COM672 – COMPUTER VISION

Coursework Two

Brógán McShane

Task One – Frame Differencing

Upon the first attempt to implement a frame differencing algorithm through the use of spatial filters, it was observed that immediate sequential frames (i.e. frame 6 and frame 7) had little differences and would not effectively detect motion within the video. For this reason, a question had to be asked: could the motion detection in a video be more accurate if the frames were further spaced? i.e. would frame 4 show a more accurate difference when compared to frames 7, 8 and 9 for moving objects? For this reason, different ranges were considered when comparing frames which would be used to differentiate from each other. The most notable steps in ranges were: 25 (the video's frame rate), 12, 7 and 5.

In the experiment, comparing frames 1-24 with the initial frame proved to be effective, however, when frames 25-49 were compared with the 25th frame, the motion tracking algorithm would struggle with accurate detection for the frames that were furthest from the frame which will be used for differentiation – deeming that this comparison range is too large. A similar result proved to be the outcome when frames 1-11 were compared with the initial frame, and 12-23 with the 12th frame, concluding that a smaller range would be necessary in order to increase the accuracy of the motion detection algorithm.

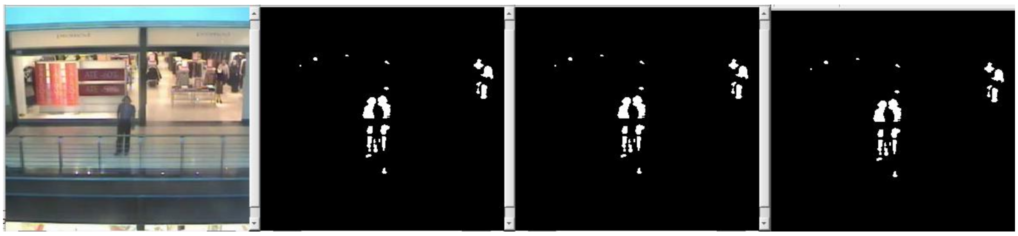
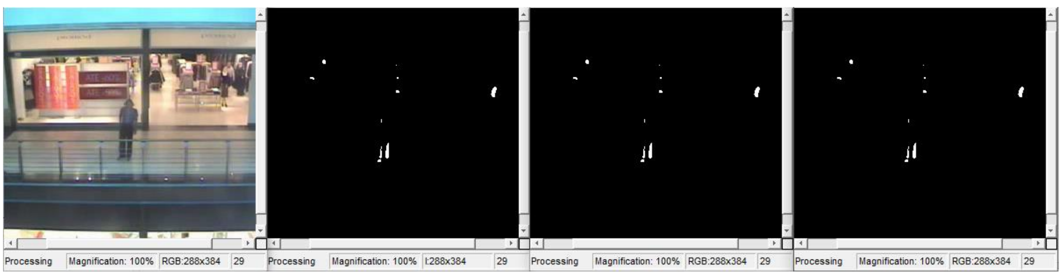


Figure 1: Motion detection algorithm applied on frame 44 when differentiated with frame 25 through average filtering (filter size: [2,4]).

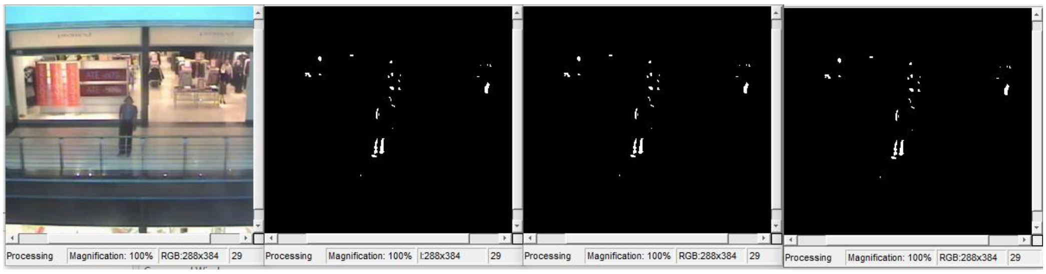
The next ranges to be considered in the experiment were steps of 7 and steps of 5. However, first the author needed to address the method of determining which frame would be used as a differentiator from the first. As previously described, the initial method would compare all frames within a valid range with the same one differentiating frame (i.e. in the range of 7, all frames from 7-13 would be compared with the 7th frame), meaning that the first frames within said range would show little to no movement if any took place. To overcome this, an alternative method would be put in place, which would determine the differentiating frame by considering the current frame's position index in the video and comparing it against the frame n indices before it (i.e. for ranges of 7, frame 12 would be compared against frame 5, frame 13 against frame 6 etc.)

After experimenting with the new method of comparing frames, it was concluded that the frame differencing algorithm would use steps of 7 when comparing the current frame against the differing frame, as it would allow enough time to detect subtle movements as well as more drastic changes in the two frames.

Next, different spatial filters of different sizes would be used as part of an experiment to determine: which spatial filter is the most effective and which size is most effective per filter. To begin, a mean filter of size 7 by 7 would be applied during the process of differentiating frames and would see success in more drastic changes in movement, however it would be inaccurate to the more subtle changes in the video. Next, a mean filter of size 4 by 4 would be applied during the process of differentiating frames, and while effective in representing both subtle and more drastic changes, it seemed that this size of filter would too often represent small changes as large movements when compared to actual moving objects, as well as overrepresenting the areas of change from the differentiating frame.



Following this, different mean filters would be experimented with to determine the most effective size to apply for the frame differencing algorithm. The first mean filter's size was 3 by 3, and seen similar results to the 4 by 4 filter, in that it would overrepresent small details of change between the two frames. Next a filter size of 2 by 2 would be evaluated, and although this filter would represent minute details in the two frames, it would also represent moving objects with more detail. See the below example: although changes in the surrounding "background" (i.e. shared similarities between the two frames) are being represented by the filter – which may not be desired, there is a clear, more detailed outline of the areas of objects which have moved, including the closest person's head, shoulder, arm and legs and the furthest person's body, and more identifiably, their leg.

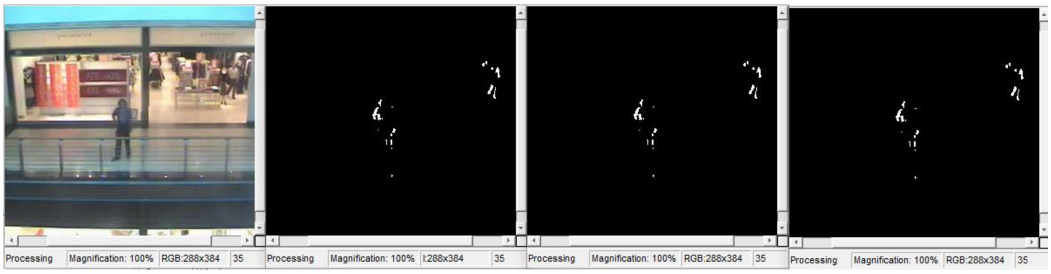


Task Two – Frame Differencing Through Temporal Filtering

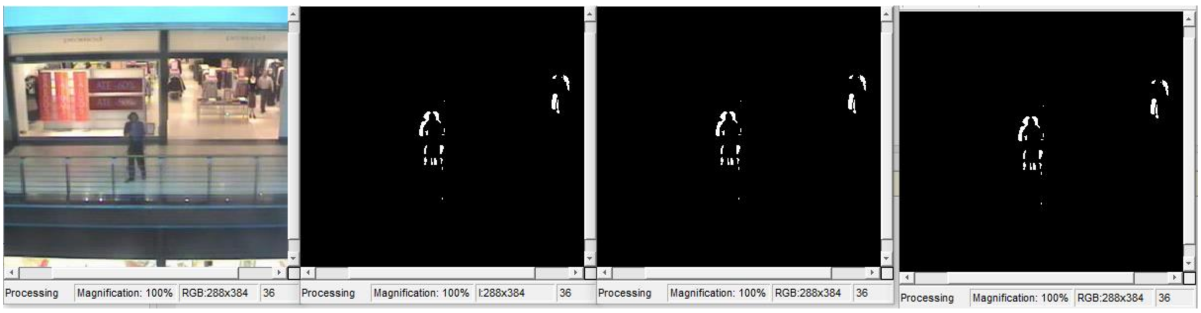
Next, temporal filtering would be investigated as a method to deploy for frame differencing. Initially, the first four frames would make up the temporal filter against which the current frame would be compared to. After every frame has been iterated for in the entire video, the frames that make up the temporal filter will reflect this (see below for visual representation of this algorithm):

Current frame index:	Temporal filter will contain frames at indexes
5	1,2,3,4
6	5,1,2,3,4
7	5,6,2,3,4
8	5,6,7,3,4

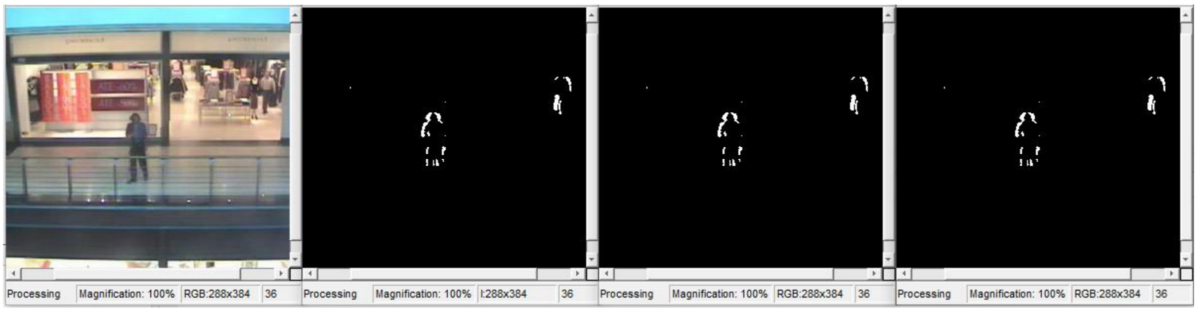
As mentioned, the initial implementation of this algorithm would contain an average image of the previous four frames, the resulting comparison did show where the current frame differentiated from the temporal filter, however, it was determined that although this filter was accurate, improvements could be made to represent the moving object in reference to the temporal filter.



Next, the temporal filter would consist of only frames that are at least three frames away from the current frame. This method of improved on the previous temporal differencing implementation, as there is a more accurate representation of where both the closest and furthest person's bodies have moved in comparison to the temporal filter.



As a means of experimentation, temporal filters containing both three and ten frames would be analysed to determine the optimal size of the filter for frame differencing. Initially, the filter which contains three frames did show similar results to the temporal consisting of five, however it was more sensitive to background noise as seen below on the same frame in figure x.



Furthermore, when the temporal filter containing ten frames was evaluated it was determined that this range of frames that construct the filter was too large when compared against the current frame, as although it was effective in detecting motion, the accuracy of the detection was lower when compared to the previous two implementations.

Task Three – Exploring and overcoming noise

Upon initial investigation, the frame differencing motion detection algorithm struggled in distinguishing motion against salt and pepper noise as well as gaussian noise when both were individually applied to the current video frame and each frame that makes up the temporal filter.

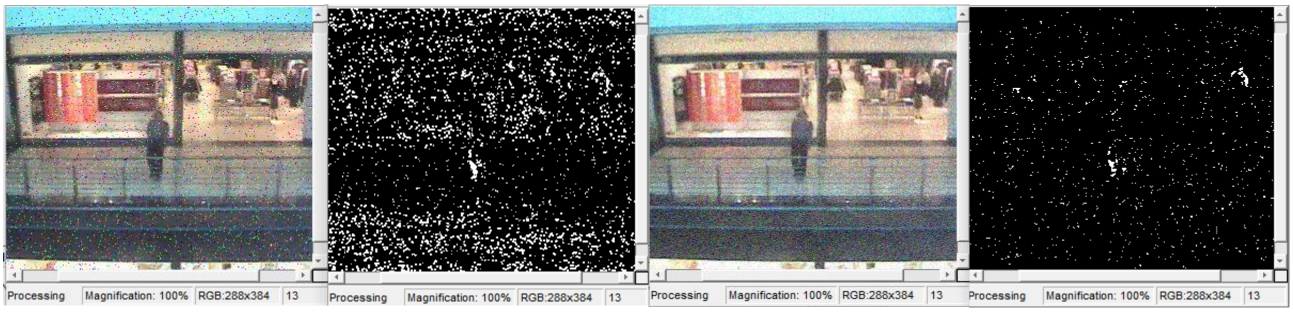


Figure x (from left to right): Frame with Salt and Pepper noise (0.05), Frame differencing on current frame against salt and pepper frame, Frame with Gaussian noise, Frame differencing on current frame against Gaussian frame

To overcome differentiating noise from object motion in the current frame against the temporal filter, experiments would take place that involved the removal of noise on the current frame as well as each frame that constructs the temporal filter. The first problem to be overcome was added salt and pepper noise: the initial approach to attain effective noise removal of this kind involved the use of a median filter on each layer of every frame involved in the frame differencing, which seen a level of success and accuracy in detecting object motion, however there would still be small areas of the frame that the algorithm would consider to be motion when in reality, the “motion” was salt and pepper noise. To improve on the motion detection’s algorithm on distinguishing the motion from the salt and pepper noise, different thresholding levels and different kernel sizes were analysed, the most effective filter was sized at 3 by 3 and the most effective thresholding level was 48 (previously set to 45).

Figure x (from left to right): Current frame with increased Salt and Pepper noise (0.1), Current frame cleansed of salt and pepper noise, Frame difference between cleansed current frame and cleansed temporal filter

Next, Gaussian noise needed to be addressed in all frames that make up the temporal filter and the current frame. Two methods of removing Gaussian noise were investigated: the first being through the use of the wiener filter, and the latter being through the use of average filtering. Using a Wiener filter on each layer of the RGB frame involve the evaluation of different sizes of kernels, with the most effective being a 4 by 4 kernel. The result of removing Gaussian noise through the Wiener filter proved to increase the frame differencing algorithm's performance drastically, and surprisingly included precise motion detection for subtle movements of objects. Like the Wiener filter, the use of Averaging filtering also removed the Gaussian noise to a proficient degree, and additionally displayed high accuracy when motion detection was analysed.

Figure x (from left to right): Current frame with Gaussian noise (0.1), Current frame cleansed of Gaussian noise using the wiener filter, Current cleansed frame differed from the cleansed temporal filter.

Figure x (from left to right): Current frame with Gaussian noise (0.1), Current frame cleansed of Gaussian noise using the wiener filter, Current cleansed frame differed from the cleansed temporal filter.

Task Four – Exploring and overcoming smaller videos

By looking at how the motion detection algorithm performs on two different sizes of the same frame, it is clear that the algorithm struggles with accurately representing movement on smaller videos.

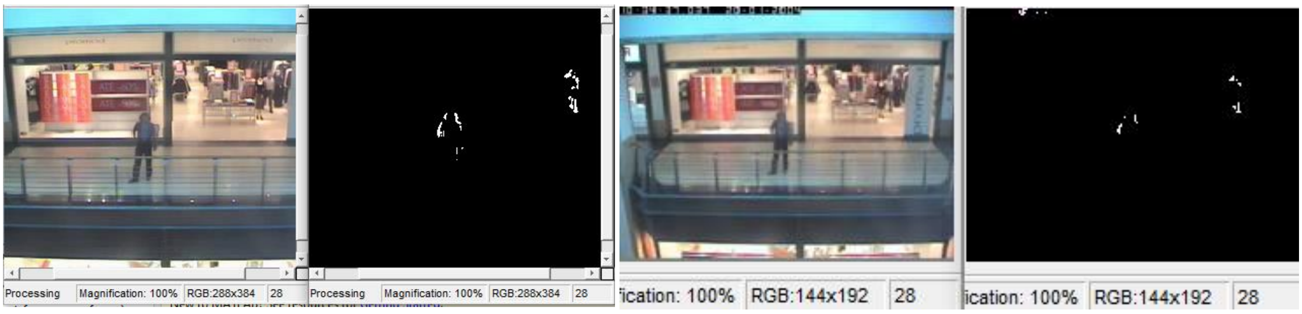


Figure x (from left to right): Full-Scale Frame 28 (size: 288x384), Frame 28 differed from temporal filter, Frame 28 resized to 144x192 (half of its original size), Resized Frame 28 differed from temporal filter.

As a means of experimentation, smaller spatial filters would be used as an investigation into if the motion detection algorithm's accuracy on smaller sized frames – however, this would be short-lived as the current spatial filter is only a 2 by 2 kernel. Upon initial analysis, the 2 by 1 kernel in addition to the 1 by 2 kernel both displayed decreased accuracy when compared to the original kernel's accuracy when detecting object motion. For this reason, alternative methods need to be investigated to improve the motion detector's accuracy when analysing smaller video frames.

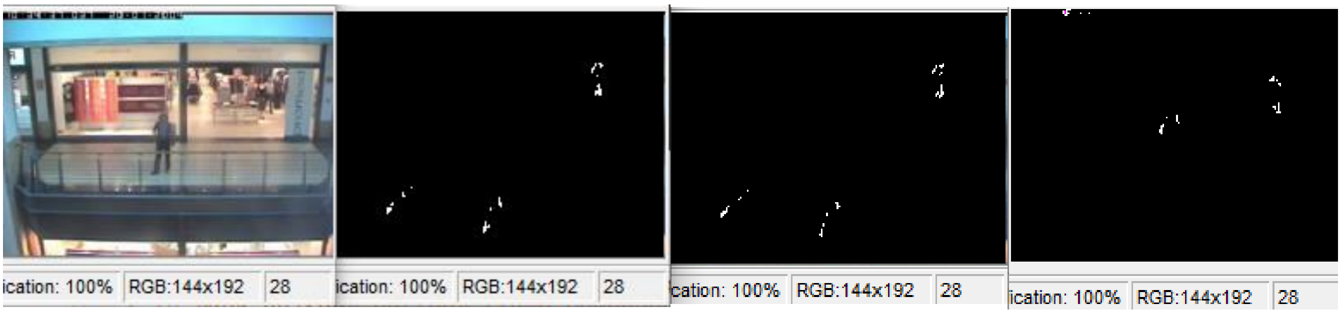
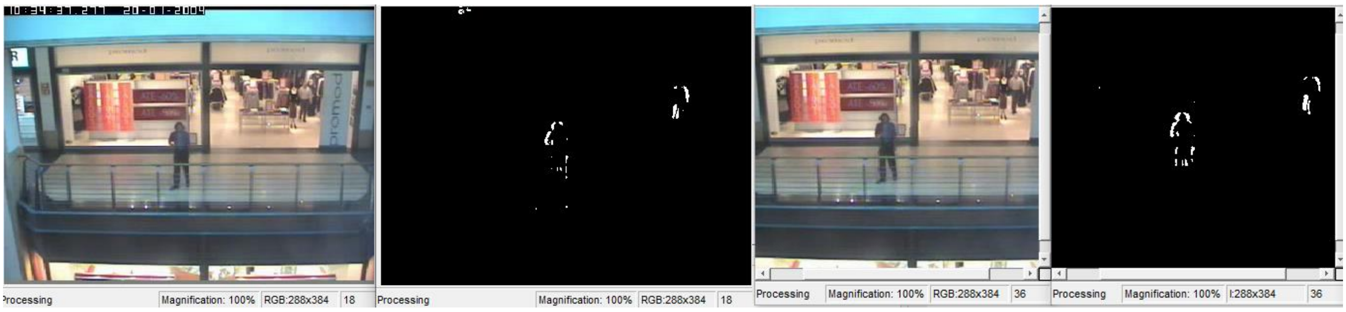


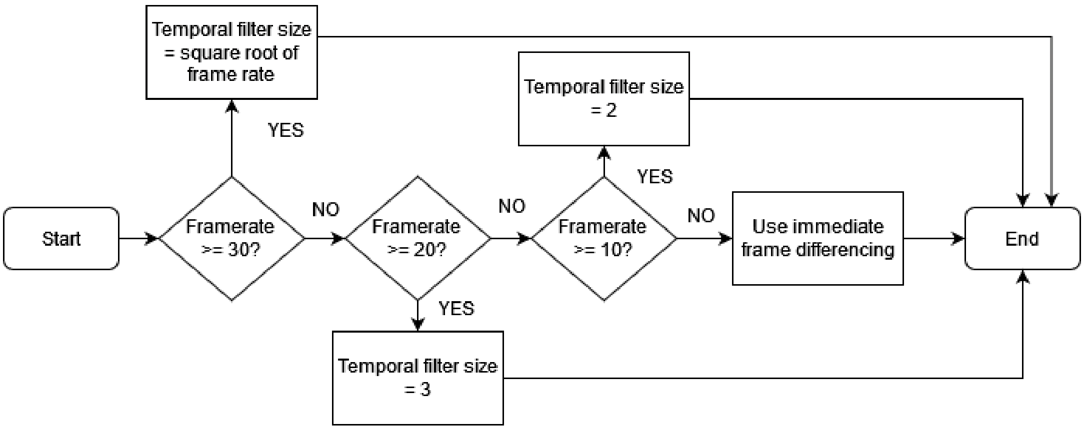
Figure x (from left to right): Frame 28 (half of original size), Frame 28 differed from temporal filter with a 2x1 average filter kernel, Frame 28 differed from temporal filter with a 1x2 average filter kernel, Frame 28 differed using original 2x2 average filter kernel

Task Five

When the original frame rate of the video is split in half (i.e. from 30 to 15), and frames are compared against three of the previous frames which make up a temporal filter, the accuracy of the motion detection algorithm is slightly less accurate in capsulating the true movement of the object being tracked. As the value of frame rate shrinks, the motion tracker has access to less detail illustrated by a higher frame per second value, meaning more subtle movements won't be as accurately detected and will appear as more drastic changes in the motion detector, this decrease in accuracy will continue to decline as the rate between frames grows smaller.



For this reason, an algorithm will need to be developed to determine whether to use temporal filtering (as well as determining the number of frames which make up the filter) or frame differencing, illustrated below.



This updated motion detection algorithm was the most optimal approach in determining the appropriate frames (either in the form of temporal differencing or frame differencing) which would be used to show the difference between frames. This dynamic approach to overcoming videos with lower frame rates proved more useful than the previous demonstration of the motion detection algorithm as although there would still be less accuracy in showing the true motion of an object, appropriate frames would be considered in correlation to the frame rate in order to achieve as high an accuracy as possible.

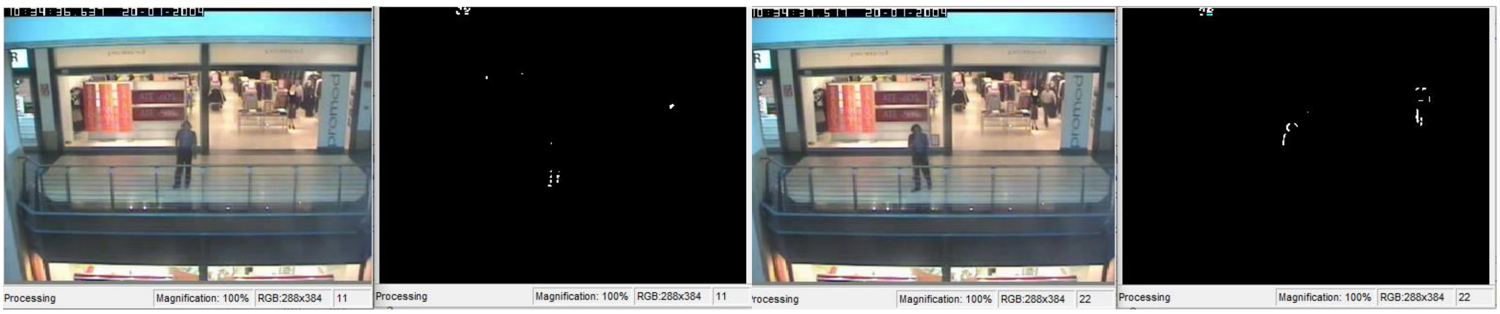


Figure x: Illustration of motion detection algorithm's capability of representing small motion differences between the current frame and the temporal filter

Task Six – Background Subtraction

As Video2.mp4 contains motion throughout its entirety, the current task requires the development of a method which will determine which frame(s) to consider to be the background. Initially, the "background" to be differentiated against would be a mean image of the first and final frames in the video, however this experiment was unsuccessful in determining an appropriate "background" as the current frame would detect motion where there the current frame didn't feature any movement.

The next approach to determining an appropriate background image was by investigating the mean frame of every frame in the video. By writing every frame to a jpg file, constructing a matrix containing all frame jpgs and dividing that matrix by the number of frames in the video, a suitable background was created and didn't feature the moving objects. This background image performed exceptionally well in comparison to the previous experiment, and even detected the shadows of the moving objects on the ground.

Figure x: The mean frame of all video frames in Video2.mp4 to be used as the background

Figure x - Background subtraction (from left to right): Frame 187, Frame 187's movement determined by background subtraction, Frame 65, Frame 65's movement determined by background subtraction.

Task Seven – Combining Background Subtraction and Frame Differencing

In order to identify an occurrence of an object being brought into the scene and left unattended, the background (i.e. the first frame which contains no movement) and the previous frame would construct a temporal filter against which the current frame would be differentiated from. Although this approach was successful in detecting moving objects, and unattended objects (see figure below), the algorithm struggled with detecting when the second person picked up an object and carried it off-scene and could be improved upon to more accurately represent how an object has moved during the duration of the video.

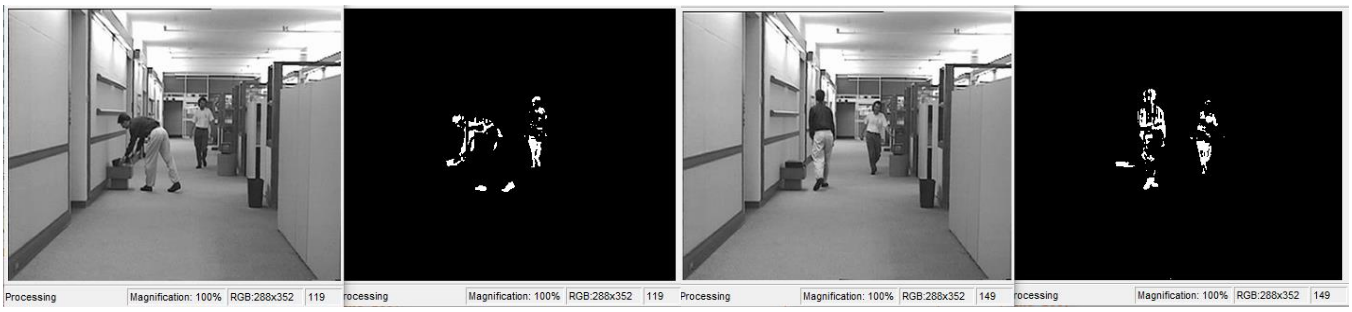


Figure x - Background subtraction (from left to right): Person is abandoning object (Frame 119), Current frame compared against temporal filter, Person moving away from abandoned object and another person coming on-scene, Detection of abandoned object and two moving people

In an attempt to overcome the algorithm's weakness of detecting objects being taken off-scene, different temporal filters would be investigated. The first temporal filter would consist of the mean of the previous two frames (i.e. 3 and 4 for a current frame 5), then, the background image (frame 1) would be added to the temporal filter, which would then be divided by two – this process was followed in order to give the background frame more weight when differentiating between the current frame and the temporal filter. This new temporal filter did improve upon the accuracy of the previous filter in representing where an object has moved to and where an object had been abandoned as well as detecting when an object was picked up and proceeded to be carried off-scene.

Figure x (from left to right): Person is abandoning object (Frame 119), Current frame compared against temporal filter, Person moving away from abandoned object and another person coming on-scene(Frame 149), Detection of abandoned object and two moving people

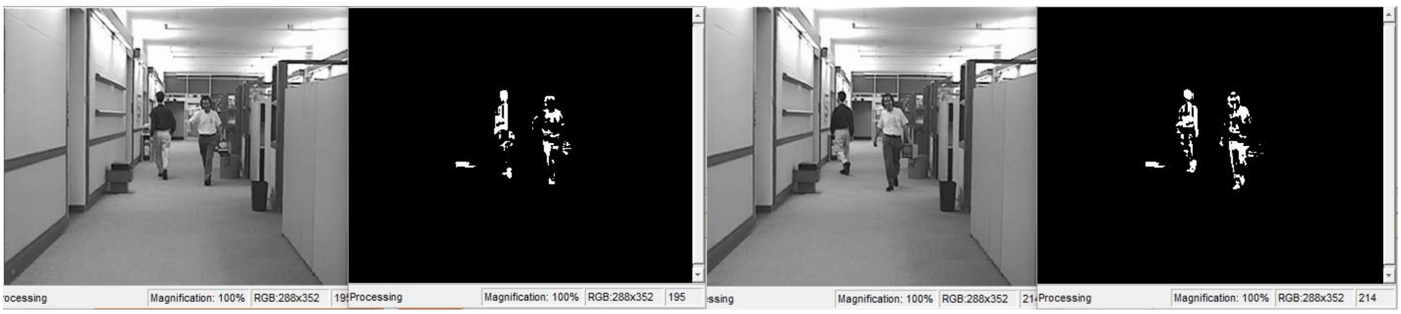


Figure x (from left to right): Person (furthest to the right) carrying object, current frame compared to temporal filter, person still carrying object, detection of movement and abandoned object.