Abandoned and Removed Object Classification

Ruairi Mullally – 22336002

Abstract:

This report presents the design, implementation, and testing of a dual-speed Gaussian Mixture Model (GMM) system for detecting abandoned and removed objects in static surveillance scenarios. The system employs two parallel GMMs with distinct learning rates to create a temporal filter that separates transient motion from persistent scene changes. Morphological filtering, and temporal persistence analysis refine detections and eliminate noise to produce a stable 'persistent mask' that highlights consistent change. Connected component analysis and spatial tracking are applied to track and merge detections across frames and stabilise object boundaries. Object classification is achieved through edge correlation analysis by comparing edge-maps of the persistent mask, the current frame, and the slow-GMM background. The relative overlap between mask edges and the current and background frames provides a robust classifier metric, with clear visual interpretability. Evaluation across multiple videos (n=10) demonstrate a high classification accuracy of 100% (for correctly identified objects), and a moderate spatial detection performance, with an average intersection-over-union of 69% and mean detection delay of 8.5 seconds. The model performs with an f1 score of 0.609 for strictly abandoned and removed objects, or an f1 score of 0.438 when accounting for objects that are moved within a scene (neither abandoned nor removed). Failure cases such as small objects, video initialisation artefacts, and failure to detect abandoned and subsequently removed objects highlight the model's sensitivity to lighting changes, dynamic backgrounds, and the system's dependence on pre-tuned parameters and video properties. Future improvements could include adaptive GMM learning rates, enhanced feature-based tracking, and additional classification metrics to handle more complex object interactions.

Introduction:

Automated detection of abandoned and removed objects is a key problem in video surveillance and public safety. The ability to distinguish between temporary motion and persistent scene changes are essential. Achieving this reliably is challenging due to varying lighting conditions, occlusions, and the dynamic nature of human activity. Simple background subtraction techniques often fail under these conditions. GMMs have become a standard probabilistic method for background model which provide some robustness against gradual scene changes (lighting, weather) that are not objects. In recent years the field has shifted from purely background-subtraction-based detection toward deep learning (such as YOLO) and context-aware methods [1] [2]. The report describes the framework of a dual-speed GMM and its associated post-processing stages, including morphological refinement, temporal persistence analysis, and simple object tracking. It then details the classification process based on edge correlation. A comprehensive evaluation is presented using multiple videos, illustrating both successful detections and failure cases. The system's detection accuracy, robustness, and limitations are analysed, and potential improvements are proposed for enhancing generalisation and adaptability in more complex environments.

Description of Object Detection:

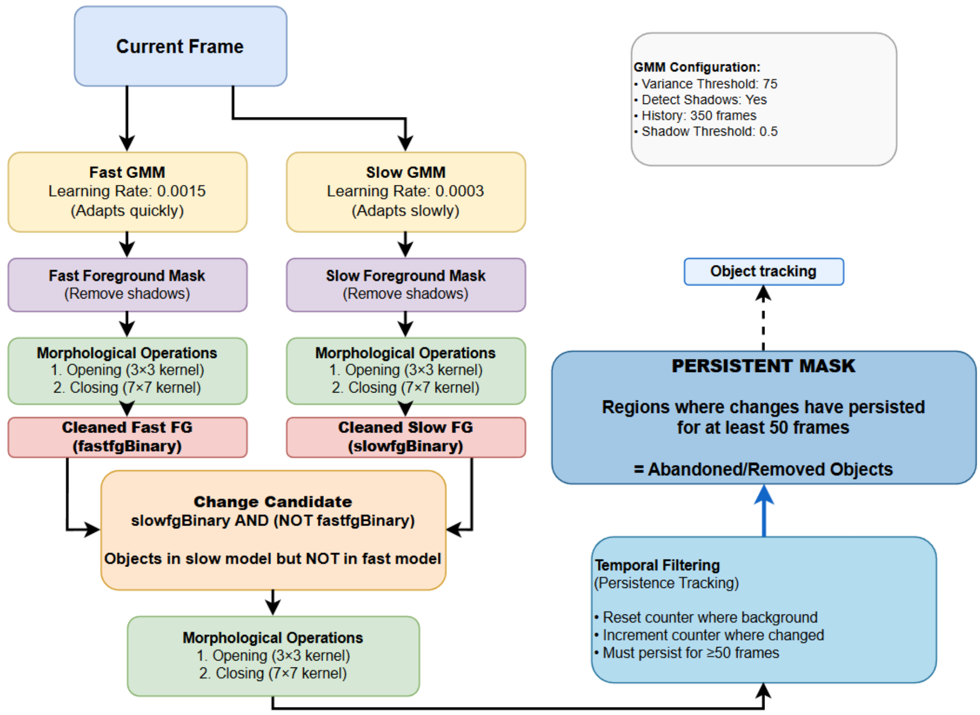

Fig 1 Dual-speed GMM Background Subtraction algorithm for abandoned/removed object detection.

Changes are detected by dual-speed background subtraction. The system uses two Gaussian mixture models (GMMs) running in parallel at different learning rates. The fast-GMM (0.0015), adapts quicker to persistent changes than the slow-GMM (0.0003). This pairing acts as a temporal filter: transient motion is absorbed by the fast model but not by the slow one, allowing sustained differences, such as abandoned or removed objects, to be isolated from ordinary movement. The probabilistic nature of GMMs provides robustness to gradual lighting variations and sensor noise.

Fig 2 Processing performed on GMM foreground masks.

For each frame of the video, the foreground masks for both GMMs are obtained, and thresholded to remove shadow pixels. Morphological opening and closing with 3x3 and 7x7 kernels respectively are applied to remove small artifacts and merge fragmented regions. A ‘change candidate’ is defined as the bitwise AND of the slow-GMM mask and the negated fast-GMM mask, highlighting areas where long-term and short-term backgrounds disagree. The same morphological operations refine the change candidate. This identifies areas that have changed from the original background but have remained consistently changed for some time. Temporal filtering is used to compute a ‘persistent mask’, whereby the difference must remain for at least 50 frames to be added. This makes the system more robust to slow-moving objects in the scene that were not necessarily abandoned or removed.

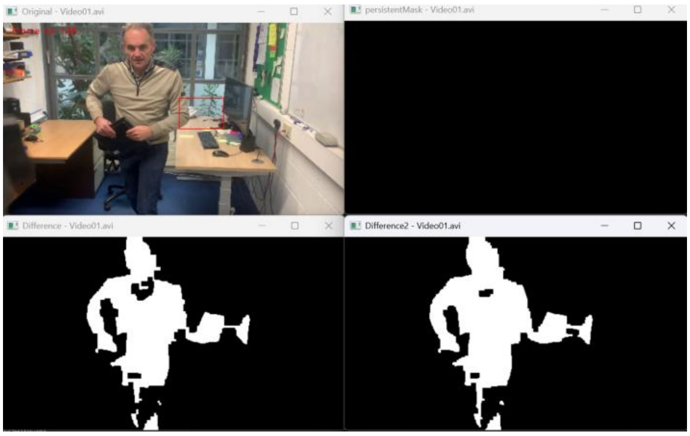

Fig 3 Dual GMM model detecting transient motion.

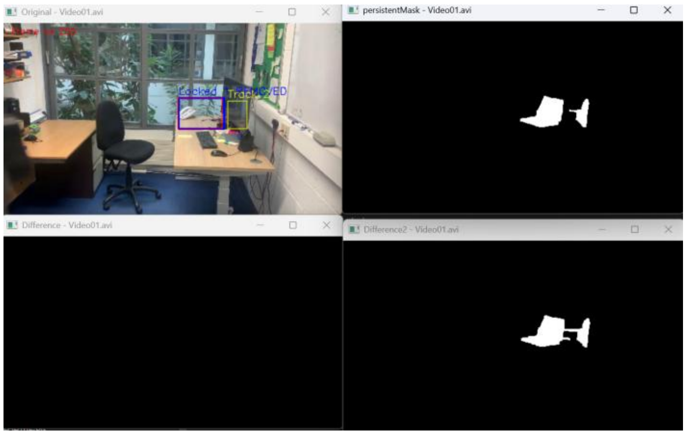

Fig 4 Dual GMM model detecting removed object.

Figures 3 and 4 depict the process. In figure 3 the fast GMM (Difference) and the slow GMM (Difference2) are seen capturing the transient and removed objects in their masks. In figure 4 the removed object has been absorbed into the fast model, but persists in the slow model, causing the object to be added into the persistent mask, producing a stable detection.

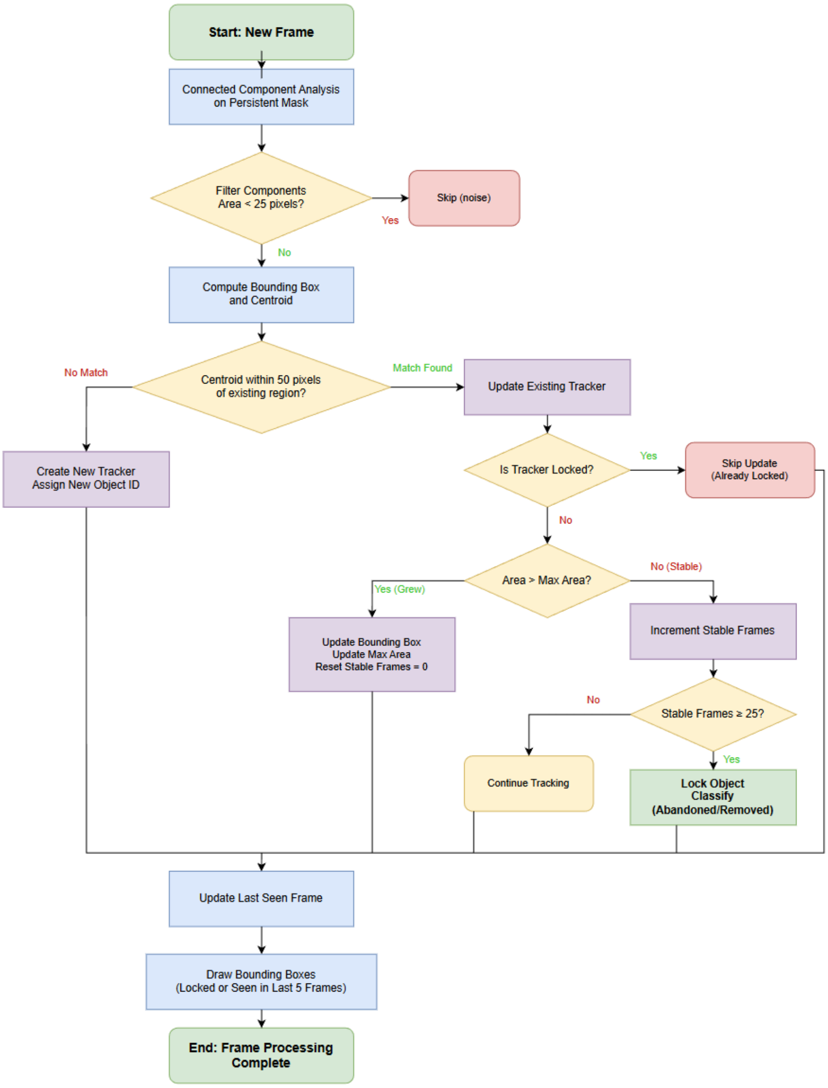

Fig 5 Object tracking algorithm.

The system implements a spatial tracking system to monitor detected regions across frames and identify stable objects. Connected component analysis is performed on the persistent mask, grouping contiguous pixels into candidate objects. Regions with area under 25 pixels are assumed to be noise and are ignored. Bounding boxes are computed for all valid regions, as well as their centroid. The regions for this frame are attempted to be matched with existing regions (from previous frames), if the centroid of a region is within 50 pixels of an existing region, they are assumed to be the same region, and the existing region is updated. If the region does not change (i.e. is stable) for 25 frames it is 'locked' and considered a detected object. If no match is found, a new region is added to the tracker. This process allows objects to stabilise before being detected and classified.

Description of Object Classification:

Classification is based on edge correlation, exploiting the fact that structural boundaries of an abandoned object appear in the current frame but not yet in the slowly updating background. Conversely, when an object is removed, its edges persist in the background but not in the current frame.

For each detected object, the region of interest (ROI) is extracted from: the current frame, the slow-GMM background, the persistent mask. The current video frame and background frames are converted to greyscale (persistent mask is a binary image). A Gaussian blur is passed over the three frames to remove noise and encourage continuous edges – a 5x5 kernel for the persistent mask, and a 3x3 kernel for the current video and background frames. Canny edge detection is applied on all three frames with a suppressed threshold of 50, and a strong threshold of 150. These edge images are dilated with a 2x2 kernel to account for small spatial misalignments in edges between images.

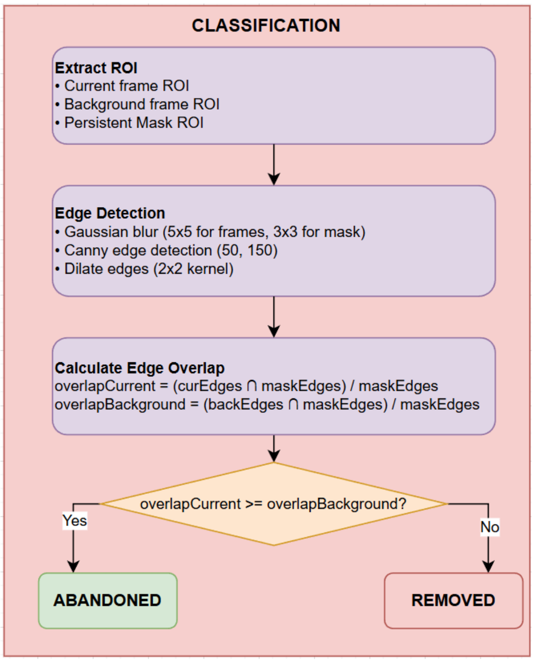

Fig 6 Object classification algorithm.

Fig 7 Persistent mask, current frame, background edges for a removed object in Video 1.

Figure 7 shows the resulting images after processing for a removed object in Video 1. The (left to right) persistent mask edges, the current frame edges, the background model edges. As the object is removed in this case, the mask better matches the background than the current frame, and thus is correctly classified.

The edge images are used to calculate two ratios which underpin the decision:

$$currRatio = (currEdgePixels \cap maskEdgePixels) / maskEdgePixels$$
$$backRatio = (backEdgePixels \cap maskEdgePixels) / maskEdgePixels$$

If the current ratio is greater than or equal to the background ratio, then the object is classified as abandoned. The inverse is true for removed objects.

Results:

Green boxes in the video space represent abandoned objects in the ground truth, and red boxes represent removed objects in the ground truth. Blue, labelled boxes represent detections by the system. Detected objects are matched to ground-truth objects for a minimum intersection over union (IoU) value of at least 0.3. The IoU and delay between ground truth and detections are tracked for each object. Classification accuracy is calculated based on successfully matched detections.

Successful detections:

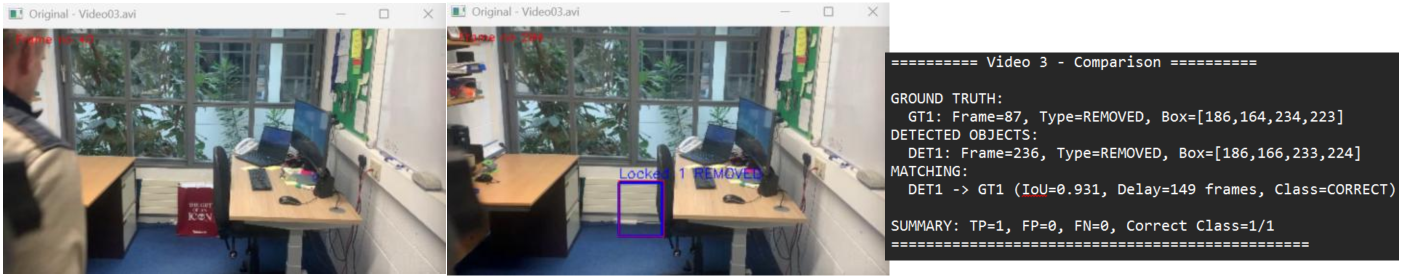

Fig 8 Successful removed detection.

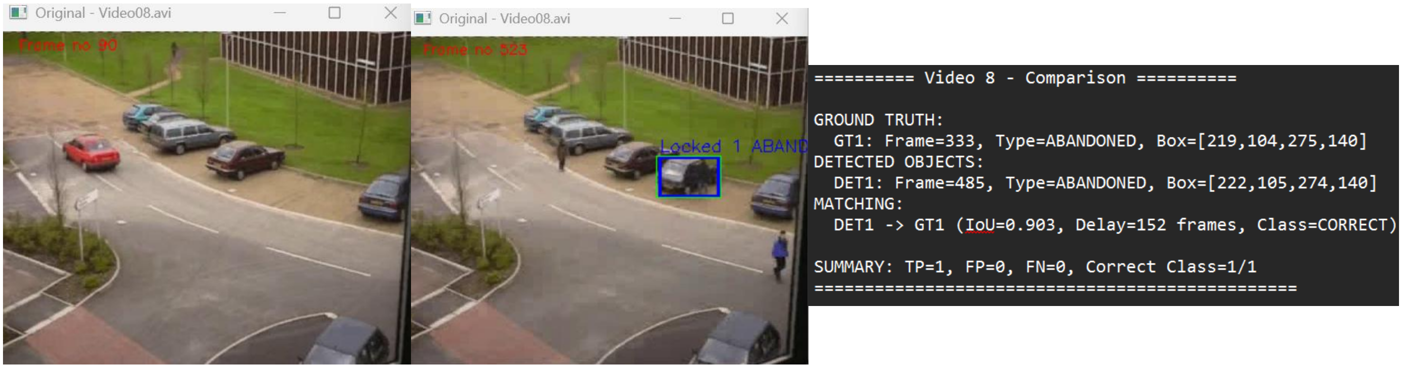

Fig 9 Successful abandoned detection.

Figure 8 and figure 9 depict the successful detection of a removed object and the successful detection of an abandoned object. Detection accuracy is high with an IoU exceeding 90% in both cases. The average delay in detection for these cases was 10.75 seconds for a 14fps video (as used in the development). This represents the best performance of the system, demonstrating strong detections and classification, dealing minor lighting changes and many transient objects in the scene.

Failure Cases:

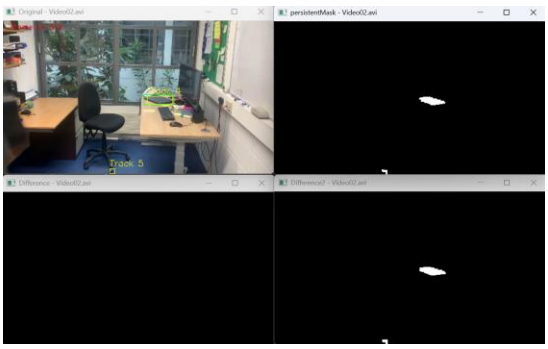

Fig 10 Failure to detect abandoned object.

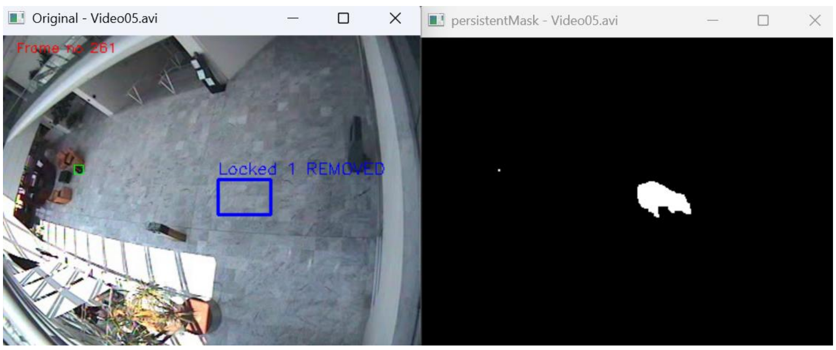

Fig 11 False positive, and false negative.

Figure 10 shows a false negative, in which an abandoned object is missed because the video ends before the object can stabilise and lock. This failure arises from dataset limitations rather than the model design: a continuous video stream would have enabled detection once the object remained static.

Figure 11 illustrates more genuine breakdowns. The false positive of the removed object occurred when a moving person was in frame when the video started. The slow-GMM does not adapt to this quickly enough and an object is detected. Experimenting with a faster learning rate at the beginning of a video solved this problem, however introduced other artifacts on the rest of the video set for an overall negative impact and was not included in the final design. The small dot on the left of the persistent mask represents a false negative, in which the object is too small to be detected. This is a clear limitation of the system, illustrating a trade-off between sensitivity and noise-suppression. This is a clear limit set in the parameters of the model in which objects under 25-pixel area are ignored.

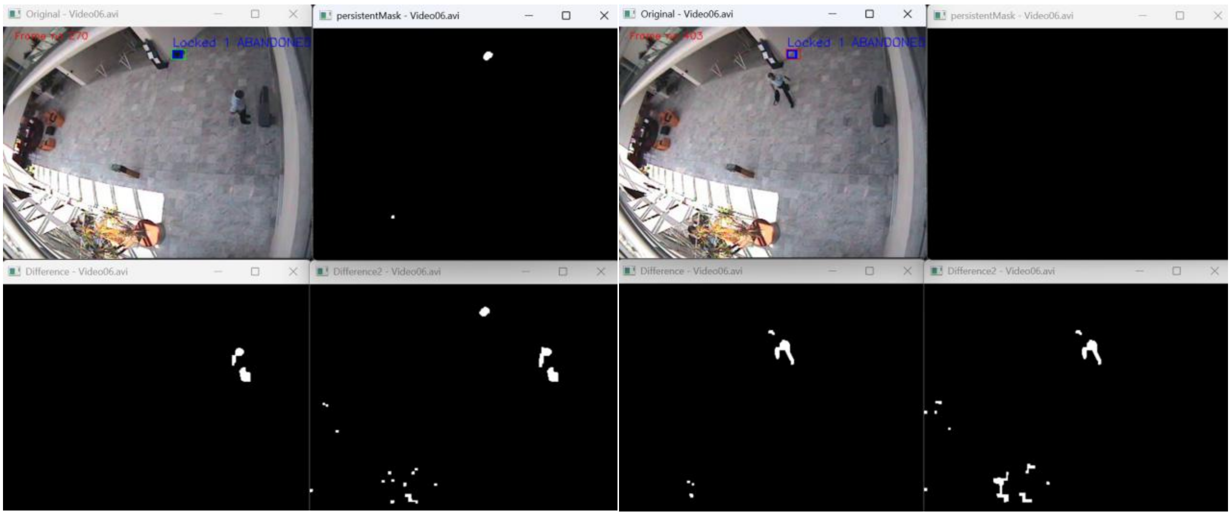

Fig 12 Failure to detect sequence on same object.

In Video 6, an object is abandoned and later removed. They system correctly identified the abandoned state but failed to recognise the removal. The GMMs do not incorporate the object quick enough to be able to detect the subsequent removal. This leads to a false negative on the removal. Experimenting with temporarily increasing the learning rate after a detection to incorporate those changes into the background, yielded some success. Unfortunately, OpenCV does not allow for a mask of the background model to operate at different learning rates, so this increased learning, while allowing the subsequent removal to be detected in the persistent mask, introduced lighting artefacts into the rest of the scene that negatively affected overall performance.

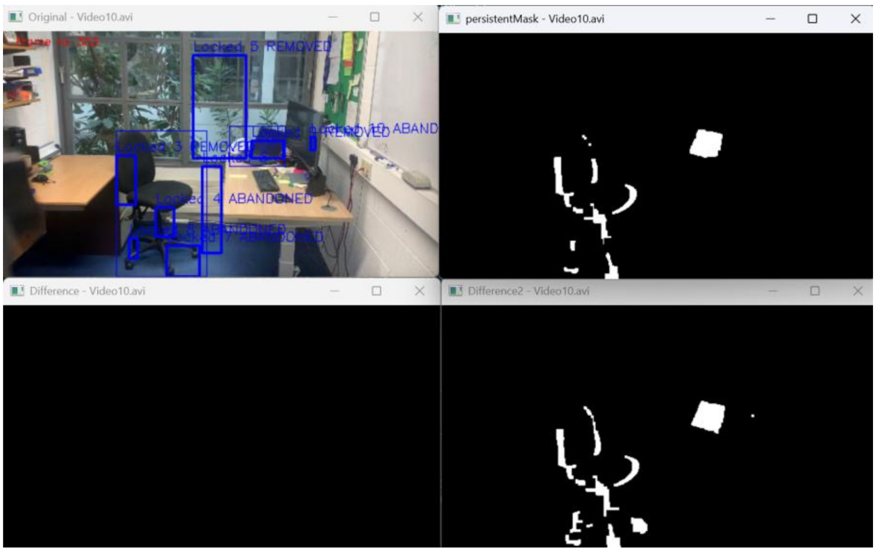

Fig 13 System breakdown for moved objects within a scene.

Video 10 introduced moved (but not removed or abandoned) objects. Additional descriptors such as texture and chamfer distance were tested, but they reduced overall classification accuracy. Consequently, the system was restricted to abandoned and removed object detection.

Performance evaluation:

```
##################################################
##        AGGREGATE RESULTS - Videos 1 - 9      ##
##################################################

DETECTION PERFORMANCE:
  Total Ground Truth Objects: 12
  True Positives:  7
  False Positives: 4
  False Negatives: 5

OVERALL METRICS:
  Precision: 0.636
  Recall:    0.583
  F1-Score:  0.609
  Classification Accuracy: 1.000 (7/7 correct)
  Average Detection Delay: 119.00 frames
  Average Spatial Overlap (IoU): 0.694

##################################################
```

Fig 14 Aggregate system results (videos 1-9).

```
##################################################
##        AGGREGATE RESULTS - ALL VIDEOS        ##
##################################################

DETECTION PERFORMANCE:
  Total Ground Truth Objects: 14
  True Positives:  7
  False Positives: 11
  False Negatives: 7

OVERALL METRICS:
  Precision: 0.389
  Recall:    0.500
  F1-Score:  0.438
  Classification Accuracy: 1.000 (7/7 correct)
  Average Detection Delay: 119.00 frames
  Average Spatial Overlap (IoU): 0.694

##################################################
```

Fig 15 Aggregate system results (all videos).

The evaluation of the system focused on detection performance (correct localisation and identified as existing) and classification performance (correctly categorised as abandoned or removed).

Detection Performance Metrics:

- Total Ground Truth Objects: The number of objects annotated in the ground truth data for all evaluated videos.
- True Positives (TP): The number of ground truth objects that were successfully detected and spatially matched to a system output.
- False Positives (FP): The number of detections produced by the system that did not correspond to any real object in the ground truth.
- False Negatives (FN): The number of ground truth objects that were missed by the system.

The system correctly classified all correctly identified objects with 100% accuracy. The moderate precision and recall indicate that many objects were missed or incorrectly flagged. This is exacerbated when considering video 10, as the system does not deal with objects that are neither abandoned nor removed, considerably impacting the overall performance. The average IoU in correctly detected objects was 69%, with an average detection delay of 8.5 seconds. The imperfect IoU can mostly be attributed to challenging lighting conditions, where objects were in dark areas of the scene, or matched the background quite well in terms of colour and brightness. 8.5 seconds is a reasonable detection time for classifying an object as abandoned or removed, however this could be semantically interpreted differently based on the use-case/context.

Appropriateness, Robustness and Improvements:

The use of a Gaussian Mixture Model (GMM) for background subtraction is an appropriate and well-established approach for detecting stationary foreground objects in relatively static surveillance scenes [3]. However, they are not inherently robust to sudden illumination changes, dynamic backgrounds, or short video durations, which limits their suitability in less-controlled environments. Also, their reliance on pixel-level statistics makes them less effective for complex scenes with overlapping objects, occlusions, and significant texture variation.

Overall, the system demonstrates strong classification reliability once detections occur but limited spatial and temporal robustness. The model is unlikely to generalise well beyond the development dataset. Parameters such as kernel size, learning rate, and temporal thresholds are tightly coupled to the resolution and framerate of the seen videos. Re-optimisation would be required for different datasets. The model is sensitive to video length, as well as lighting variation.

The robustness of the system could be greatly improved by allowing for variable learning rates in the GMM models. Specifically, at the beginning of videos, and when detections occur. There are challenges to doing this in a generalisable way, but initial exploration yielded positive results. This could improve most major problems, with video initialisation, moving objects at video inception, and missed detections on previously abandoned objects. Additional techniques like feature matching comparing persistent mask features to the background models could be used to infer that an object has moved in a scene and falls into the 'neither abandoned nor removed' class. This would greatly improve detection and classification robustness.

Conclusion:

The dual-speed Gaussian Mixture Model (GMM) system developed for abandoned and removed object detection successfully demonstrates the viability of adaptive background subtraction as a foundation for scene understanding in static surveillance environments. By coupling fast and slow GMMs, the system effectively differentiates transient motion from sustained scene changes, allowing it to isolate stationary or missing objects with temporal persistence. The classification mechanism, based on edge correlation between the current frame, background, and persistent mask, provides reliable distinction between abandoned and removed objects once detections occur.

Evaluation showed that the model achieved strong classification accuracy (100% on valid detections), and moderate localisation performance, with an average intersection-over-union (IoU) of 69%, and a mean detection delay of 8.5 seconds. The model fails to address and classify objects that are moved within a scene but not abandoned or removed. Several limitations constrain the model's generalisation capability, with it being highly dependent on carefully tuned parameters that are specific to the type of video seen in development.

Future work should focus on dynamic adaptation mechanisms, such as spatially and temporally variable learning rates for the GMM models to handle scene initialisation and objects that undergo abandonment and removal. Extending the framework to handle feature-matching and more robust object-tracking would enhance the robustness of detections and classifications.

References:

[1] Russel NS, Selvaraj A. Ownership of abandoned object detection by integrating carried object recognition and context sensing. The Visual Computer. 2023 Oct 2;40(6):4401–26.

[2] Liu D, Wang H, Zhang X, Zhang X, Liu Y. Fast highway abandoned object detection via block-based multi-group foreground extraction. Scientific Reports [Internet]. 2025 Oct 27 [cited 2025 Nov 12];15(1). Available from: https://www.nature.com/articles/s41598-025-20331-z

[3] Tian, Y, Feris, R, Liu, H, Hampapur, A, Sun, MT. "Robust Detection of Abandoned and Removed Objects in Complex Surveillance Videos". IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 2011; 41(5):565-576.