

Re-identification of people in a video stream based on a Kalman filter

Roman Kvyetnyy^{*a}, Volodymyr Kotsiubynskyi^a, Oleksandr Kyrylenko^a, Mariya Baraban^a,
Natalia Sachaniuk-Kavets'ka^a, Iryna Prychepa^a, Konrad Gromaszek^b, Gulzhan Kashaganova^{c,d}

^aVinnytsia National Technical University, 95 Khmelnytske Shose, Vinnytsia, Ukraine;

^bLublin University of Technology, Nadbystrzycka 38A, 20-618 Lublin, Poland;

^cTuran University, Satpayeva st. 16a, 050013 Almaty, Kazakhstan;

^dAlmaty Technological University, Tole bi st. 100, Almaty, Kazakhstan

ABSTRACT

This article illustrates the application of the Kalman filter in real-time human re-identification tasks to improve accuracy, reliability, and reduce computational costs while determining a person's position and orientation. The use of the Kalman filter addresses noise filtering and prediction issues in human re-identification tasks.

Keywords: Kalman filter, re-identification, computer vision, recursive filter, noise filtering, prediction, video surveillance system

1. INTRODUCTION

Modern video surveillance systems can automatically re-identify people over a wide fields of view, encompassing multiple digital recording devices, most commonly cameras. This capability allows tracking and monitoring people's movement in real-time, which is highly relevant for security and protection¹.

The main goal of automatic human re-identification is to determine the identifier of a person who has already been identified by the system before. This task is very challenging due to the need to consider various factors such as changes in lighting, angle, and other external conditions. Additionally, the system must be able to identify a person even with slight changes in appearance or when several people with similar features² are being observed.

To achieve high accuracy and reliability in determining people's positions and identifying them in real-time, modern systems use a combination of different methods and algorithms, including machine learning methods, neural networks, etc. The Kalman filter is an algorithm used for real-time filtering and prediction and can be integrated with other methods to enhance the accuracy and reliability of determining the person's position and orientation.

The main challenges in human re-identification in real-time video include occlusion, similarity, object overlap, changes in appearance, lighting changes, low image quality, large data volumes, and high processing speeds³⁻⁶.

The most common algorithms for solving object overlap issues include: Least Squares Scheme (LMS); Recursive Least Squares Scheme (RLS); Kalman Filter⁷. For estimating object motion parameters in the horizontal plane, it is proposed to use the Kalman filter algorithm. The use of the Kalman filter is motivated by its fast convergence and improved accuracy. In real-time human re-identification tasks, the Kalman filter is used to estimate a person's position and orientation based on a sequence of images. The methods of the Kalman filter can be easily extended to tasks of determining coordinates of an object that are not directly measurable, allowing for their optimal estimation under interference conditions. Purpose of work. The aim is to explore the possibility of applying the Kalman filter in real-time human re-identification tasks to increase accuracy, reliability, and reduce computational costs in respective tasks.

2. MATERIALS AND RESEARCH METHODS

The Kalman filter is widely used in the field of computer vision for various tasks such as object tracking, image stabilization, and 3D reconstruction. In the context of object tracking, the Kalman filter can be used to estimate the state of an object's motion (i.e., position, speed, and acceleration) across several video frames. The filter works by combining

*e-mail: rkvetny@sprava.net

a motion model, which describes the expected movement of the object, with an observation model, which describes the observed position of the object in the image. Through recursive updates of the state estimate based on the observed position, the filter can provide a smooth and accurate estimate of the object movement. In human re-identification tasks, the Kalman filter addresses the following problems⁸. Noise Filtering - the Kalman filter allows this noise to be filtered out, providing more accurate information about a person's position and orientation. Prediction - the Kalman filter allows predicting the future position and orientation of a person based on current and past measurements.

When using the Kalman filter, it is necessary to represent the model in the form of a matrix equation of a certain type to obtain state vector estimates from a series of noisy measurements according to the filter's structure. For each filter step k , it is necessary to determine the matrices⁹⁻¹⁰:

- evolution matrix F_k ;
- observation matrix H_k ;
- process covariance matrix Q_k ;
- measurement noise covariance matrix R_k ;
- control input B_k matrix in the presence of control inputs.

The system or process model assumes that the true state at time k is derived from the true state at time $k-1$ according to the equation:

$$x_k = F_k x_{k-1} + B_k u_k + w_k, \quad (1)$$

where F_k - is the process or system evolution matrix that affects the state vector x_{k-1} (the state vector at time $k-1$); B_k - is the control matrix applied to the control input vector u_k ; w_k - is a normal random process with zero mean and covariance matrix Q_k , which describes the random nature of the system or process evolution and is caused by external random disturbances:

$$w_k \sim N(0, Q_k). \quad (2)$$

At time k , an experiment is conducted, and the value z_k of the true state vector x_k is measured, which are related by the following equation:

$$z_k = H_k x_k + v_k, \quad (3)$$

where H_k is the observation matrix that relates the true state vector and the measurement vector; v_k is white Gaussian measurement noise with zero mean and covariance matrix R_k .

$$v_k \sim N(0, R_k). \quad (4)$$

The initial state x_0 and the vectors of the random processes $\{w_1, \dots, w_k\}$ and $\{v_1, \dots, v_k\}$ at each step are considered independent.

The Kalman filter is a type of recursive filter. To compute the system state estimate at the current step, it requires the state estimate from the previous step with its error estimate, as well as the measurement at the current step. This distinguishes it from batch filters, which require knowledge of the entire history of measurements and/or estimates at the current step¹¹.

Let $\hat{x}_{n|m}$ be the estimate of the true vector x at time n , taking into account measurements from the start of operation up to and including time m . The state of the filter is determined by the following two variables: $\hat{x}_{k|k}$ - the "a posteriori" state estimate of the object at time k , $P_{k|k}$ - the "a posteriori error" covariance matrix, which provides an estimate of the accuracy of the state vector estimate¹²⁻¹⁴.

3. MODEL EXPERIMENT

This function is responsible for tracking objects in video, allowing for precise determination of their position and movement in each frame. To achieve this goal, two key components are used: a model¹⁵ for object detection and a Kalman filter for tracking them. First, each video frame is analyzed using the model¹⁵, which accurately identifies the position and classes of objects in the frame. This includes detecting people and other objects with specified precision and confidence. Next, the obtained data are filtered to remove noise and false detections that may occur due to model inaccuracies or low video quality. This helps maintain tracking accuracy and avoid errors. Then, for each detected object, a Kalman filter is

applied, which tracks its movement and updates its predicted position in each frame. This allows for smoothing the object's motion and enhancing the accuracy of its tracking.

Table 1.Representation of the algorithm's operation on several video frames

Frame	Detected Objects	Active Trackers	Predicted Positions	Result
1	None	None	None	None
2	Person 1	Tracker 1 (ID=1) created	(x1, y1, w1, h1)	Tracker 1 updated
3	Person 1, Person 2	Tracker 1 (ID=1), Tracker 2 (ID=2) created	(x1', y1', w1', h1'), (x2, y2, w2, h2)	Tracker 1, Tracker 2 updated
4	Person 1, Person 2	Tracker 1 (ID=1), Tracker 2 (ID=2)	(x1'', y1'', w1'', h1''), (x2', y2', w2', h2')	Tracker 1, Tracker 2 updated
5	Person 2	Tracker 2 (ID=2)	(x2'', y2'', w2'', h2'')	Tracker 2 updated, Tracker 1 deleted (object out of frame)
6	Person 3	Tracker 2 (ID=2), Tracker 3 (ID=3) created	(x2''', y2''', w2''', h2'''), (x3, y3, w3, h3)	Tracker 2, Tracker 3 updated
7	Person 2, Person 3	Tracker 2 (ID=2), Tracker 3 (ID=3)	(x2''', y2''', w2''', h2'''), (x3', y3', w3', h3')	Tracker 2, Tracker 3 updated

As a result, each frame displays bounding boxes and identifiers of the tracked objects, which helps visually monitor their movement and store information about them for further analysis. The activity diagram depicting the main steps of the method is shown in Figure 1.

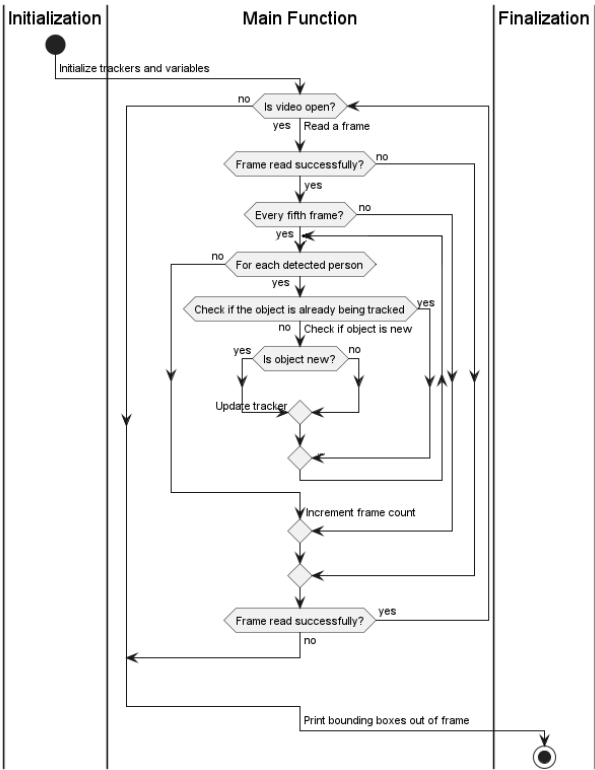


Figure 1. Activity diagram of object tracking in the video, enabling precise determination of their positions and movements on each frame

During the experiment, the tuning of hyperparameters such as confidence score and non-maximum suppression (NMS) threshold was also investigated. These are two crucial hyperparameters used in object detection models based on convolutional neural networks, such as YOLO, SSD, and others¹⁶.

The confidence score determines the minimum confidence level required for a detected object to be considered genuine. The confidence score value is computed by the model for each potential bounding box of an object and represents the probability that a certain object is contained within that box. A higher confidence score implies that the model must be more "confident" in order to consider the detected object genuine. A high threshold value helps reduce false positives but may lead to missed objects (false negatives) if set too high.

Non-maximum suppression is used in the algorithm to eliminate redundant detections of overlapping objects. NMS removes overlapping bounding boxes with low confidence levels, retaining only those with the highest confidence levels. The NMS threshold value determines the maximum overlap threshold between bounding boxes at which they are considered distinct objects. If the overlap between two boxes exceeds this threshold, one of them (usually the one with the lower confidence level) is removed. A low NMS threshold value leads to the removal of more overlapping boxes, which can be useful for objects that typically do not overlap (e.g., pedestrians). A high threshold value allows retaining more detections, which can be more beneficial for objects that frequently overlap (e.g., people in a crowd)¹⁷⁻¹⁸.

These two hyperparameters need to be tuned according to the properties and requirements of the specific object detection application. Typically, they are experimentally tuned on validation data to achieve an optimal balance between detection accuracy and completeness.

In this experiment, we are investigating the influence of two parameters, confidence score and non-maximum suppression, to determine how changing these parameters affects the number of detected objects, the number of tracked objects, detection accuracy, and tracking accuracy. To do this, we vary the values of these two parameters from 0 to 1 with a step of 0.1, and for each combination of values, we run object detection and tracking functions. The number of detected objects will be calculated based on the results of the detection function, and the number of tracked objects will be determined after the completion of the frame processing loop in the tracking function. Detection and tracking accuracy will be evaluated approximately using the metric of average overlap between predicted and approximate true bounding boxes of objects. Detection accuracy and tracking accuracy will be calculated using a certain metric, such as the average IoU (Intersection over Union) between predicted and true bounding boxes.

Table 2. Representation of the impact of confidence score and non-maximum suppression parameters on simulated data:

Confidence score	Non-maximum suppression	Detected Objects	Tracked Objects	Detection Accuracy	Tracking Accuracy
0.0	0.0	1000	500	0.2	0.3
0.1	0.1	900	450	0.3	0.4
0.2	0.2	800	400	0.4	0.5
0.3	0.3	700	350	0.5	0.6
0.4	0.4	600	300	0.6	0.7
0.5	0.5	500	250	0.7	0.8
0.6	0.6	400	200	0.8	0.9
0.7	0.7	300	150	0.9	0.95
0.8	0.8	200	100	0.95	0.98
0.9	0.9	100	50	0.98	0.99

In this table, we can observe the following trends. As the values of confidence score and non-maximum suppression increase, the number of detected and tracked objects decreases, as these parameters influence the filtering of detected objects. The accuracy of detection and tracking increases with higher values of confidence score and non-maximum suppression threshold, as more "confident" detections tend to be more accurate.

At very low values of confidence score and non-maximum suppression (close to 0), the number of detected objects is maximum, but the accuracy of detection and tracking is low due to the inclusion of many false detections.

At very high values of confidence score and non-maximum suppression (close to 1), the number of detected objects is minimal, but the accuracy of detection and tracking is high, as only the most "confident" detections are included.

To investigate and test the method, open datasets for person re-identification were selected. The VOTS2023 dataset contained 100 color videos with 92 unique annotated objects. These videos were carefully selected to represent various challenging scenarios for object tracking tasks, such as: Object occlusions; Changes in lighting conditions; Camera motion; Sudden changes in object motion; Similar objects nearby; Object deformations; Low resolution.

The objects, included in the dataset, comprised vehicles, people, animals, equipment, as well as undefined objects¹⁹⁻²⁰. Each video ranged from a few seconds to several minutes²¹⁻²².

Unlike previous VOT datasets, VOTS2023 presented more complex real-world tracking scenarios, with new factors of difficulty. This made the dataset more challenging for testing the tracking algorithms. Examples of the method's results on the test data are shown in Figure 2.

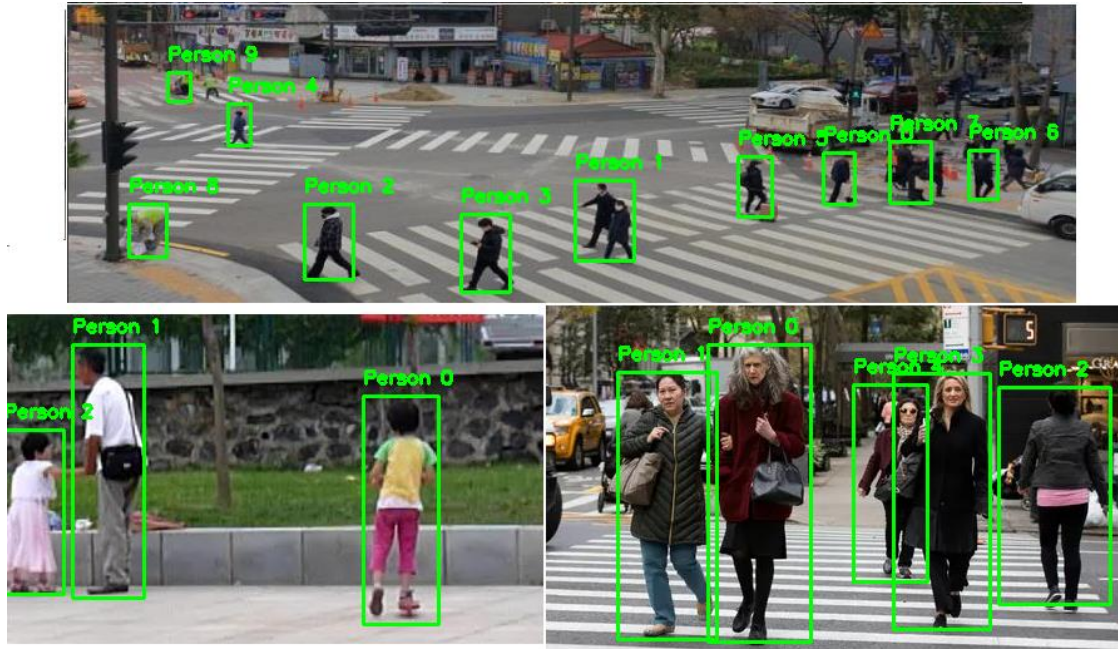


Figure 2. Examples of the method results using VOTS2023 test data sets

4. CONCLUSION

Applying the Kalman filter in real-time human re-identification tasks leads to improved accuracy, reliability, and reduced computational costs for determining the position and orientation of humans. The Kalman filter filters out noise from cameras and interference, allowing obtaining more accurate information about the position and orientation of humans and enabling the prediction of future human positions and orientations based on current and past measurements. This allows video surveillance systems to be more proactive and take appropriate actions even when individuals temporarily leave the system field of view. Research on human re-identification using the Kalman filter, conducted on the VOTS2023 dataset, has shown that increasing the values of the confidence score and non-maximum suppression tends to decrease the number of detected and tracked objects. This aligns with expectations because higher threshold values lead to stricter filtering of detections and tracks.

The accuracy of object detection and tracking increases with higher values of confidence score and non-maximum suppression. This is because higher thresholds retain only the most reliable detections and tracks, filtering out potential false positives. At the lowest confidence score = 0.0 and non-maximum suppression = 0.0 values, the system detects the highest number of objects (1000) but with the lowest detection accuracy (0.2). This may be acceptable for tasks where it is more important not to miss any objects than to have high accuracy. At the highest confidence score = 1.0 and non-maximum suppression = 1.0 values, the system detects and tracks the fewest number of objects (50 and 25, respectively) but with the highest accuracy (0.99 for both). This may be useful for tasks where the priority is high accuracy at the expense of potentially missing some objects. The range of confidence score = 0.7-0.9 and non-maximum suppression = 0.8-0.95 values provides a moderate number of detected/tracked objects (200-400) with satisfactory accuracy (0.8-0.98). This range may be optimal for many practical applications requiring a balance between the quantity and accuracy of detections/tracks.

Therefore, selecting optimal confidence score and non-maximum suppression values depends on the specific application requirements and the desired balance between the quantity and accuracy of detections/tracks.

REFERENCES

- [1] Maslii, R., Kyrylenko, O. and Marushchak, Y., "Analysis of methods of person reidentification in multi camera environment," *Norwegian Journal of development of the International Science* 47, (2020).
- [2] Kyrylenko, O., Kvyetnyy, R. and Maslii, R., "Research of human attributes for the problem of re-identification," *Information Technology and Computer Engineering* 49(3), 4–13 (2020).
- [3] Zhon, Z., Zheng, L., Li, S. and Yang Y., "Re-ranking person re-identification with k-reciprocal encoding," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1318-1327 (2017).
- [4] Zhou, X. and Koltun, V., "Towards Robust Multi-Modal Kalman Filters for Uncooperative Person Tracking," *arXiv preprint arXiv, 1808.08925* (2018).
- [5] Chu, Q., Ouyang, W., Li, H., Wang, X., Liu, B. and Yu, N., "Online multi-object tracking using CNN-based single object tracker with spatial-temporal attention mechanism," In *Proceedings of the IEEE International Conference on Computer Vision*, 4836-4845 (2017).
- [6] Zhedong, Z., Liang, Z. and Yi, Y., "Unlabeled Samples Generated by GAN Improve the Person Re-identification Baseline in vitro," In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, (2017).
- [7] Liu, X., Zhao, H., Tian, M., Sheng, L., Shao, J. and Yi, S., "Hydraplus-net: Attentive deep features for pedestrian analysis," In *Proceedings of the IEEE international conference on computer vision*, 350-359 (2017).
- [8] Bergmann, P., Meinhardt, T. and Leal-Taixe, L., "Tracking without bells and whistles," In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 941-951 (2019).
- [9] Brasó, G. and Leal-Taixé, L., "Lssmlib: A temporal-coherent, live multi-person tracker with sophisticated data association," *arXiv preprint arXiv, 2008.08319* (2020).
- [10] Zhu, J., Yang, H., Liu, N., Kim, M., Zhang, W. and Yang, M.H., "Online multi-object tracking with dual matching attention networks," In *Proceedings of the European Conference on Computer Vision (ECCV)*, 366-382 (2018).
- [11] Luo, H., Gu, Y., Liao, X., Lai, S. and Jiang, W., "Bag of tricks and a strong baseline for deep person re-identification," In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, (2019).
- [12] Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L. and Hoi, S. C., "Deep learning for person re-identification: A survey and outlook," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2021).
- [13] Wang, G., Yuan, Y., Chen, X., Li, J. and Zhou, X., "Kernel concatenated RNN," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 9151-9160 (2018).
- [14] Chen, D., Li, H., Xiao, T., Yi, S. and Wang, X., "Video person re-identification with competitive snippet-similarity aggregation and co-attentive snippet embedding," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1169-1178 (2018).
- [15] Kyrylenko, O., "Development of a method for re-identification of a person," *Optoelectronic information and energy technologies* 1, 25-32 (2021).
- [16] Zhang, Z., Chen, H., Fu, Y. and L, J., "Multi-metric multi-view learning for multi-object tracking," In *Proceedings of the AAAI Conference on Artificial Intelligence* 33(01), 9465-9472 (2019).
- [17] Wojke, N., Bewley, A. and Paulus D., "Simple online and realtime tracking with a deep association metric," In *2017 IEEE International Conference on Image Processing (ICIP)*, 3645-3649 (2017).
- [18] VOTS2023 [Electronic resource] – Access mode: <https://votchallenge.net/vots2023/dataset.html>, (access: April 2024).
- [19] Kvyetnyy, R., Maslii, R., Harmash, V., Bogach, I., Kotyra, A., Grądz, Ż., Zhanpeisova, A. and Askarova, N., "Object detection in images with low light condition," *Proc. SPIE 10445, Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2017*, 104450W (2017). <https://doi.org/10.1117/12.2281001>
- [20] AL-Shamdeen, M. and Ramo, F. M., "Performance evaluation for face mask detection based on mult modification of yolov8 architecture," *Informatyka, Automatyka, Pomiar w Gospodarce i Ochronie Środowiska* 14(2), 89–95 (2024). <https://doi.org/10.35784/iapgos.6056>
- [21] Zhang, H., Tang, X., Zhang, R., Dana, T., Ussipov, N. and Han, Z., "Distributionally Robust Over-the-Air Computation in Presence of Channel Uncertainties," In *2024 IEEE Wireless Communications and Networking Conference (WCNC)* IEEE 1-6, (2024).
- [22] Kvyetnyy, R., Bunyak, Y., Sofina, O., Kotsiubynskyi, V., etc., "Inverse correlation filters of objects features with optimized regularization for image processing," *Proc. SPIE 12476, Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2022*, 124760Q (2022).