# LAB 3 - OBJECT DETECTION AND CLASSIFICATION

## APPLIED VIDEO SEQUENCES ANALYSIS

Zsófia Sára Budai and Juan Manuel Peña Trapero
{budai9902, juanmaptcg}@gmail.com

## I. INTRODUCTION

Kalman Filter is a mathematical algorithm that is used to estimate the state of a dynamic system in the presence of uncertain and noisy measurements. It is widely used in various fields, including robotics, control systems, and computer vision. In this lab, we will be exploring the performance of Kalman Filter for object tracking in both Toy videos and real videos.

## II. EXPERIMENTAL METHOD

The execution of the experiments for tasks 2 and 3 have been carried out using an iterative method in which the control parameters of the kalman filter have been modified at different levels to check which combination yields the best visual results for tracking.

Taking into account that the objective of the Kalman filter is to track the motion of the object, we are going to evaluate how close the measurement and prediction are. It is necessary to take into account that the measurements employed by the Kalman filter are not perfect and, in some cases, contain a lot of noise. For this reason the evaluation will also take into account how the kalman reacts when the measurements are very noisy. This is because the extraction of blobs is not perfect and in some cases introduces many erroneous measurements. For this reason, in some of the sequences the blob detection module will be slightly adjusted so that the measurements are not so erratic.

Before diving further into the experimental setup it is important to define and understand what is going to be modified in the Kalman filter. The set of modifications are used to adjust how much to trust the predictions and measurements made by the filter. Here's a detailed explanation of each of the modified parameters:

1) MEAS_NOISE_COV: This variable represents the covariance matrix of the measurement noise. It is a measure of how noisy the measurements are. A small value for MEAS_NOISE_COV means that we have a high degree of confidence in the measurements and rely less on the predictions to update our state estimate. On the other hand, a large value for MEAS_NOISE_COV means that we have a low degree of confidence in the measurements and rely more on the predictions to update our state estimate.
2) ERROR_COV_POST: This variable represents the initial error covariance matrix of the system. The error covariance matrix is used to quantify the uncertainty of the state estimate. A high value for ERROR_COV_POST means that we don't trust our initial prediction and rely more on the measurements to update our state estimate. On the other hand, a low value for ERROR_COV_POST means that we have a high degree of confidence in our initial prediction and rely less on the measurements to update our state estimate.
3) The variables MEAS_NOISE_COV, NOISECOV_VEL, and NOISECOV_ACC are all used to model the uncertainty in the system being tracked. A small value for each of these variables means that we have a high degree of confidence in the corresponding aspect of the system (measurement noise, velocity, or acceleration) and expect it to change very little between frames. On the other hand, a large value for each of these variables means that we have a low degree of confidence in the corresponding aspect of the system and expect it to change more between frames.

## III. TASK 1: MERGING OUR CODE

The structure of the code has been modified taking into account the feedback received in the Kalman filter part. A new class called Kalman Filter has been created which encapsulates all the part that was previously in the main function. Now the code is simpler and more readable as it is only necessary to instantiate the class to initialise the kalman class internally with the chosen parameters. On the other hand, the numerical values that were introduced directly in the code have been eliminated and have been modified by constants to make the code more understandable both in the main and in the Kalman class.

## IV. TASK 2: ANALYSIS WITH TOY DATA

### A. Video2

The toy video 2 (111 frames total, ball is visible from 50 to 63) is composed of a single ball running across a floor describing a linear trajectory, the ball shape is starched because the shutter speed of the camera is not fast enough to "freeze" the ball for each frame. This causes the blob extraction algorithm to detect a larger blob thus getting a wrong measurement. In this case the Kalman filter cannot correct easily the measurements because the same segmentation error is experience for all the frames. For this reason, the Kalman parameters can be adjusted to not perform a strong prediction and just rely on the measurements for this sequence. The baseline and final parameters are displaid in

TABLE I

SUMMARY OF MOST RELEVANT TEST OVER SEQUENCE VIDEO2

| Mod | Meas Noise Cov(R) | Error Post Cov(P) | Process Noise Covariance (Q) | | | EOM params | Result |
|---|---|---|---|---|---|---|---|
| | | | *Pos* | *Vel* | *Acc* | | |
| Con Vel | 25 | 1e7 | 25 | 10 | 1 | UnMod | Bad |
| Con Acc | 10 | 10e6 | 200 | 50 | 1 | Mod | Good |

the Table I. As can be seen in the Figure 2 the measurement are completely trusted for this simple case. In addition to the Kalman filter, the EOM has been modified with a larger learning rate and bigger opening in order to avoid throwing measurements on the wall due to noise and ilumination variations.

Fig. 1. Kalman filtered prediction and measured points for sequence video 2 with best parameters (last row of table).

### B. Video3

This video shows a ball rolling into a wall and bouncing out. This video is difficult because of the high variability of the background due to noise. The ball does not have a very high speed and is easily identifiable. On the other hand, the trajectory of the ball is not completely linear but it is composed of two linear trajectories. The deceleration of the ball due to the bouncing on the wall and the friction with the ground seems to indicate that the Constant Acceleration Model is the most suitable.

For this video, the recommended parameters for the EOM were outputting extremely noisy measurements due to the noise and lightning variability produced across the sequence. For this reason, the parameters are altered to avoid extremely confusion on the Kalman filter when analyzing the measured trajectory. The table II shows the best parameters for this video sequence in the last row. It should be noted that the model selected for this sequence is the constant velocity model, although it may seem to require a more complex model. This model was chosen because the constant acceleration model predicted the wrong trajectory when the ball disappeared from the scene. Due to the fact that this scene is quite simple and has correct measurements, the simplest model obtains correct results, leaving aside the first measurements in which there is a greater error, as can be seen in Figure 2.

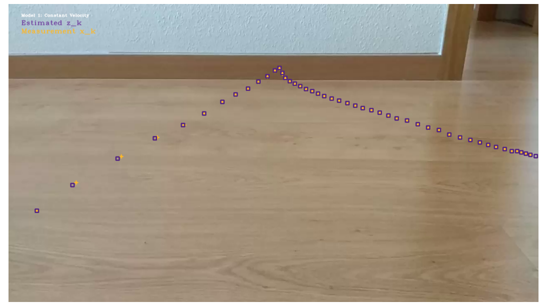

Fig. 2. Kalman filtered prediction and measured points for sequence video 2 with best parameters (last row of table).

TABLE II

SUMMARY OF MOST RELEVANT TEST OVER SEQUENCE VIDEO3

| Mod | Meas Noise Cov(R) | Error Post Cov(P) | Process Noise Covariance (Q) | | | EOM params | Result |
|---|---|---|---|---|---|---|---|
| | | | *Pos* | *Vel* | *Acc* | | |
| Con Acc | 25 | 1e6 | 25 | 10 | 1 | UnMod | Bad: kalman cannot filter enough noisy measuremtns |
| Con Acc | 5 | 1e2 | 50 | 8 | 1 | Mod | Bad: the model predicts that the ball bounces back (Figure) |
| Co Spd | 5 | 1e6 | 50 | 8 | 1 | Mod | Good |

### C. Video5

This sequence captures the ball hitting the floor after starting from a free fall, resulting in approximately nine bounces. The initial descent was rapid, followed by a zig-zag trajectory along the vertical axis as the ball decelerated. Additionally, the video presented variable shadows that appeared as the ball approached the ground.

For this case the EOM is further modified including a larguer value of morphological opening due to the speed and complexity of the movement. It is interested to note how in the Table III the value of the acceleration process covariance noise is increase. This is because, for this sequence, the Kalman filter must estimate and detect a more complex motion, so the best results are obtained with somewhat more extreme parameters than for the other figures. Figure **??** shows how the trajectory described by the Kalman prediction is closer to reality even than the measured points. This is because, although the ball is not occluded at any time, the measurements obtained have a lot of noise due to the complexity of the background subtraction. On the other hand, the high speed means that it is not possible to measure the exact position of the ball as there is a considerable displacement between the capture of one frame and the next.

TABLE III

SUMMARY OF MOST RELEVANT TEST OVER SEQUENCE VIDEO5

| Mod | Meas Noise Cov(R) | Error Post Cov(P) | Process Noise Covariance (Q) | | | EOM params | Result |
|---|---|---|---|---|---|---|---|
| | | | *Pos* | *Vel* | *Acc* | | |
| Con Acc | 25 | 1e7 | 10 | 5 | 3 | Mod | Good |

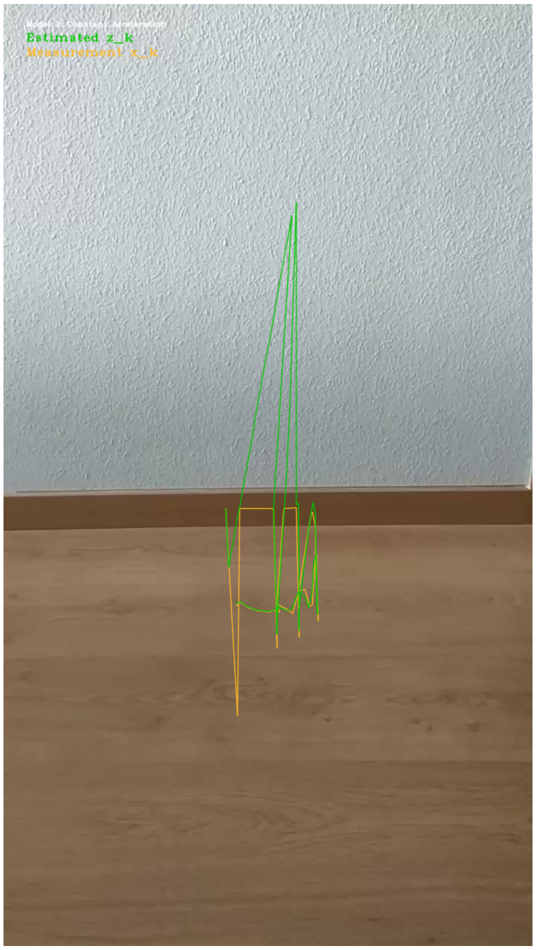


Fig. 3. Kalman filtered prediction and measured points for sequence video 5 with best parameters (last row of table).

### D. Video6

This sequence features the ball moving linearly through a dark area before becoming completely occluded. The ball re-emerged in a brighter area, with linear deceleration and clear detection in the final section of the scene.

This scene is considerably more complex due to the fact that the segmentation is not robust to changes in lighting and due to the reflection of the ball. For this sequence we decided not to modify the measurement extraction module too much in order to focus on the kalman filter parameters. In this sequence there is a trade-off between the importance given to the prediction and the trajectory obtained. This is because, as the ball comes to a complete stop, kalman must adjust the velocity accordingly so that the prediction stops as soon as the ball reaches the wall. In many of the parameters tested (e.g. those in Figure 4) it is found that the predicted ball bounces when occluded and then continues its trajectory even though it actually comes to a complete stop at the end of the sequence without leaving the scene. This is corrected with the parameters reflected in the last row of Table IV. As can be seen in Figure 7, the averages obtained are not accurate in the most illuminated part when leaving the box due to the fact that the tennis ball is much more similar to the background. This effect is compensated by a parameter selection that is sufficiently confident in the prediction. In this sense, it should be noted that the kalman filter should ideally detect that the ball follows a linear decelerated motion and should not depend too much on the averages obtained at the exit of the box (incorrect measurements). This is the trade-off of this sequence, if you rely too much on the averages you need to get the correct measurements in the frames where the ball comes to a complete stop.

It is also worth noting that then MOE block has been modified making the size smaller the minimum size of blobs due to the reflection problems and the smaller blob obtained due to the background similarity on the bright part.

TABLE IV

SUMMARY OF MOST RELEVANT TEST OVER SEQUENCE VIDEO6

| **Mod** | **Meas Noise Cov(R)** | **Error Post Cov(P)** | **Process Noise Covariance (Q)** | | | **EOM params** | **Result** |
|---|---|---|---|---|---|---|---|
| | | | ***Pos*** | ***Vel*** | ***Acc*** | | |
| Con Acc | 25 | 1e7 | 25 | 3 | 5 | UnMod | Bad: trajectory continues going right after stopping |
| Con Acc | 28 | 1e5 | 25 | 3 | 2 | Mod | Good |

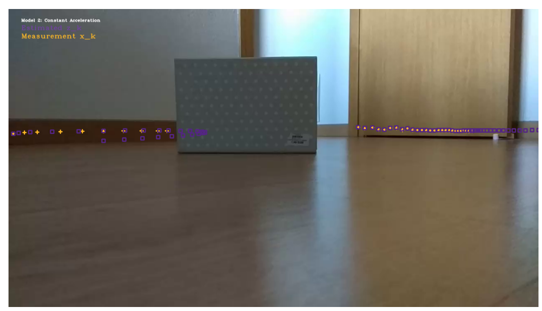


Fig. 4. Kalman outputing wrong predictions points for video 6 due to high acceleration (first row of Table IV).

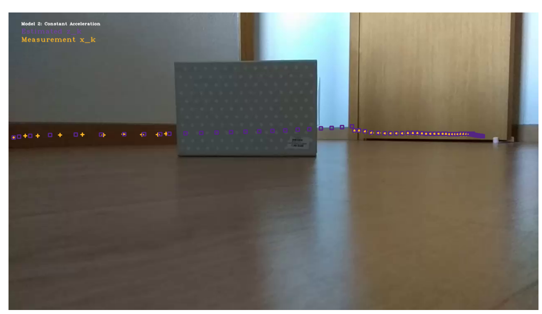


Fig. 5. Kalman predictions points for video 6 with best parameters (last row of Table IV).

## V. TASK 3: ANALYSIS WITH REAL DATA

This section shows the same structure that the previous one but including real video sequences. Each subsection is named after the name of the file that was provided for this lab.

### A. Abandoned Box

This sequence shows a surveillance camera with rather poor image quality (a lot of noise and interlaced video coding synchronisation problems). In the sequence a cyclist appears at a moderate speed and rides through half of the scene in a few seconds to stop at a traffic light.

At first the cyclist is not detected correctly so the MOE block is modified in order to obtain some accurate measurements of the cyclist's position. The history value has been increased substantially, we also included opening-closing-opening and higher connectivity. Consequently, we also increased the threshold. The goal is to obtain a larger and more constant blob that can be easily detected, otherwise the measurements are extremely noisy.

In this scene the MOE block has a difficult task because the bicycle has transparent parts (the background is seen through the wheels) and the background has a high variability due to noise in the image. This results in a large number of blobs being detected which produce a large number of noisy measurements.

This sequence has shown us the importance of knowing how to correctly configure the kalman filter in order to correct measurements that, as in this case, are extremely noisy due to their complexity.

In this case for position, speed and deceleration changes are low and the measurements are noisy. For this reason our intuition was to make the noise covariance large. The problem with this configuration (the first one in Table V) is that the prediction did not stop completely (Figure 6).

If we increase the amount of noise too much (matrix R) the prediction updates too slowly.


Fig. 6. Kalman outputting wrong predictions points for Abandoned Box due to high acceleration (first row of Table V).

### B. Boats

This sequence shows a boat moving at low speed over water in a linearly decelerated movement as it comes to a complete stop and then describes a linear movement in the reverse direction moving to the right of the frame. This sequence is quite complex due to the movement of the water and the fact that we have a road with cars moving in the background of the scene.

Because we have set our goal to track the boat we have set the MOE to get as big a blob as possible on the sail and hull


Fig. 7. Trajectories of measurements and Kalman prediction for Abandoned Box with best parameters (last row of Table V).

TABLE V

SUMMARY OF MOST RELEVANT TEST OVER ABANDONED BOX

| Mod | Meas Noise Cov(R) | Error Post Cov(P) | Process Noise Covariance (Q) | | | EOM params | Result |
|---|---|---|---|---|---|---|---|
| | | | *Pos* | *Vel* | *Acc* | | |
| Con Vel | 30 | 1e2 | 3 | 2 | 0 | UnMod | Bad: trajectory continues going right after stopping |
| Con Vel | 30 | 1e2 | 3 | 0.5 | 0 | Mod | Good |

of the boat. For this we have configured the segmentation to filter out cars and people by selecting a tall aspect ratio. Also a smaller threshold and history value since the boat is white and the reflections are complex. We also applied opening and closing with a rectangular aspect ratio.

With the given tuning of the MOE and a low confidence over the measurements the results were satisfactory as can be seen in the Figure 8 with the parameters of the Table VI.


Fig. 8. Trajectories of measurements and Kalman prediction for Boat with best parameters (last row of Table VI).

### C. Pedestrians

This sequence shows a person walking down a well-lit street. The task is simple for the segmenter and the pedestrian's speed is moderate, although he is walking with strange strides that have come to confuse the segmentation block.

TABLE VI

SUMMARY OF MOST RELEVANT TEST OVER BOAT SEQUENCE.

| Mod | Meas Noise Cov(R) | Error Post Cov(P) | Process Noise Covariance (Q) | | | EOM params | Result |
|---|---|---|---|---|---|---|---|
| | | | ***Pos*** | ***Vel*** | ***Acc*** | | |
| Con Acc | 1e4 | 1e3 | 3 | 0.5 | 0 | UnMod | Good |

TABLE VII

SUMMARY OF MOST RELEVANT TEST OVER PEDESTRIAN SEQUENCE.

| Mod | Meas Noise Cov(R) | Error Post Cov(P) | Process Noise Covariance (Q) | | | EOM params | Result |
|---|---|---|---|---|---|---|---|
| | | | ***Pos*** | ***Vel*** | ***Acc*** | | |
| Con Vel | 1e2 | 1e6 | 10 | 2 | 0 | UnMod | Good |

Because it is a person we have used the elliptical shape for the opening to select the blobs. We have also adjusted a high value for the noise of the measurements because the segmentation did not manage to obtain constant and linear measurements.

The best results are obtained with the parameters shown in Table VII and the filtered trajectory can be observed in the Figure 9. As can be seen in the figure even tough the measurements are noisy Kalman obtains a god filtered version an correctly models the movement of the pedestrian.


Fig. 9. Trajectories of measurements and Kalman prediction for Pedestrian with best parameters (last row of Table VII).

### D. Street Corner at Night

This video shows a car (distinguishable only by the light of its headlights) moving at high speed through an intersection. The sequence is short, so in each frame the object moves a considerable amount of pixels. On the other hand, we have adjusted the segmenter to obtain blobs of a size according to the size of the car. TableVIII shows the parameters of the Kalamn filter and Figure 10 shows how the model manages to obtain a correct trajectory (clearer than the measurements).


Fig. 10. Trajectories of measurements and Kalman prediction for Car with best parameters (last row of Table VIII).

TABLE VIII

SUMMARY OF MOST RELEVANT TEST OVER CARSEQUENCE.

| Mod | Meas Noise Cov(R) | Error Post Cov(P) | Process Noise Covariance (Q) | | | EOM params | Result |
|---|---|---|---|---|---|---|---|
| | | | ***Pos*** | ***Vel*** | ***Acc*** | | |
| Con Acc | 1e4 | 1e5 | 30 | 5 | 1 | UnMod | Good |

## VI. CONCLUSIONS

In conclusion, the Kalman filter proved to be a powerful tool for object tracking in both toy and real videos. Through the iterative process of adjusting the control parameters, we were able to improve the performance of the Kalman filter and obtain better tracking results. We found that the choice of control parameters depends heavily on the specific characteristics of the video sequence, such as the amount of noise and the nature of the object's motion. Additionally, we observed that the performance of the Kalman filter can be improved by tuning the parameters of the blob detection algorithm, which directly affects the quality of the measurements. Overall, this lab provided a valuable hands-on experience in applying the Kalman filter to real-world object tracking problems and demonstrated the importance of careful parameter selection in achieving optimal results.

# APPENDIX
## COLLABORATIVE PROGRAMMING EXPERIENCE

It was not the first time approaching a coding task in a collaborative manner but for these tasks, for the first time, we have coded with a single laptop. This was challenging because we were used to work with our own notes. On the other hand, the experience was more entertaining and we could focus more in the task at hand. In our specific case, since we are working in a C++ with the OpenCV library, collaborative programming was really useful to correct syntax and compilation-time mistakes on the fly.

Regarding the combination of our individual work we inmediately nocied that it was really helpful to work together as we needed to combine all the functions and understand each other's code. In our case, the kalman filter was not encapsulated in a class so we needed extra time to refactor the code. We have completed the lab in 3 intense sessions:

- Session 1 (4 hours): The goal of the first session was merging our individual previous task focusing on cleaning the code and correcting the feedback on the individual work. We kept switching roles every hour or sooner if one of us was tired of typing.
- Session 2 (5 hours): On this session we focused on the toy videos and we pursued to get a deeper understanding on the kalman filter algorithm by combining our knoledge and trying to understant the set of parameters we were going to play with. On this session we were changin roles less frequently because we were just testing different combinations of parameters and trying to guess how to improve our results.
- Session 3 (4 hours): This session was similar to the previous one but now focusing on the real videos.

Another aspect of collaborative programming that we found to be particularly beneficial was the ability to brainstorm ideas together, even when we were not actively coding. We found that this allowed us to explore different solutions and approaches and ultimately led to a more robust and efficient codebase. We would like to highlight is that we found this method so interesting that we believe it could have been introduced for the previous assignments in the subject. This is because when we had access to the guidelines for collaborative programming, we had already been working in 2 labs in a way that was closer to divide and conquer.

Overall, the experience of practicing real collaborative programming for the first time in an image processing task was a valuable learning experience. We found it to be an enjoyable and effective way of coding and would highly recommend it to others. In particular, we found that it was far superior to coding paired with Google as it allowed us to work more closely together and share our ideas more effectively.