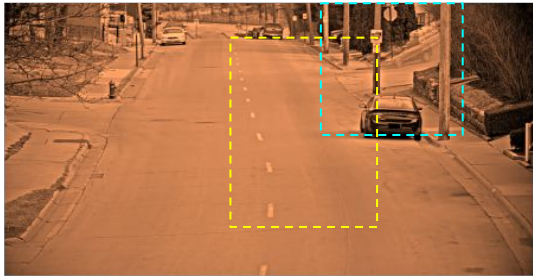
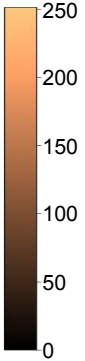
Task 1.1: Gaussian modelling (Team 5) - (1/3)

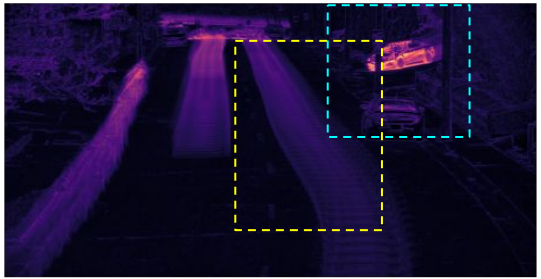
Statistics from the first 25% of frames (1 to 535) to define background:

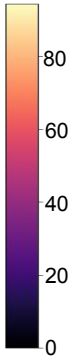
Mean values μ



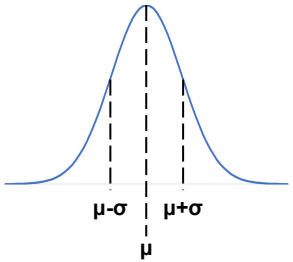


Standard deviation σ





Gaussian distribution model for every **pixel value**



The trajectories of cars that pass through the road during the frames is apparent as…

A car shows up after some frames and remains static for a while, leaving…

…slightly darker **(lower) mean values** due to the darker tone of most cars compared to the asphalt, as well as lighter **(higher) variance levels** compared to areas that experience minimal change.

Task 1.1: Gaussian modelling (Team 5) - (2/3)

The constructed distributions can be interpreted in a way that allows distinguishing background from foreground. It involves following the next algorithm:

for all pixels i **do**
  if $|I_i - \mu_i| \geq \alpha \cdot (\sigma_i + 2)$
  (A = $|I_i - \mu_i|$; B = $\alpha \cdot (\sigma_i + 2)$)
  then
    pixel → Foreground
  else
    pixel → Background
  end if
end for

The absolute difference between a given frame and the established mean (*A*) is evaluated for every pixel…

…and then compared to a weighted (*α*) version of the standard deviation (*B*)

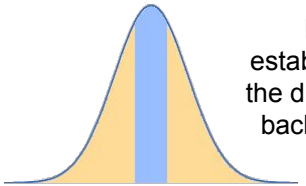
Note: A constant of 2 is summed to the standard deviation to avoid excessively low values

If the evaluated pixel is different from its mean value to the point where term *A* is equal or higher than term *B*…

…it is considered to be part of the foreground

On the contrary, if the difference in term *A* is lower than term *B*…

…it is considered to be part of the background



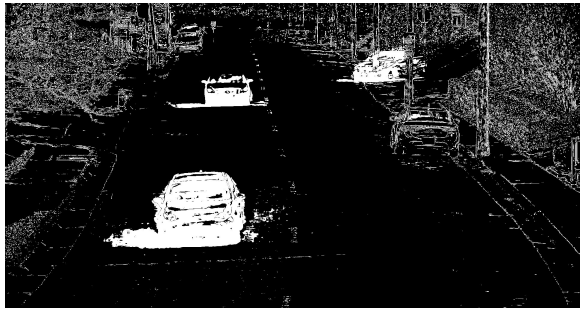
Basically a matter of establishing which portions of the distribution are considered background and foreground based on *α*.

Background Foreground

Task 1.1: Gaussian modelling (Team 5) - (3/3)

The effect of weight α can be well appreciated on the masks for background and foreground distinction:

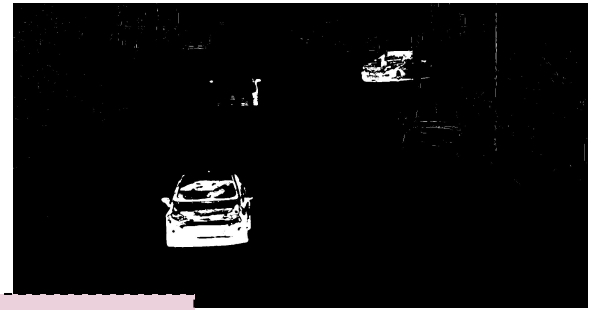
$\alpha = 2$



$\alpha = 4$

$\alpha = 6$

$\alpha = 8$



Examples for frame 548

Lower values result in areas with detailed texture (e.g., trees, stone walls, etc.) to easily be identified as foreground as soon as minor noise is present between frames.

\+ Higher detail - More noise

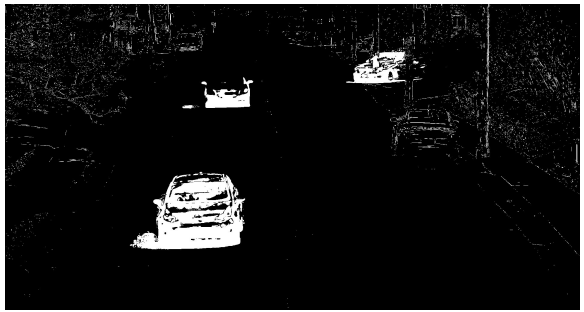
Higher values are more robust to noise throughout frames. However, parts of what should be considered as foreground are more easily confused with the background (e.g., reflections in parts of the cars).

\- Missing parts + Less noise

Task 1.2: mAP vs alpha (Team 5) - (1/3)

Since the masks obtained by simply applying the presented algorithm contain many imperfections, a series of steps are implemented before proceeding to the evaluation of the system:

Original mask



Applying ROI mask

A region-of-interest (ROI) mask is provided, which is added to the original mask in a bitwise_and manner to ensure no pixels are detected as foreground in the areas where there should not be any important activity.

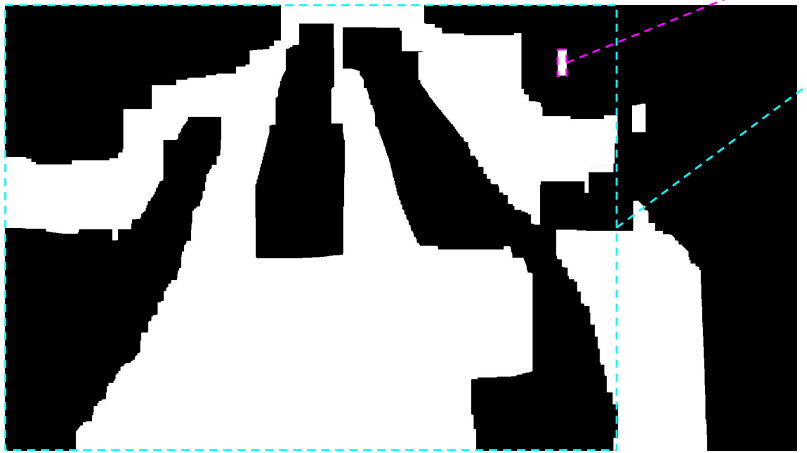
Morphological operations

Several morph. ops. are applied:

1. Opening (5x5 kernel)
   For removal of small noise
2. Closing (1x80 kernel)
   For vertical filling
3. Closing (80x1 kernel)
   For horizontal filling
4. Opening (10x60)
   Removal of elements like shadows

Task 1.2: mAP vs alpha (Team 5) - (2/3)

Using only the first 25% of frames to determine the characteristics of the background implies that the system will increasingly struggle over time due to progressive and sudden changes.



Example of failed background/foreground distinction in a situation with a significant illumination change

Example of a small erroneous detection

Example of a large erroneous detection

In an attempt to mitigate the influence of erroneous detections caused by these kinds of issues, several restrictions are enforced over the detected bounding boxes:

- **Minimum size**: width: 29px, height: 12px
- **Maximum size**: width: 593px, height: 442px
- **Aspect ratio range**: 0.4 < width/height < 2.5

Based on available ground truth

Task 1.2: mAP vs alpha (Team 5) - (3/3)

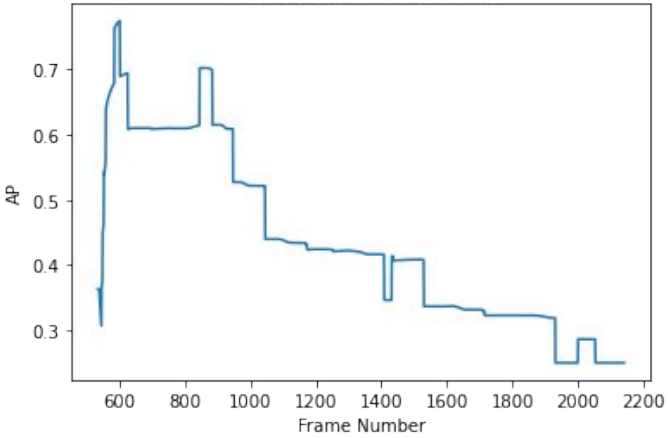
The AP of the system has been evaluated for different values of the weight α. The effect of implementing two different kinds of low-pass filters on each frame prior to the explained process is also evaluated.

$AP_{0.5}$ results for different α and blur

	α						
	3	**3.57**	**4.14**	**4.71**	**5.29**	**5.86**	**6.43**
No blur	0.228	0.242	0.195	0.203	0.209	0.160	0.155
Gaussian blur							
5x5	0.219	0.234	0.197	0.164	0.163	0.159	0.158
7x7	0.195	0.226	0.197	0.162	0.161	0.157	0.155
Median filter							
5x5	0.222	0.242	0.213	0.200	0.209	0.162	0.151
7x7	0.220	0.243	0.207	0.202	0.212	0.167	0.151

An α of ~3.57 appears to yield the best results, although this depends on aspects such as the used morphological operations. The low-pass filters seem to contribute to a slightly better performance, especially the median one set to a kernel of 7x7.

Accumulated AP over time (best case)



As expected, the system becomes less precise as time passes and the initial characteristics of the background become less representative of the current frame.

Task 2.1: Adaptive modelling (Team 5) - (1/2)

An adaptive model is proposed as a solution to the limitations of the presented model. The idea here is to once again use the 25% frames to extract background statistics for each pixel, but then to slightly update these on the go. This is implemented according to the following algorithm:

for all pixels i **do**
  if $|I_i - \mu_i| \geq \alpha \cdot (\sigma_i + 2)$ **then**
    pixel $\rightarrow$ Foreground
  else
    pixel $\rightarrow$ Background
  end if
end for

Same principle used in the previous model

if pixel $i \in$ Background **then**

$$\mu_i = \rho \cdot I_i + (1 - \rho) \cdot \mu_i$$

$$\sigma_i^2 = \rho \cdot (I_i - \mu_i)^2 + (1 - \rho) \cdot \sigma_i^2$$

end if

For every pixel that is identified as being part of the background, its mean is updated based on a weight ρ that determines how much of an influence its current value should have on the mean.

High $\rho \rightarrow$ More influence on mean

The same process applies similarly to the variance of the pixel in question

High $\rho \rightarrow$ More influence on variance

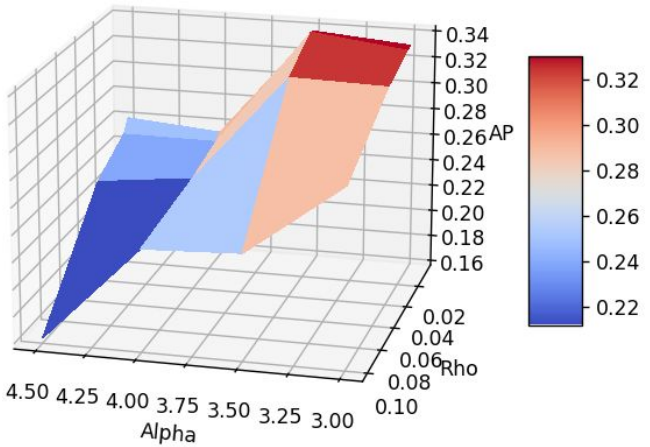
Task 2.1: Adaptive modelling (Team 5) - (2/2)

There are now two hyperparameters to adjust, *α* and *ρ*. A couple of options can be pursued to approximate which combination of these lead to the best performing model:

- **Estimate *α* for a non-recursive model and then *ρ* for a recursive one**. While this option would be rather simple and not require as much resources, the results would not be as optimal due to how these two hyperparameters influence each other.
- **Estimate both *α* and *ρ* at once through a grid/random search**. This second option is a lot more resource intensive, but adjusting both hyperparameters together should help reach a more optimal setting.

A grid search is in this case used, providing the following insights:

		ρ			
		0.005	0.01	0.04	0.1
α	3.0	0.327	0.326	0.317	0.285
	3.5	0.333	0.335	0.318	0.230
	4.0	0.249	0.245	0.236	0.227
	4.5	0.255	0.245	0.228	0.155



The search could be more fine-grained, but it was limited to the space above because of time limitations.

Task 2.2: Comparison adaptive vs non (Team 5) - (1/3)

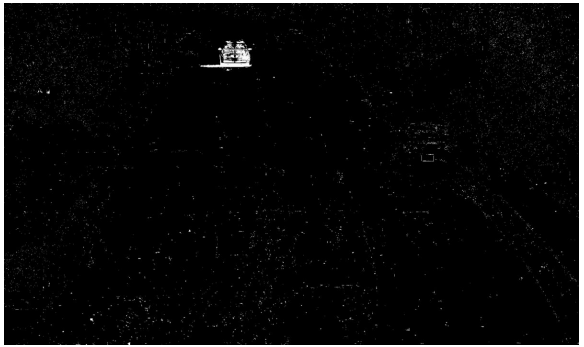
By introducing this capability to update the background statistics throughout frames, the model is now more robust to changing conditions such as the alterations in illumination mentioned before. The animations below showcase an example where this improvement can be clearly appreciated:

Original clip

Set of frames with a significant change in scene lighting

Normal model (α = 4)

Adaptive model (α = 4, ρ = 0.04)



While the normal model is not capable of performing a good separation of foreground from background when the general values in a frame deviate from the rigid initial statistics, the adaptive version is able to update these statistics and perform more consistent estimations.

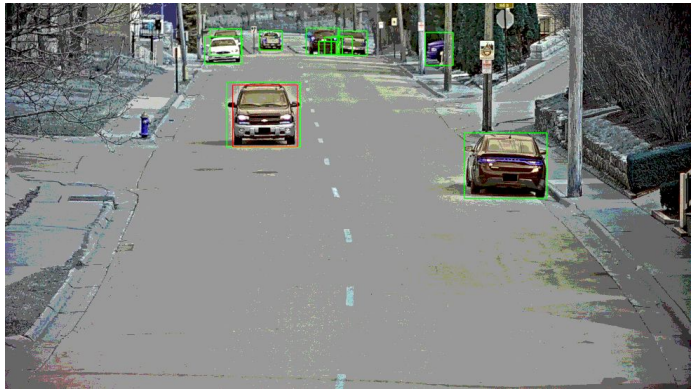
Task 2.2: Comparison adaptive vs non (Team 5) - (2/3)

This increase in robustness in terms of background and foreground distinction helps in the determination of bounding box detections, as the masks remain more consistent throughout frames.

Normal model (α = 4)

Plenty of incorrect detections in several frames

Adaptive model (α = 4, ρ = 0.04)

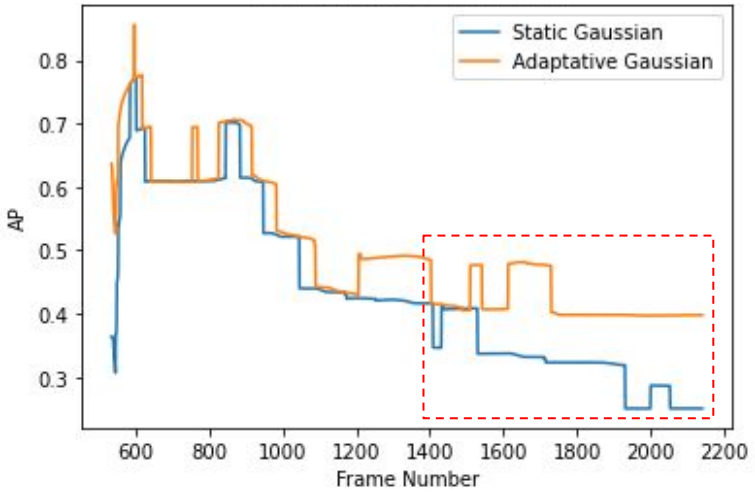


Reduction in erroneous bounding boxes

Task 2.2: Comparison adaptive vs non (Team 5) - (3/3)

A further comparison between the normal and the adaptive models is displayed in the graph below:

Accumulated AP over time for normal and adaptive models (best cases*)



The general AP values appear to be higher for the adaptive model, as it is to be expected.

It is particularly interesting to see how the difference is especially noticeable towards the final frames (after around frame 1500), where the AP values of the adaptive model remain significantly higher than those of the normal model. This is in line with the behavior of the bounding boxes observed in the previous slide.

*Best cases meaning:

- Normal/static with α = 3.57 and median filter (7x7)
- Adaptive with α = 4, ρ = 0.04 and median filter (7x7)

T3 Comparison with state-of-the-art for (Team 5) - (1/3)

OpenCV Libraries

MOG is a Gaussian Mixture-based Algorithm. It models each background pixel by a mixture of K Gaussian distributions (**K = 3 to 5**). The weights of the mixture represent the time proportions that those colours stay in the scene. The probable background colours are the ones which stay longer and more static.

MOG2 improves MOG by selecting the appropriate **number of gaussian distribution** for each pixel. This leads to better adaptability to varying scenes due sudden illumination changes.

KNN[2] introduces a **new kernel method** with large kernels in the areas with a small number of samples and smaller kernels in the densely populated area.

BGSLibrary [1] is part of a comprehensive study on background subtraction, in which it compares the performance of **29** background subtraction algorithms.

Due to time constraints we only included the following methods in this slide:

- **Adaptive Background Learning (ABL)**
- **Frame Difference (FD)**
- **Static Frame Difference (SFD)**
- **Weighted Moving Mean (WMM)**
- **Weighted Moving Variance (WMV)**

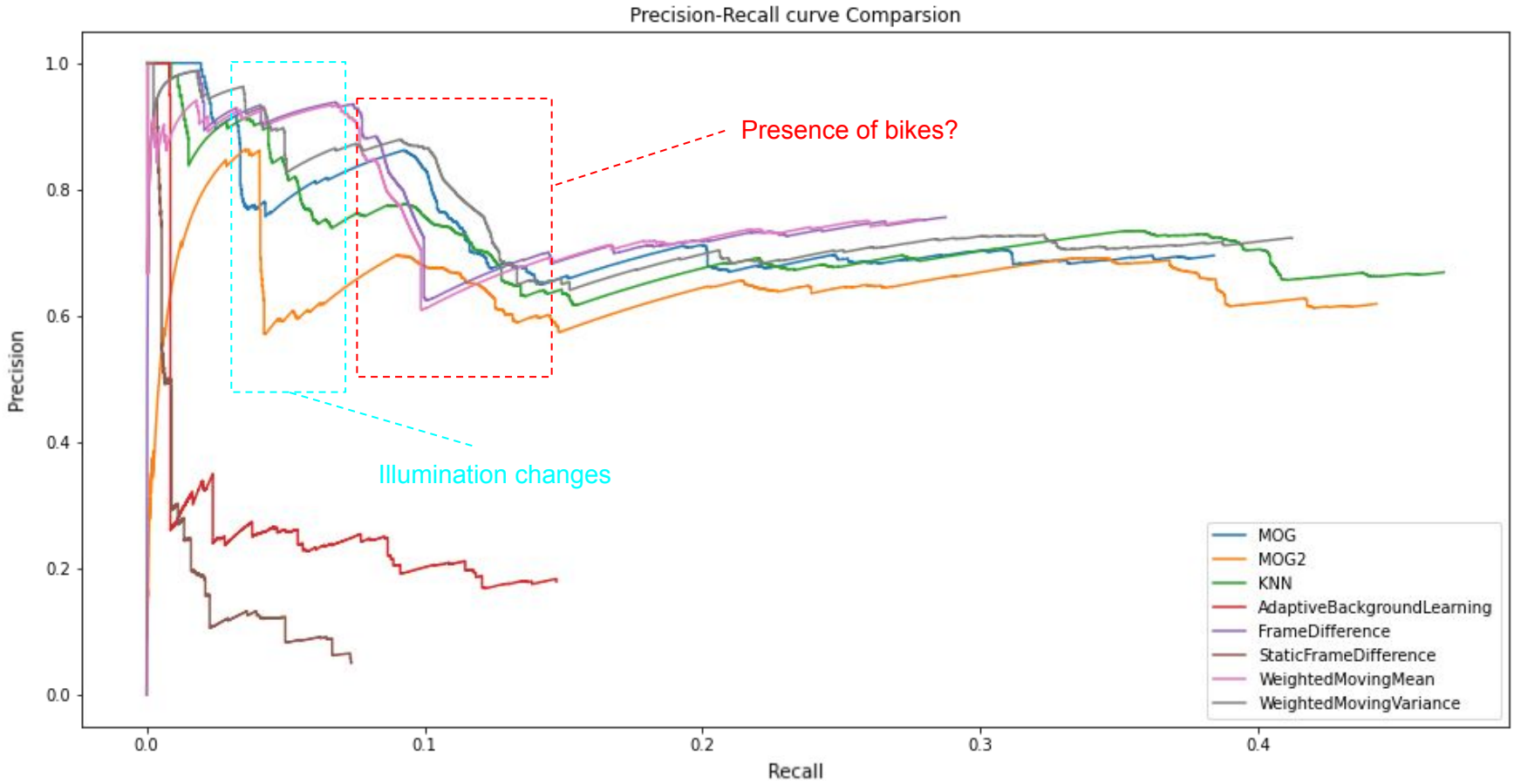
Some extra results on github.

Note: the same **post-process filters** were used for all the methods before detection..This means the result shown does not truly reflect the performance of these algorithms, as some might benefit more from the filtering than the others.

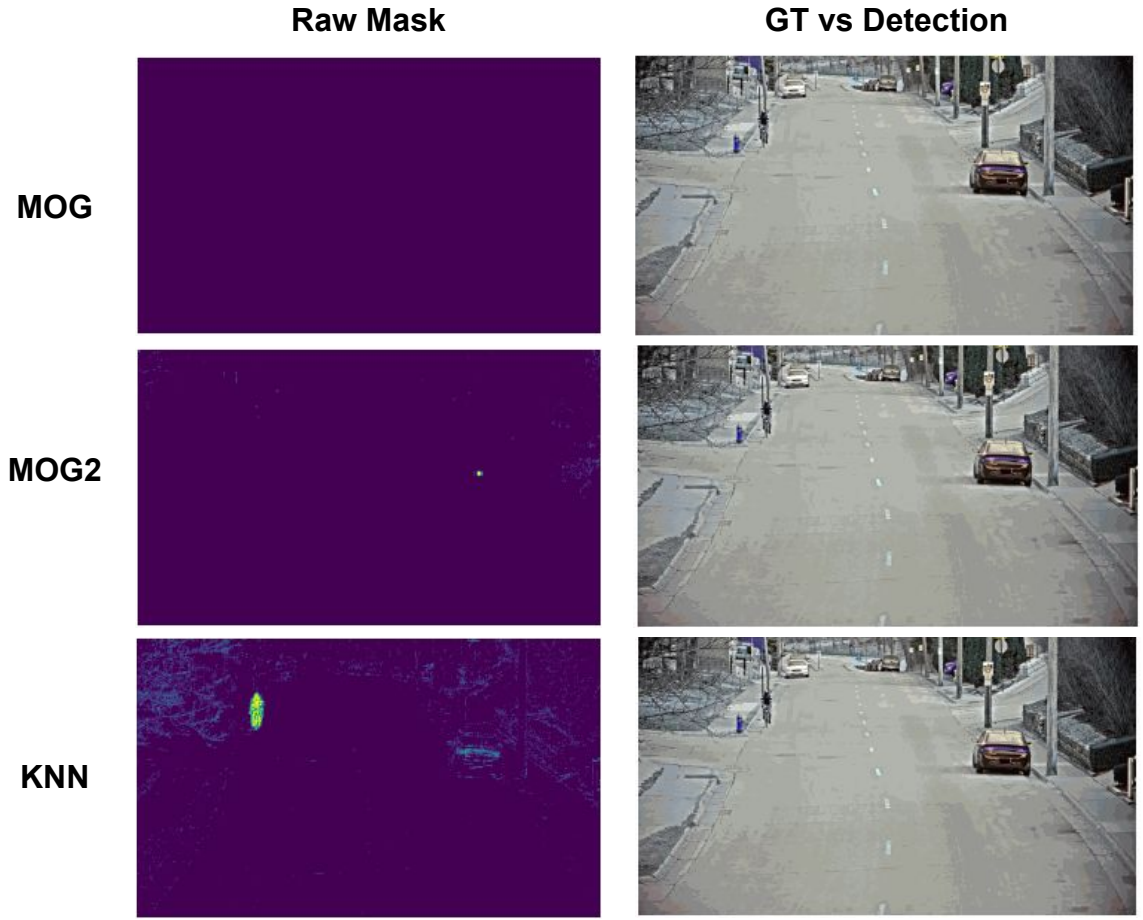
Method	AP
MOG	0.29
MOG2	0.32
KNN	0.35
ABL	0.11
FD	0.23
SFD	0.09
WMM	0.23
WMV	0.36

[1] Sobral, Andrews, and Antoine Vacavant. "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos." *Computer Vision and Image Understanding* 122 (2014): 4-21.
[2] Zivkovic, Zoran, and Ferdinand Van Der Heijden. "Efficient adaptive density estimation per image pixel for the task of background subtraction." Pattern recognition letters 27.7 (2006): 773-780.

T3 Comparison with state-of-the-art for (Team 5) - (1/3)



T3 Comparison with state-of-the-art for (Team 5) - (2/3)



MOG can not distinguish shadow from the foreground object, which results in oversized bounding box.

MOG2 and **KNN** have built-in shadow detection capabilities and it significantly improves the localization accuracy. The use of pre-computed background statistics (history=200) of these methods can cause a surge in false positives soon after a sudden illumination change in the scene.

The low **AP** score across the three methods can also be explained by the false positives from detecting the moving **bikes** that are not part of the ground truth.

T3 Color (Team 5) - (1/2)

Adaptive methods with color frames

RGB: All channels contain chroma and lightness information.

LAB: Chroma is independent of the light intensity

HSV: Chroma (contained in H) is independent of the light intensity.

YUV: Chroma is independent of the light intensity.

Methodology

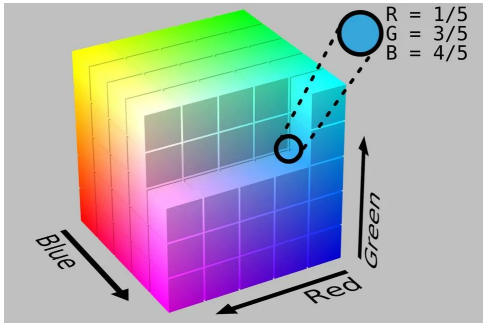
We use 20% of the frames instead of 25% due to memory limitations. Statistics (μ and σ) extracted and updated per channel. Mask is computed in regions where selected channels “agree”.

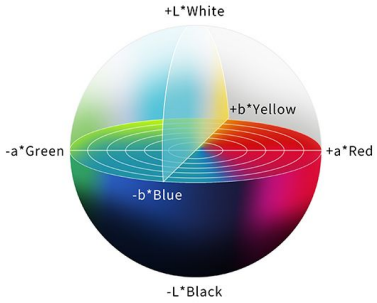
RGB : All channels are used because they contain both chroma and lightness information.

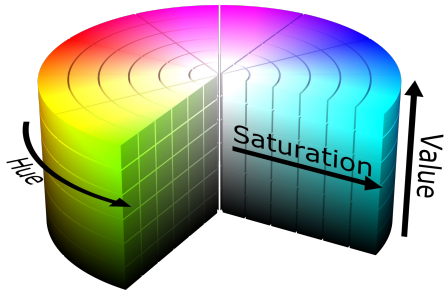
LAB : Chrominances AB used (Luminance discarded).

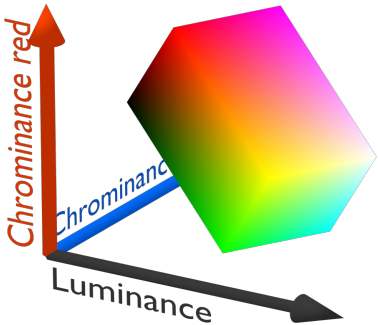
HSV : Hue and Saturation used (Value discarded).

YUV : Chrominances UV used (Luminance discarded).









T3 Color (Team 5) - (1/2)

Grid Search

Limited search space due to computation time.

Color spaces: [RGB, HSV, LAB]

α: [0.5, 1.5, 3]

σ: [1, 2, 3]

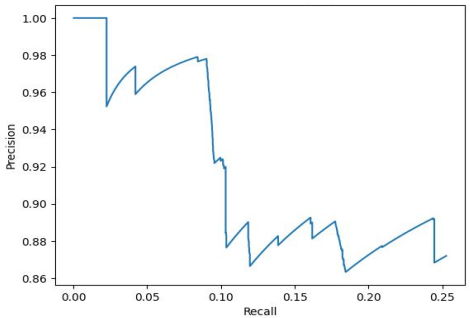
ρ: [0.005, 0.01, 0.1]

Results per color space

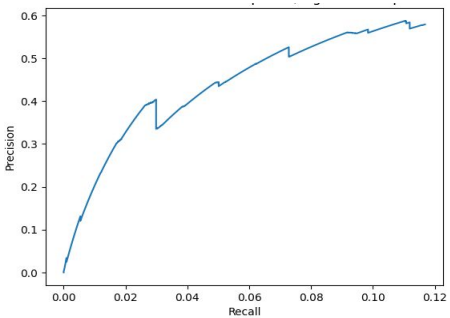
The best results obtained during the grid search over the defined search space are here shown*. RGB appears to yield the best results, although the improvement compared to the greyscale version of the model is quite limited.

	α	***σ***	***ρ***	**AP**
RGB	3	3	0.01	0.256
LAB	3	1	0.1	0.107
HSV	3	3	0.1	0.041

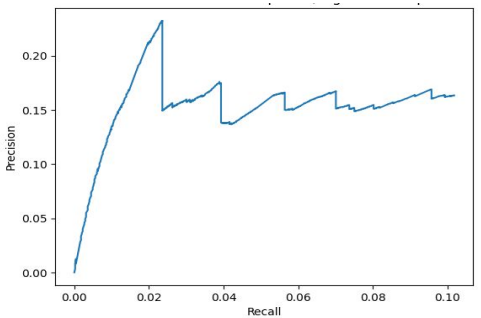
RGB



LAB



HSV



*These results were based on adapting an earlier version of the adaptive model with a maximum AP of 0.244 to 3 channels. An updated search could not be carried out due to time limitations.