COL780 Assignment 1
Background Subtraction

Aditi Khandelwal Shubham Mittal
2018EE10434 2018EE10957

September 27, 2021

In this assignment, we discuss various background subtraction techniques for different scene categories such as static (our baseline), moving background, jitter, Pan-tilt zoom videos and changing illumination. We report our best mIOU scores for the given video sequences in Table 1. Link to the generated output masks for each video sequence can be found here[1].

Table 1: Performance of best model on different scene conditions

Category	mIOU
Baseline	0.7213
Illumination	0.4936
Jitter	0.6770
Moving Background	0.4353*
Pan-Tilt-Zoom	0.0638

*For moving background condition, few evaluation frames had no foreground because of which the score dropped. If we do not take into account those frames in calculating the mIOU, we get a score of 0.65.

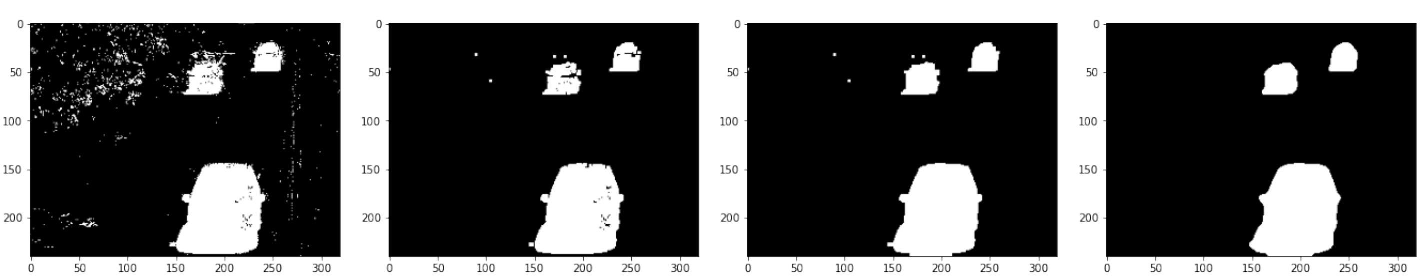
1 Baseline - Static Background

For this scene, the camera is assumed to be fixed and steady. The background is assumed to be static and have no illumination changes between any two video frames.

1.1 Proposed Method:

Our method consists of following steps:

1. KNN Model
2. Morphological transformations
3. Foreground enhancement using median filtering



(a) KNN output (b) Morphology Open (c) Morphology Close (d) Median Blur

Figure 1: Pipeline to detect foreground from static background

[1] Code will be released at https://github.com/sm354/Background-Subtraction

1.2 Discussion:

Our background model is shown in Figure 2(a). The outputs of various steps of our proposed method are shown in Figure 1. We give following reasoning behind our pipeline architecture:

1. We use KNN model as empirically it turned out to be better than GMM model (like MOG2 in OpenCV). The background model obtained using KNN model is shown in Figure 2(a). There is some level of salt and pepper noise which we deal with later in the following steps of the pipeline.

2. Upon observing the given video carefully we notice that the background is not completely static - there are moving leaves in it. Such noises are inevitable in real data, thus we use morphological operations to remove the noise. We use morphology open and close operations to remove the noise in background and foreground respectively.

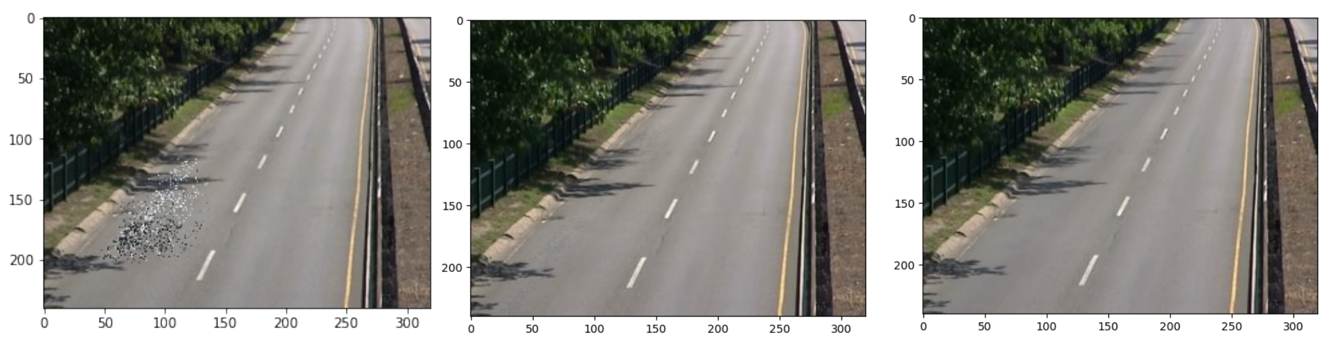
3. Finally, we smoothen the foreground boundary and remove the remaining (salt and pepper) noise using median filter.

1.3 Other methods tried

In order to build a strong baseline model (that will be used for the other scene categories), we try various different models, and techniques of pre-processing and post-processing.

1.3.1 Background subtraction models

We experiment with different models shown in Figure 2. The averaging model is obtained by taking a mean of all the frames in the train set. We can observe that both the MOG2 and averaging model are "smoother" and noise-free than the KNN model. However, upon adding the pre/post-processing techniques (discussed above and below also), the KNN model turned out to be the best.



(a) KNN (0.55) (b) MOG2 (0.48) (c) Averaging (0.49)

Figure 2: Different Background Models with mIOU scores (scores obtained without any pre/post processing and just the model itself)

1.3.2 Pre and Post Processing Techniques

We define pre-processing techniques as those transformations applied on images passed into the model (like KNN, MOG2), and post-processing techniques as those applied on the output of the model i.e. on the predicted masks.

Pre-processing techniques: we experimented with noise removal techniques like blurring using gaussian kernel but we got negligible improvements in the mIOU score. We also experimented with image sharpening kernels (made using LoG kernel) but this also didn't give considerable improvement in the scores.

Post-processing techniques: the predicted masks obtained from the model were observed to contain lot of noise (as shown in Figure 1). We experimented with Gaussian kernel, median filter, and morphological operations (erosion and dilation). Morphological operators turned out to be the best for noise removal in the predicted masks. After passing the masks through morphological operations, we again apply median filter to remove the left out noise.

2 Illumination

Having built a strong baseline model, we add illumination invariance to it for this scene category. Since the given video sequence also has static background except that it has changing illumination, our approach is to remove/normalize the illumination information so that we can pass the pre-processed frames into the baseline pipeline (proposed above).

2.1 Proposed Method:

Our method consists of two pre-processing steps that counter the varying illumination problem, and finally the baseline model pipeline (discussed above):

1. Histogram Equalization
2. Channel-wise Adaptive Thresholding
3. Baseline model pipeline: KNN + Morphological transformations + Foreground smoothening

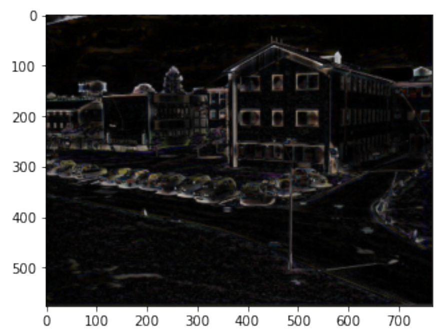


Figure 3: Background Model

2.2 Discussion:

Our background model is shown in Figure 3. We explain our proposed method pipeline:

1. **Histogram Equalization (HE)**
   (a) We use histogram equalization for tackling the varying illumination problem. We first convert the color scheme to YCrCb which separates the brightness/illumination information of image unlike RGB or BGR color scheme. We may also use HSV color scheme but empirically using YCrCb turned out to be better.
   (b) After going into the correct color scheme we use Clahe's algorithm for histogram equalization on the Y-channel[2]. Clahe's algorithm is effective in performing piece-wise (localized) histogram equalization. Finally we convert back to BGR color scheme.
2. **Channel-wise Adaptive Thresholding (AT)**: We implement adaptive thresholding (described below) on each channel (of bgr image) and merge all the processed channels to get final image.
   (a) we apply dilation operator on the channel followed by median blur. the dilation operation with kernel of 3 x 3 increases the blockiness and median kernel does the final smoothening.
   (b) then we take the absolute difference of above processed channel and the original channel to get the prominent features (like edges)
   (c) finally we do normalization on the processed channel
3. The pre-processing stages are shown in Figure 4 and 5 for bright and dark condition frames respectively. We can observe that both the pre-processed frames are similar and varying illumination information is removed.

[2] this channel represents the brightness in the image

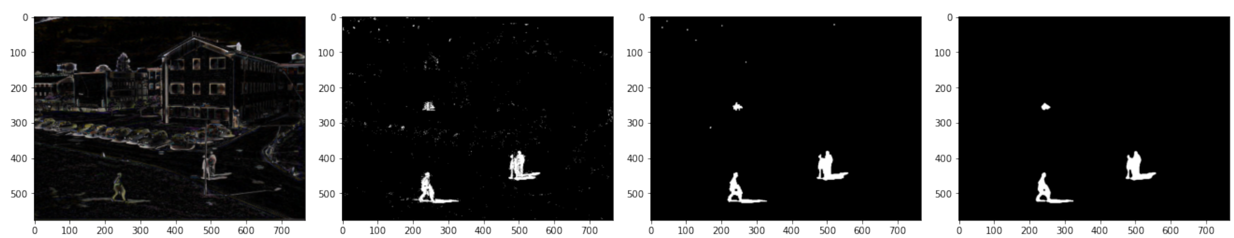
(a) Bright Condition Frame (b) Histogram Equalization (c) Adaptive Thresholding

Figure 4: Pre-processing on Bright Condition Frame

(a) Dark Condition Frame (b) Histogram Equalization (c) Adaptive Thresholding

Figure 5: Pre-processing on Dark Condition Frame

4. Finally, after pre-processing the given frame, we pass it through the baseline model pipeline like earlier. The illumination-invariant background subtraction pipeline is shown in Figure 6.



(a) Pre-processed (HE+AT) (b) KNN output (c) Morphology Operations (d) Median Blur

Figure 6: Pipeline to detect foreground from background with changing illumination

5. **Limitation** we observe in Figure 6(d) that the model also includes shadows in the foreground.

3 Jitter

The background model for video sequences with jitter can be seen as static background + jitter and no other noise like varying illumination. Thus, we aim to improve upon the baseline model pipeline with emphasis on removing the noise incurred due to the jitter in the predicted masks. Our model pipeline for this scene category is the baseline model pipeline with extra targetted noise removal techniques. Our background model for the given video sequence is shown in 7

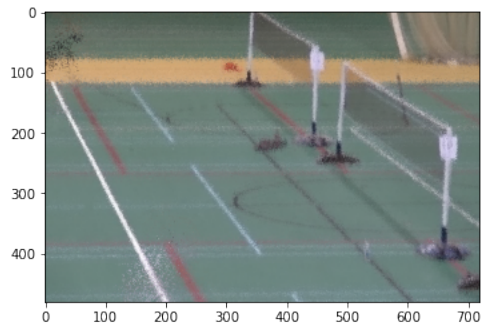
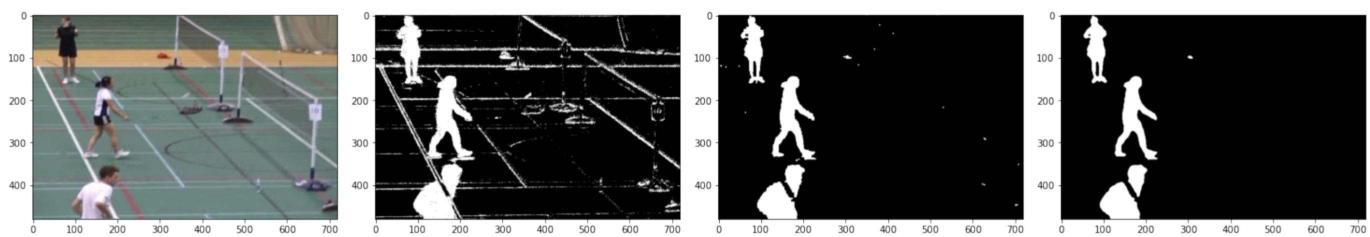


Figure 7: Background Model

3.1 Discussion:

The outputs of various steps of our proposed method are shown in Figure 8. As discussed earlier, we apply KNN model for jitter scene condition as well. We can see that there are edges (prominent badminton court lines) also present in addition to the salt and pepper noise that was there in Figure 1(a). Thus, to remove the edges we use erosion for thinning (and eroding away) such type of noise. We still observe salt and pepper noise in the obtained image for which we apply erosion followed by dilation which is implemented in OpenCV as morphology open. The foreground also has noise, which is tackled by the morphology close. Finally, we apply median blurring to smoothen the foreground and remove left out noise.



(a) Input frame (b) KNN output (c) Morphology (d) Median Blurring

Figure 8: Pipeline to detect foreground from background with jitter

4 Dynamic Background

For this scene category we again improve upon the baseline model pipeline by removing the noise incurred due to ever moving background. Our proposed model pipeline is the baseline model pipeline with extra targetted noise removal techniques. Our background model for the given video sequence is shown in 9

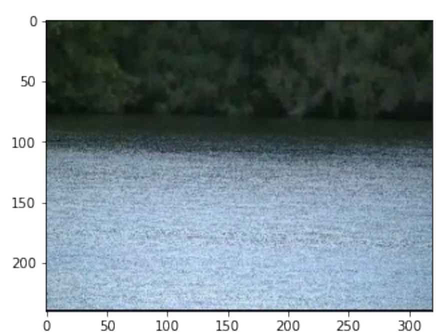
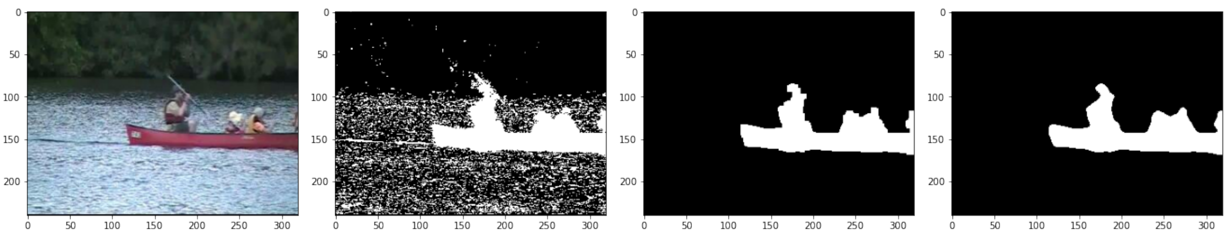


Figure 9: Background Model

4.1 Discussion

The outputs of various steps of our proposed method are shown in Figure 10. KNN model is applied which produces the foreground as seen in Figure 10(b). For this dynamic condition, initial foreground obtained has a lot of noise because the background is ever moving. We apply the following noise removal techniques to obtain noiseless foreground:

1. The moving water ripples induce thick noise in the image. To handle such kind of a noise, we erode the image obtained twice to lessen the noise. This also erodes away the foreground which we tackle by applying dilation twice to rectify the change in the foreground. On doing such a morphology, we were able to get rid of all the noise present in the image.
2. The foreground obtained after morphology has very rough edges which are smoothen by applying the median filter which also helps us get rid of any salt and pepper noise if left.
3. We also experimented with contour filling to fill the noise due to ripples. This noise removal gave similar performance like morphology operators but we dropped this to make our model pipeline generic.

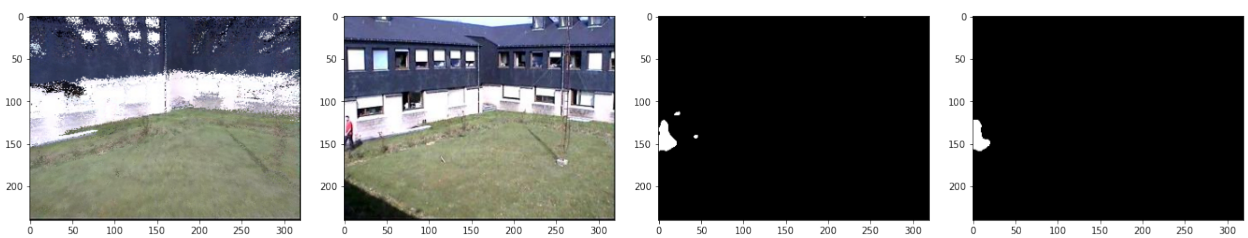


(a) Sample Image (b) Foreground Obtained (c) Morphology (d) Median Blur

Figure 10: Pipeline to detect foreground from moving background

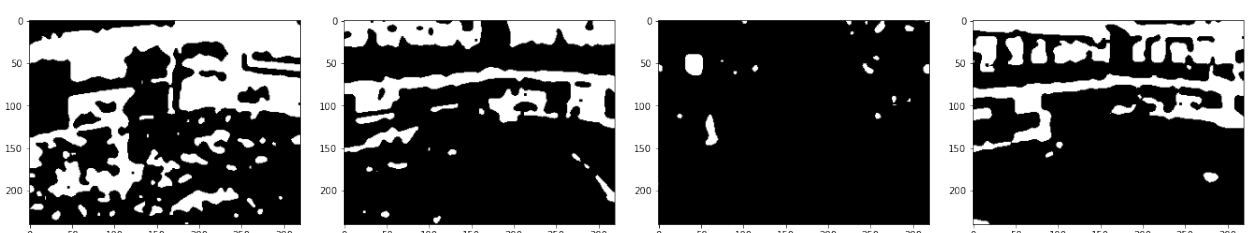
5 Pan-Tilt-Zoom

We applied various techniques (discussed in Section 1.3) to obtain the foreground in PTZ condition. The results presented below are obtained from our baseline methodology as proposed above. As can be seen from the Figure 12 that the obtained foregrounds are extremely noisy. The best foreground obtained can be witnessed on Figure 11. We conclude that the baseline works only for a small set of images and for the rest, it performs very poorly.



(a) Background Model (b) Sample Image (c) Morphology (d) Median Blur

Figure 11: Pipeline to detect foreground in PTZ video condition



(a) Noisy Foreground Sample 1 (b) Noisy Foreground Sample 2 (c) Noisy Foreground Sample 3 (d) Noisy Foreground Sample 4

Figure 12: Performance of PTZ on large set of images