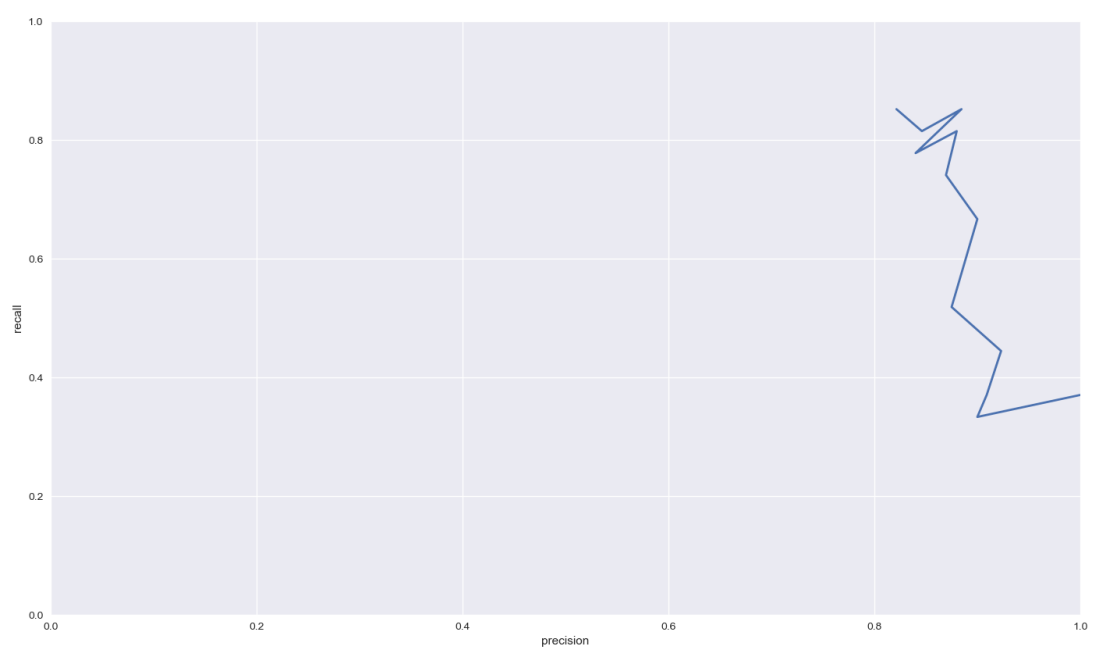
當 Threshold 訂在 270

(表當 match 點距離超過 270 pixels 時，就視為斷點)

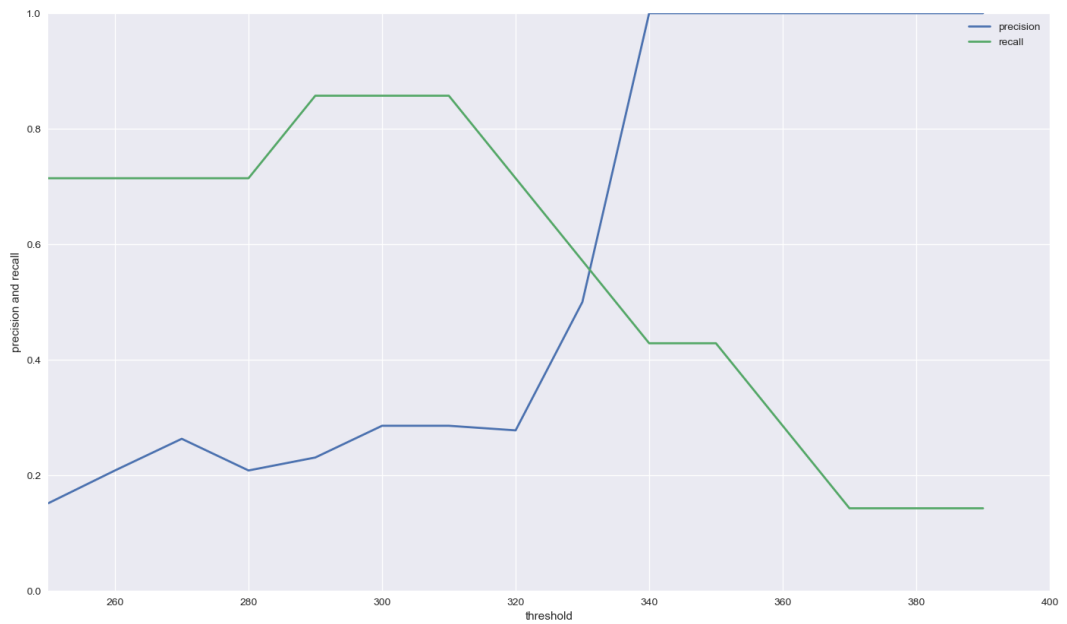
Ratio Threshold 定為 0.4

Precision 0.8846

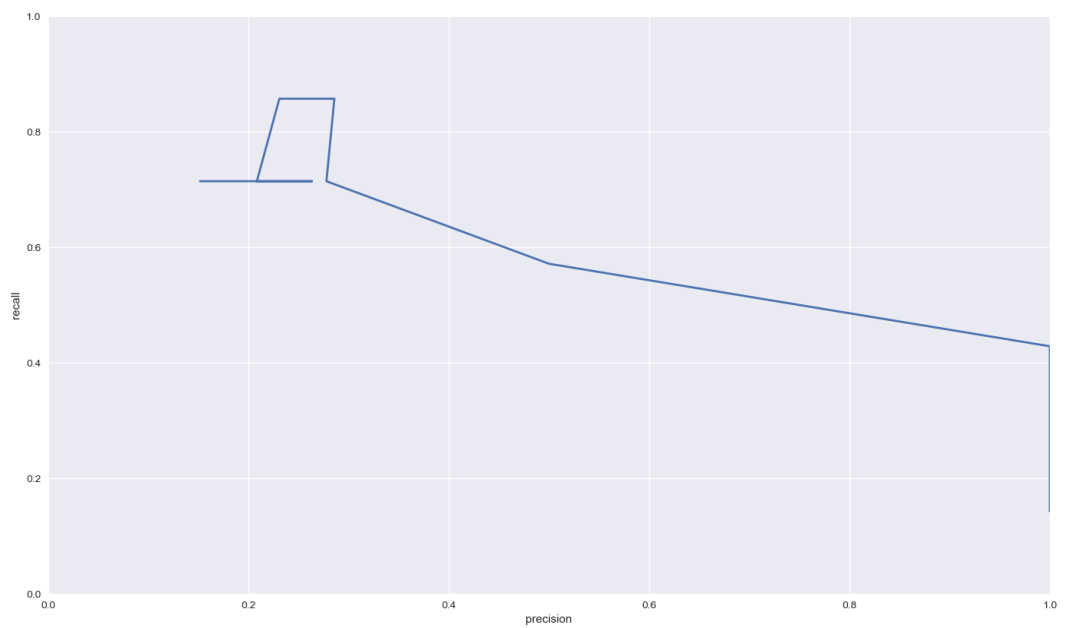
Recall 0.8519



4.2.3 news out



1. 並沒有最佳狀態
當 Threshold 訂在 330
(表當 match 點距離超過 330 pixels 時，就視為斷點)
Ratio Threshold 定為 0.7
Precision 0.5
Recall 0.5714
2. 當 Threshold 訂在 330
Precision 0.8571
3. 當 Threshold 訂在 340
Precision 1.0
4. 此演算法在此資料集 precision recall 的 tradeoff 尤其明顯



References

- [1] J. Majumdar, M. Awale and K. L. K. Santhosh, "Video Shot Detection based on SIFT Features and Video Summarization using Expectation-Maximization," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, 2018, pp. 1033-1037, doi: 10.1109/ICACCI.2018.8554662.