CSC420 Project

Object Detection and Tracking in Video

Huan Wang & Zhengyuan Xie

Problem Definition

- Sometimes, we want to track a particular person or object from a video. Beforehand, we do not know whether the object will appear in the video or not. But we have a set of images about the object we want to track.

- E.g. tracking pedestrains or tracking person's face from a video

Topics from courses

- Object Detection with SVM
- Non-Maxima Suppression
- Finding SIFT keypoints, descriptors and matching
- Computing Homography (Affine Transformation)
- RANSAC

Role of each student

- Object Detection: Zhengyuan Xie
- Video Processing & Object Tracking: Huan Wang

Methods

- **Detection:**
    - Collections with positive and negative datasets
    - With Pedestrain Dataset from:http://pascal.inrialpes.fr/data/human/
    - Histogram of Oriented Gradients Descriptor (HOG)
    - Train Support Vector Machine (SVM) model
    - Hard Negative Mining to learn from false postive patches
    - Using Non-Maxima Suppression to reduce redundant bounding box

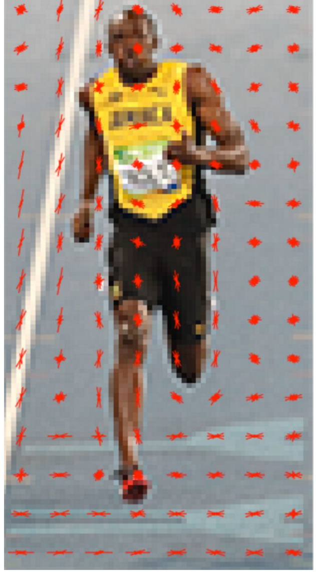
- ## Dataset

Some postive data images: with person (with various posetures) inside (already croppped to 128*64)

Some negative data images: with no person inside

Dataset from:http://pascal.inrialpes.fr/data/human/

- Histogram of Oriented Gradients Descriptor (HOG)



HOG is a histogram storing frequency of each direction

and magnitude among a set of pixels.

image from: https://www.learnopencv.com/histogram-of-oriented-gradients/

How to get HOG

Grayscale and gamma correction

Calculate gradient magnitude and direction for each pixel (Sobel)

Build a histogram based on pixels in each cell

Combine histograms in one block to build a descriptor for one block

Collect vectors from all blocks into a big vector

- Support Vector Machine

SVM can be used for classification problem.

With this model, we aim to find the best hyperplane which separate different classes with the largest distance from the plane to nearest points.

To get better weights, we need to train SVM on examples of different classes.

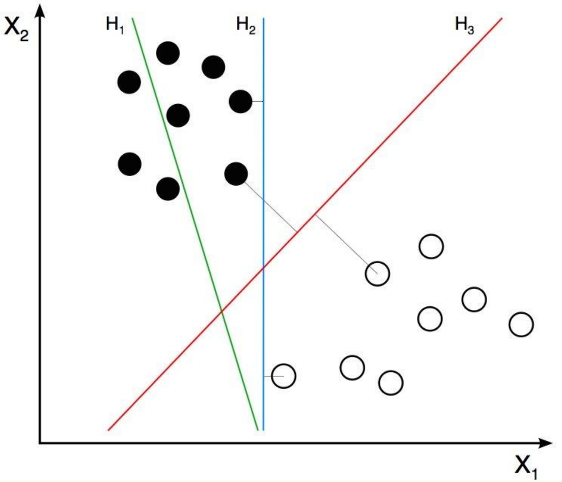
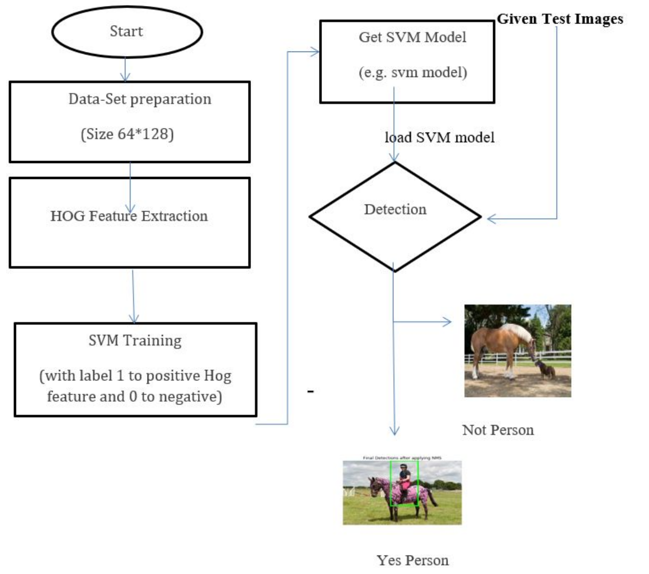


image from: https://www.learnopencv.com/image-recognition-and-object-detection-part1/

Combine SVM and HOG



Get HOG descriptors from positive examples

Get HOG descriptors from negative examples

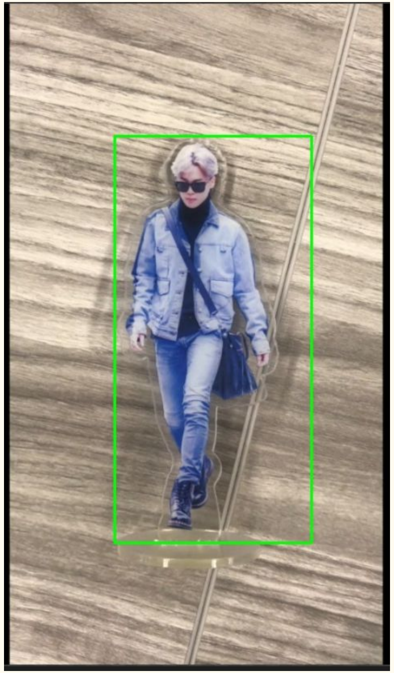
Train a svm detector on these descriptors

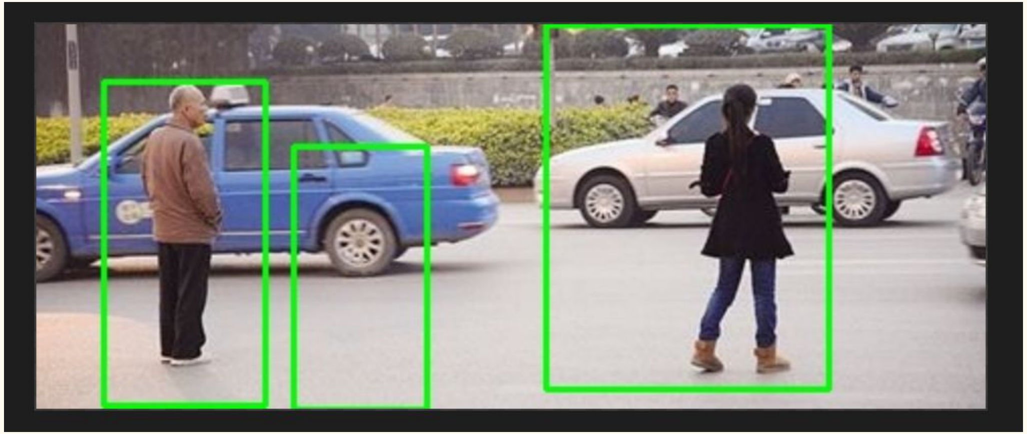
Run the svm detector on some negative examples. Find false rectangles the model detects and resize these rectangles images to a new negative example.

Train svm model again on the extended examples.

Images from: https://medium.com/@richa.agrawal228/person-detection-in-various-posture-using-hog-feature-and-svm-classifier-2c3a3991022c

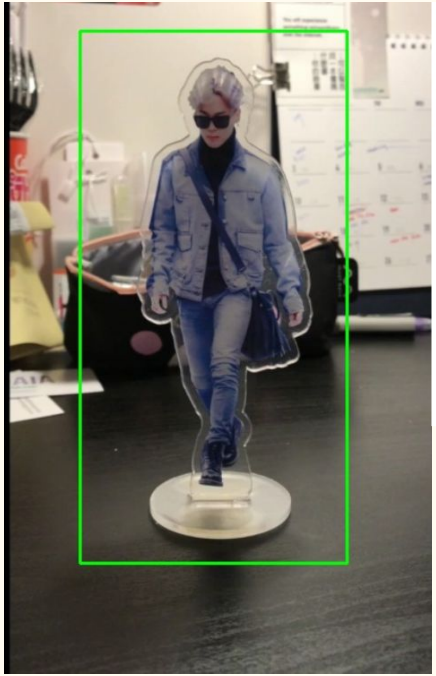
Detection Results

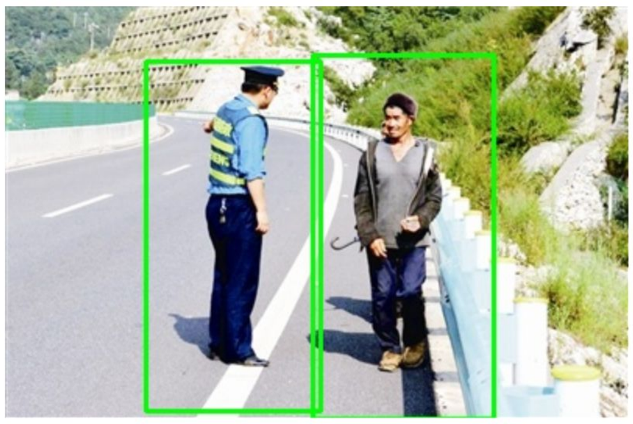


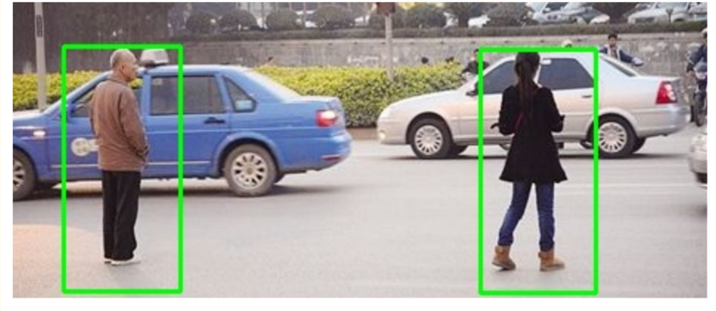


100 positive
200 negative
no gamma correction
(but with false postive detect patches)

Detection results (with more training data)

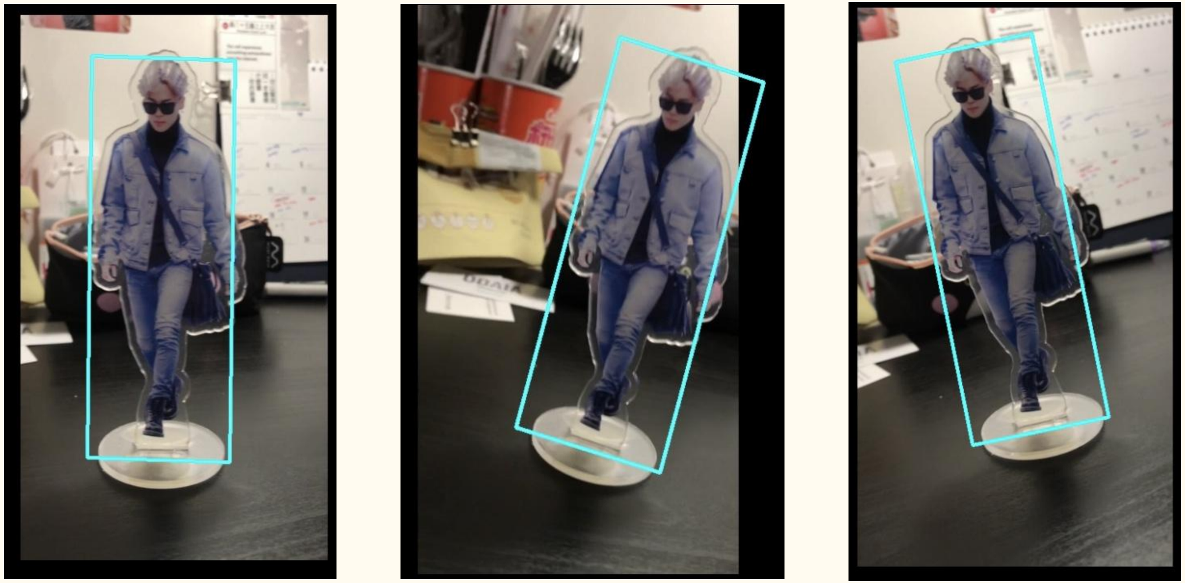




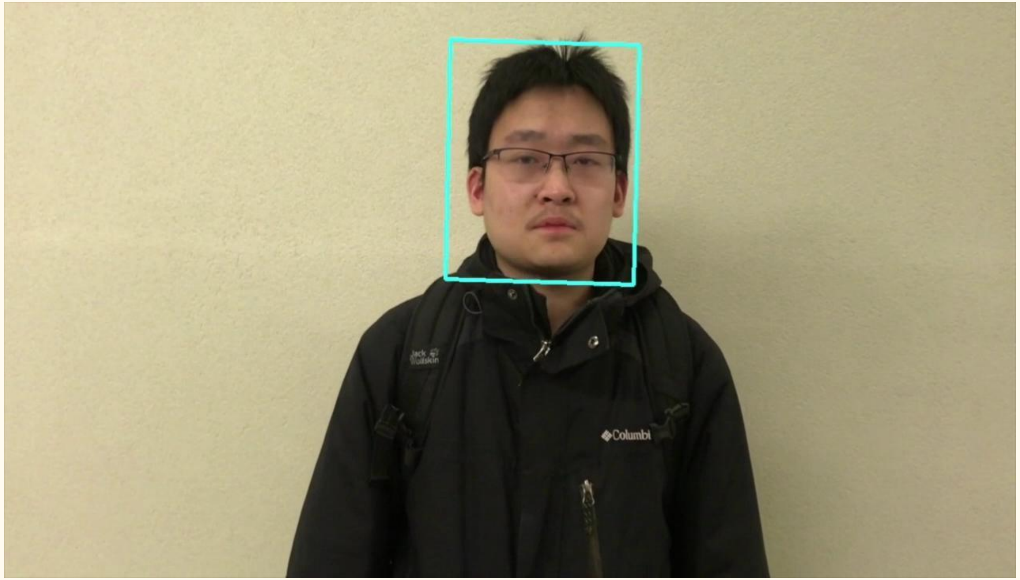


2000 positive examples, 4000 negative examples, with gamma correction but no merging (less false positive detect)

- **Tracking:**
  - SIFT descriptors and Matching
  - Robust method to compute Homography between each frame with RANSAC



- Tracking Result



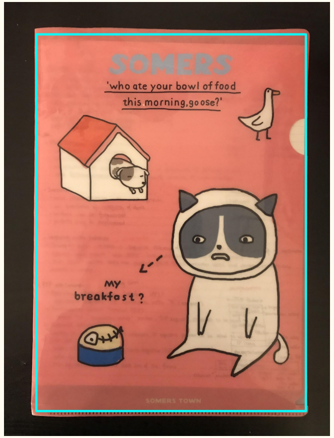
From video, we could see that rectangles could be drawn for horizontal movement successfully, but it will fail when person turns his head or rotates.

Since object is not planar, viewpoint changes will have a huge impact on matching and estimating homography. And it fails since no enough matching and inliers.

Comparison

- Self-implemented RANSAC to find homography and RANSAC within skimage

For this project we use sef-implemented RANSAC method to find homography and have a similar result with RANSAC from other build-in library.



target image

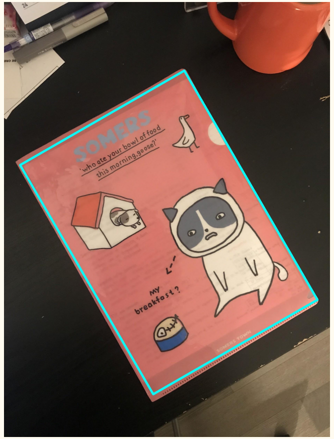


image from skimage's built-in RANSC

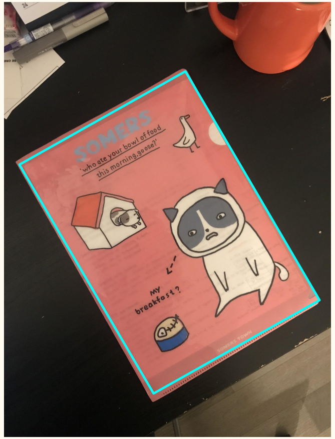
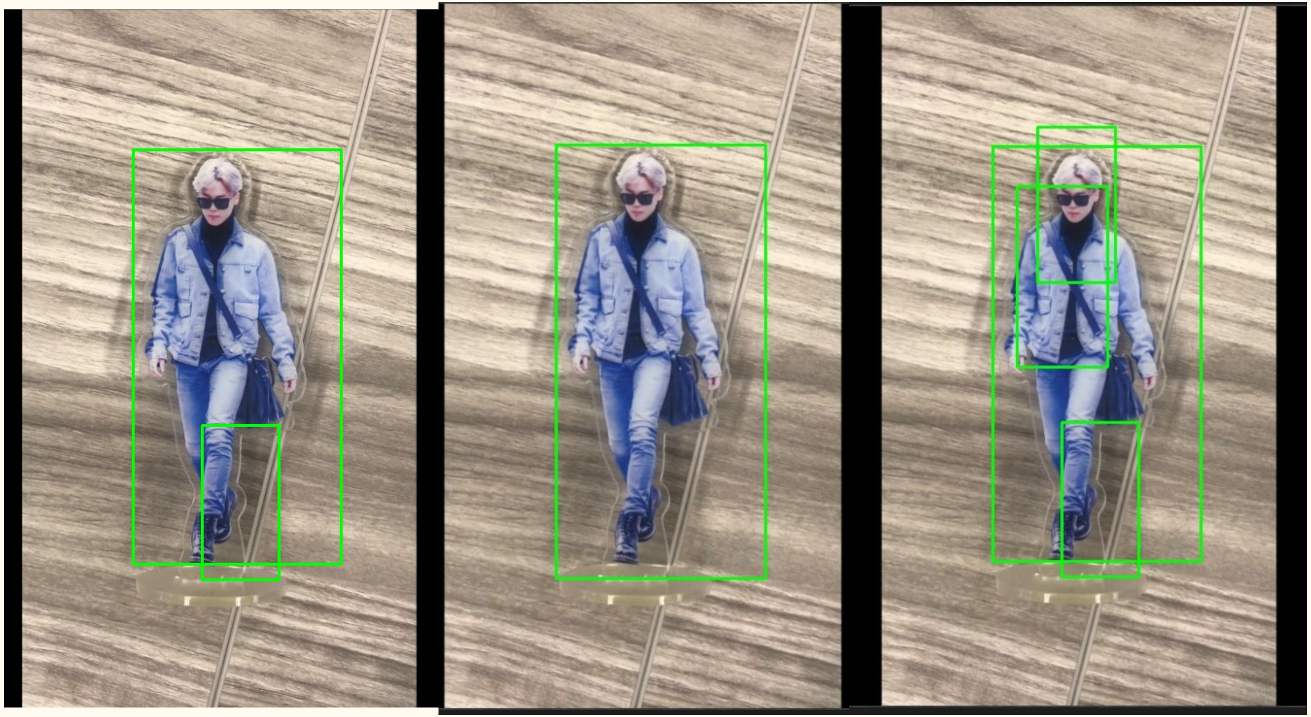


image from self-implemented RANSAC

- Non-Maxima Suppression

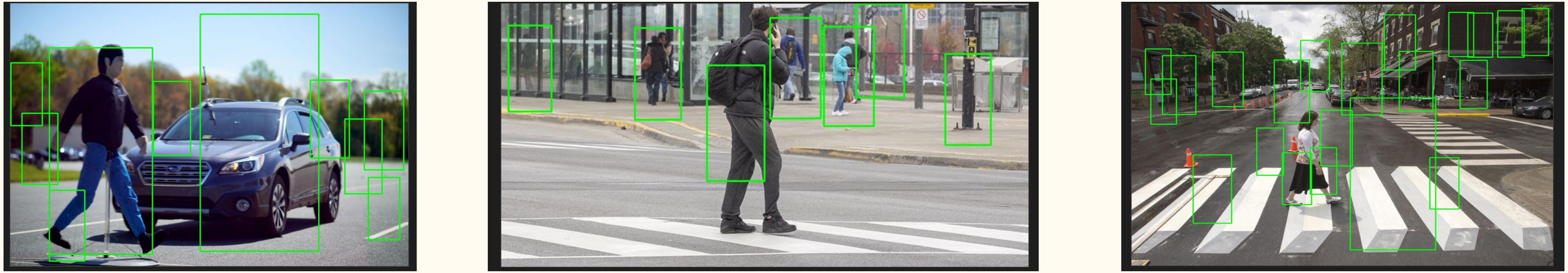


NMS removes small rectangles which share large overlap with rectangles (on right).

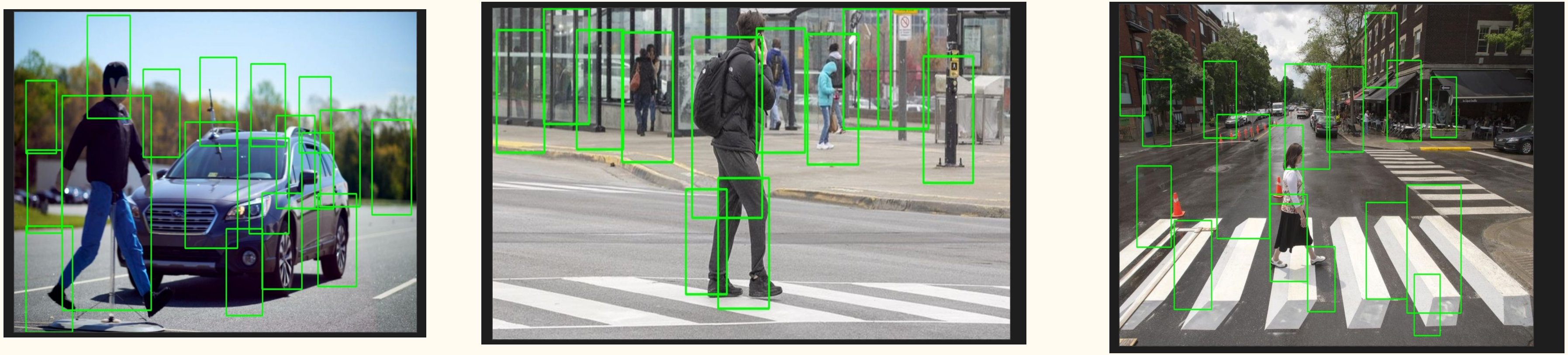
Left image uses NMS. Middle image uses NMS + merge

Right image uses neither NMS nor merge

- With different training data size



100 positive, 200 negative, with NMS, no merging



200 positive, 400 negative, with NMS, no merging

Result Video

Thank You Very Much!