COL780 Fall 2021
Assignment 1
Background Subtraction

Due Date: TBD

1 Introduction

Background subtraction is a well-known technique for extracting the foreground objects in images or videos. The main aim of background subtraction is to separate moving object foreground from the background in a video, which makes the subsequent video processing tasks easier and more efficient. Usually, the foreground object masks are obtained by performing a subtraction between the current frame and a background model. The background model contains the characteristics of the background or static part of a scene. Figure 1 shows a high level schematic of the background subtraction technique. In this assignment, you will be using background subtraction to find the foreground object masks for different scene conditions. We expect you to tackle the following four scene conditions.

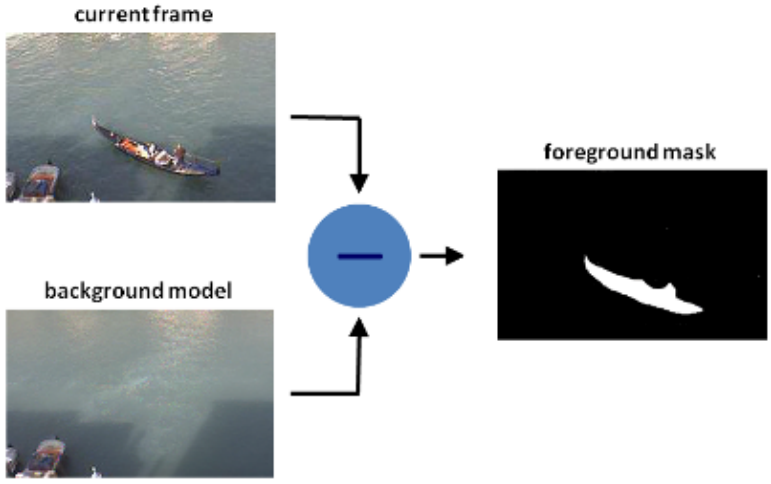

Figure 1: Key idea of the Background Subtraction technique.

2 Baseline

This category contains simple videos. The camera is fixed and steady. The background is static, and there is no change in illumination conditions between the video frames. Handling the baseline data is the first step towards building a robust background subtraction system.

3 Illumination changes

We add more complexity to the input videos. The lighting conditions may vary between the frames of a video. For example, an outdoor video may get affected by the variation of the sunlight. Changes in lighting conditions can also introduce shadows of the foreground objects, which need to be ignored.

4 Camera shake (jitter)

The camera may not always be steady, and it can shake due to the vibration of the mount or the unsteady hand of a photographer. In this step, your task is to ignore the changes introduced due to camera shake while detecting a foreground object mask.

5 Dynamic scenes: changing background

As the name suggests, changing backgrounds refer to videos where the background is not static anymore. The change in the background in this case is due to movements of background objects. For example, consider detecting a boat in the sea. The background sea will be dynamic due to the waves. Or movement of tree leaves due to wind.

6 Extra Credit: Pan-tilt-zoom videos

In a more advanced setting, the camera view can be dynamic. Although this is not a part of the main assignment, you may choose to attempt this part for extra credit. Pan-tilt-zoom (PTZ) cameras expand the level of flexibility since the desired camera viewpoint can be selected at run-time, which is not possible with standard fixed-camera solutions. PTZ cameras are applied in numerous applications, such as video surveillance.

7 Dataset

The validation dataset can be downloaded from this link. The dataset contains the following five folders - `baseline, illumination, jitter, moving_bg,` and `ptz`. In each folder, you will find the required video sequences. Note that the `ptz` folder contains the video sequences for pan-tilt-zoom, which you may optionally attempt as extra credit. The `input` sub-folder contains the extracted video frames, and the `groundtruth` sub-folder contains the ground-truth binary masks. The black regions (pixel value: 0) are the background, and the white regions (pixel value: 255) are the foreground objects. The `eval_frames.txt` file contains the starting and ending frame that will be used for evaluation. All input frames within this range (including the starting and ending frames) are considered evaluation frames for the video sequences. The remaining frames are background frames.

8 Evaluation

The efficacy of your solution will be judged by the intersection-over-union (IOU) scores between the ground truth and the predicted masks for each frame. Use the `eval.py` script to get the mean IOU scores for each video sequences.

```
python eval.py  --pred_path=<path to generated masks folder> \
                --gt_path=<path to groundtruth masks folder>
```

You need to complete the `main.py` to generate the masks. You may create additional source files that will be used by the `main.py` to compute the foreground masks. The usage of the `main.py` should be the following

```
python main.py  --inp_path=<path to input frames> \
                --out_path=<path to generated masks> \
                --eval_frames=<path to eval_frames.txt file> \
                --category="<b/i/j/m/p>"
```

The `category` argument is used for determining the scene category. `b, i, j, m,` and `p` refer to baseline, illumination, jitter, moving background (dynamic scene), and PTZ categories, respectively. We will evaluate your method on held-out test video sequences for efficiency, apart from the provided validation videos. The `eval.py` can be downloaded from here and `main.py` can be downloaded from here.

9 Submission

You are required to submit your source code (including the completed `main.py`) and a report containing the following information:

1. The methods you used to generate the foreground masks for each scene category.
2. The mean IOU scores for each video sequence.
3. A link to the generated output masks for each video.
4. Discussion of the bonus part, if attempted.

Note: Generate IOU scores and masks only for the evaluation frames of a video.
You must zip the code and report in a single zip file, rename the zip file as `<Your-Entry-Number>.zip` (e.g., `2019CSZ8406.zip`), and submit in Moodle.

10 Rubric

1. Full marks - 10
   (a) Baseline – 3 Marks
   (b) Lighting changes – 2 Marks
   (c) Camera shaking – 2 Marks
   (d) Moving background – 2 Marks
   (e) Report – 1 Mark
2. Extra credit: PTZ – 1 Mark

11 Constraints

Apart from the standard libraries and packages, you are allowed to use only `OpenCV` as the third-party library. No other third-party library is allowed. You are allowed to use either `Python 3.6+` (preferred) or `C++`.