CSE/ECE 344/544: Computer Vision
Assignment-2

Max Marks: 100 Due Date: Mar. 29, 2017, 11:59PM

Instructions

- Please complete all questions.
- Keep collaborations at high level discussions. Copying/Plagiarism will be dealt with strictly.
- Start early, solve the problems yourself. Some of these questions may be asked in Quiz/Exams.
- Submission: Backpack.
- Late submission penalty: 25% credit per day.

---

1. (50 points) Implement the bag of words using HOG/SIFT to classify vehicles and non-vehicles from the dataset available.
   (a) **Split Data**: Randomly split the given data into training (80% of both class) and testing (20% of both class) sets.
   (b) **Feature Extraction**: Extract HOG or SIFT features from the vehicles and non-vehicles training dataset. You can also extract other features which you think will help improving the accuracy. Take N random samples from the image of size $m \times n$, smaller than the image.
   (c) (15 points) **Dictionary Computation**: Run k-means (k will be dictionary size) clustering on a subset of all training features to learn the dictionary centers (i.e. visual word).
   (d) (15 points) **Feature Quantization and Histogram Computation**: Quantize all the clusterpoints to the cluster centroid. For each feature extraction in training or test image perform similar method as part (a). Find the index of the nearest cluster center in the dictionary. Represent each image by the histogram of the visual word indices. Normalize the Histogram.
   (e) (10 points) **Classifier Training**: Implement nearest neighbor (NN) classifier to classify each test image to one of the two classes (vehicle/non-vehicle).
2. (50 points) You have to build a point based tracker for object tracking in the given video and evaluate tracker's peformance on the tracking video dataset. The dataset contains five videos with their groundtruth bounding box in a text file. The groundtruth has format $(X_0, Y_0, Width, Height)$. To build a point based tracker follow the steps:
   (a) Use the groudtruth bounding box around the object of interest in the first frame.

(b) (5 points) Detect feature points (e.g. Harris Corner etc.) in the object region, which you will be tracking for the rest of the video.

(c) (5 points) Use Lucas Kanade algorithm to estimate the feature point location in the next frame.

(d) (5 points) Use the feature points from the previous frame and current frame (i.e. computed in the previous step) to estimate an affine transformation between the frames.

(e) (5 points) Use the affine transform computed to get the new points for the bounding box.

(f) Retain the bounding box information of the frame, we will need it in the next step for analysis. Perform step (c) to (f) for every frame.
(**Note:** You are not permitted to use the feature detector or the ground truth bounding box information after the first frame.)

(g) After you have populated the bounding box information for all videos by performing previous steps, you have to evaluate the performance of your tracking algorithm using Intersection-over-Union (IoU) evaluation metric (ranges from 0 to 1). IoU metric computes overlap of the predicted bounding box with the groundtruth bounding box (for more information read the link provided). Here, you will be needing the groundtruth information from the dataset.

(h) Compute and save the IoU metric for all frames in each of the videos.

(i) (10 points) In your report, you have to include the plot of tracking success (S) vs IoU Threshold (τ).

- You will vary τ from 0 to 1.
- For each value of τ find the Number of Frames in the each video that have IoU above IT value. Let's assume for a video-i, total frames are TF_i and out of which NF_i frames have IoU value above τ.
- Suppose while evaluating on video-1 and video-2, 50 out of 100 frames from video-1 and 40 out of 50 frames from video-2 have IoU value above τ. So, for video-1 $NF_1 = 50$ and $TF_1 = 100$ and for video-2 $NF_2 = 40$ and $TF_2 = 50$. Then tracking success S value for that τ value will be 0.6 i.e. $\frac{(50+40)}{(100+50)}$).

$$S_\tau = \frac{\sum_{i=1}^{N} NF_i}{\sum_{i=1}^{N} TF_i}$$

  where, N is total number of videos and $\tau \in [0, 1]$ is the threshold value.
- Find tracking success S for at least 10 τ threshold and plot it for the given dataset. Include the plot in your report.

(j)
- (5 points) Make changes to given algorithm in order to achieve better performance.
- (10 points) Do the performance evaluation for your new algorithm and compare it with the previous algorithm. Report the comparative S vs τ plot for both algorithms.

- (5 points) Write in brief, what changes you made and explain how those changes are helping you to achieve better performance.

(**Note:** Both your trackers should be in working condition at the time of your Demo or you will get 0 in the respective sections.)