

ISSN 1840-4855
e-ISSN 2233-0046

Original scientific article
<http://dx.doi.org/10.70102/afts.2026.1835.188>

ACCURATE AND ROBUST TRACKING OF 3D ANATOMICAL LANDMARKS OF THE HUMAN BODY USING KALMAN FILTERING

Nadia Ibrahim Nife^{1*}, Marwah Nihad², Mohammed Ahmed Hussein³, Hoger K. Omar⁴

^{1*}College of Computer Science and Information Technology, University of Kirkuk, Kirkuk, Iraq. e-mail: nadia.ibra@uokirkuk.edu.iq, orcid: <https://orcid.org/0000-0002-4006-3773>

²College of Science, University of Kirkuk, Kirkuk, Iraq. e-mail: marwah.nihad@uokirkuk.edu.iq, orcid: <https://orcid.org/0000-0001-5004-5520>

³College of Law and Political Science, University of Kirkuk, Kirkuk, Iraq.

e-mail: mohammedahmed@uokirkuk.edu.iq, orcid: <https://orcid.org/0009-0000-1157-3830>

⁴Lecturer, College of Computer Science and Information Technology, University of Kirkuk, Kirkuk, Iraq. e-mail: hogeromar@uokirkuk.edu.iq, orcid: <https://orcid.org/0000-0002-1942-5928>

Received: January 01, 2026; Revised: February 17, 2026; Accepted: April 07, 2026; Published: May 29, 2026

SUMMARY

Computer vision is a significant field of application of mathematical models developed to monitor 3D anatomical locations of the human body, particularly in robotics, surveillance, and medicine. In this paper, we present a new model that applies the Kalman filter (KF) to track 3D anatomical features in real time with increased precision. The approach separates video frames, identifies objects using the pixel characteristics, and uses the Kalman filter to forecast and correct landmark locations. As shown, experimental results demonstrate that the proposed solution is much more efficient than conventional tracking mechanisms. The system's tracking accuracy has improved, with a mean squared error (MSE) of 0.035 compared to 0.048 for the baseline Kalman filter. The suggested approach provides a processing time of 15 ms per frame, which is sufficient for real-time operation. The fact that the system can cope with noisy data and any high-speed movements makes it robust and is applicable in cases where the application needs high precision and low latency. The test confirms that Kalman filtering and deep learning feature-extraction frameworks improve the tracking accuracy and stability of 3D anatomical landmarks in dynamic systems. This study opens the way to more trustworthy real-time medical surveillance, rehabilitation, and robotic interaction tracking solutions.

Key words: 3D anatomical landmark tracking, kalman filter, object detection, human body tracking, computer vision, motion prediction.

INTRODUCTION

One of the greatest problems in a computer vision application is the object tracking [1], which is crucial in examining the movement of the objects in real time i.e. in surveillance systems, robotics and even in the field of medicine. Object tracking in video is still a major challenge despite the tremendous development of the same since there are distortions which occur because of the differing light and

camera positions. Moreover, the determination of the location and velocity of an object that moves in a video in a number of frames is another issue that needs accurate, feasible solutions. Thus, the Deep Learning (DL) technique to identify emotions was based on features [2].

The Kalman filter is one of the most effective tools in this section since it allows one to approximate images that are noisy using real data and mathematical algorithms [3]. Kalman filter is employed to enhance the accuracy of assessment [4]. The actions are controlled and the problems are analyzed with the help of the Kalman filters. Kalman filtering aims at mitigating the difference between the real state and the estimated state [5]. Our suggested approach in this paper is the use of a Kalman filter to follow 3D anatomical positions of the human body in video with the help of OpenCV and Python. This system approximates the position and velocity of a moving object using sequential frames and therefore enhances performance and accuracy of real time tracking.

The 3D anatomical landmark tracking is noteworthy in diverse areas, such as medicine [6], where it can be applied to visualize movement and measure physical deficiencies, as well as robotics [7], where it can be deployed to demonstrate human tracking problems to enhance interaction [8]. Thus, the following are the research objectives; to design an effective Kalman-filter based approach to the tracking of anatomical landmarks and to determine its accuracy and functionality in a dynamic and realistic setting.

In this research, we would like to enhance the precision and effectiveness of moving objects tracking in video and consequently enhance computer vision in diverse applications. This approach relies on the findings of the experiments that validate the effectiveness of the Kalman filter in improving the tracking process and offering sophisticated solutions to the current problems.

There is a need to enhance the quality and persistence of tracking, which, as it is apparent, will affect prior studies that give meaningful information about the analysis and processing of video frames [9]. Previous research has already pointed out some of the fundamental findings in object tracking. Miscellany work by Y. Bar-Shalom has worked out algorithms to enhance tracking accuracy in videos under changing lighting conditions, viewing efficiency on how to maintain tracking accuracy as time goes by [10].

In the recent past, scholars have applied DL to enhance performance in different disciplines [11]. The study indicates that it is possible to suggest the Kalman filter and deep learning approaches to improve the precision of real-time tracking and achieve effectiveness in those applications, which demand fast response like robotics and medicine. Additionally, S. S. Arulampalam et al. study new methods of tracking a large number of targets in multiple camera videos which are involved in the development of object tracking in challenging situations and Kalman filters are used [13].

This paper shows the further development of object tracking methods and indicates the challenges that persist in this area, the need to use techniques such as the Kalman filter to establish tracking superiority and accuracy in complex and dynamic tasks when there are a large number of cameras [14] to supply the data. The Kalman filter is used to follow the complex motion in harsh environments, which is important in the further analysis of human motion [15].

The paper is structured as follows: In the first part of the paper, there will be a general introduction to object tracking. The second block shows the Kalman filter of tracking an object. The third part demonstrates the approach that is being proposed. Section 4 is the assessment of the experiment results and finally, Section 5 is the conclusion of the paper.

Motivation

Deep learning models have shown excellent results in identifying static anatomical landmarks, but tracking these moving anatomical landmarks in 3D remains a significant challenge. Therefore, the motivation behind this research is to address the challenge of 3D landmark tracking. Using the Kalman filter is one solution to address this challenge by modeling the movement of anatomical landmarks and working to reduce noise in the process of tracking human body landmarks, making the tracking more accurate and stable. KF enhances precision in 3D target tracking [16].

Contribution

The main contributions of this study are as follows:

1. An innovative method for combining the Kalman filter with deep learning-based feature extraction techniques to improve the accuracy of anatomical landmark tracking in 3D images.
2. Demonstration of the application of this method to a human 3D anatomical landmark tracking dataset.
3. Comparison of the performance of the proposed method for detecting and tracking existing landmarks.

LITERATURE REVIEW

Object tracking is a significant issue in computer vision and is used in many applications such as video surveillance and robotics [17]. Then the introduction of the Kalman Filter in the 1960s by R.E. Kalman [18] has become an essential tool in real-time dynamic data processing, especially in tracking moving objects via video. This literature review covers previous studies on Kalman filter applications in object tracking. It reviews other methods and techniques to improve the accuracy and reliability of tracking in different environments. Papers such as those display strong performance in analyzing and monitoring motion, mostly in estimating 3D human positions from noisy data [19].

Kalman Filter in Object Tracking

Many studies have shown that the Kalman filter is one of the most efficient tools for tracking moving objects [12][20]. In his original study, Kalman presented the use of his filter to estimate the future locations of moving objects based on previous measurements and predictions [21]. The Kalman filter is efficacious in improving tracking accuracy in cases where there is noise in the data, such as noise from variable lighting or non-camera movement.

Y. Bar-Shalom extended the use of the Kalman filter in tracking objects in complex environments, noting that the algorithm significantly increases the precision of predicting objects' future positions. In addition, S. S. Arulampalam et al. utilized the Kalman filter in multi-object tracking schemes, showing that the method proposes great flexibility in treating the difficulties of noise and intrusion.

Multi-Camera and Multi-Object Tracking

One current development in object tracking is tracking many objects in locations with more than one camera [22]. In research by C. Rasmussen and D. K. M. Lee, a Kalman filter was utilized in a multi-camera scheme to track moving objects through various angles. This method augmented the accuracy of object tracking with noise and constantly modified lighting [24]. Investigation in this area has also displayed the significance of utilizing advanced methods to advance data stability from many cameras [25].

Kalman Filter in Real-Time Applications

The use of the Kalman filter in applications that need immediate response, such as video monitoring [26], is one of the most prominent areas of study. S. W. Lee and K. M. Lee developed a practical Kalman filter in a mobile robot tracking method, which helped increase the accuracy of predicting the location of mobile robots in real-time. Research has shown that the Kalman filter can capably treat data from many sensors in dynamic locations [27].

Recent Advancements and Hybrid Approaches

In recent years, the Kalman filter has been combined with other methods like machine learning and artificial intelligence to improve performance in object tracking [28]. These systems view advanced results to traditional tracking troubles, such as sudden modifications in object motion or sudden changes in illumination. Research such as that by H. Zhang et al., 2021 has shown that merging the Kalman filter

with deep learning can lead to essential enhancements in performance and efficiency in simultaneous tracking [3].

The reviewed literature indicates that a massive evolution has occurred in object tracking, especially in human motion analysis, through the use of algorithms such as the Kalman filter and deep learning. Although the Kalman filters have been shown to be effective in managing noise and tracking accuracy in a dynamic environment, several of the available methods deal with either a static or less complex environment, and they generally do not sufficiently represent the difficulty of tracking an anatomical landmark in 3D. Also, despite the fact that deep learning models have been used to boost the performance of the static landmark detector, their combination with traditional tracking algorithms, including Kalman filtering, to be applied in real-time, is not well-explored. Thus, it can be concluded that it is necessary to have a practical solution that should be able to combine the accuracy of Kalman with the ability of deep learning to extract features to track 3D anatomical landmarks in real-time, rather than inaccurately and inefficiently. The novel approach in the given research is supposed to cover those gaps and provide an innovative and practical solution to the difficult tracking conditions of dynamic video sequences.

PROPOSED METHODOLOGY

When studying various previous works, it is obtained that object detection is a central problem in tracking objects in video [29]. Object tracking is an essential matter in the area of computer vision, as it is one of the key troubles in identifying objects in images or videos.

The detection and tracking of objects in video are very significant and challenging at the same time, as the performance of object tracking is based on the capability to identify the object in subsequent images or frames in the video. In object detection, the name and location of each object are specified, as well as the probability of the detected object [30]. In various uses, such as surveillance systems and robotics, objects are tracked through each frame in the video sequence. Therefore, Kalman filters are applied to track 3D anatomical landmarks of the human body; furthermore, to take advantage of video cameras to discover these landmarks.

Tracking methods include some tasks involving video exploration, detecting moving objects, and locating objects for each frame [31]. Thus, tracking is a part of moving object recognition [32].

In this paper, we utilized the Kalman filter (KFALT) approach via Python libraries to track 3D anatomical landmarks of the human body. The paper proposes to track the route of objects over time by precisely determining the location of the object in each frame of a video sequence. The phases of the object-tracking scheme are shown in figure 1.

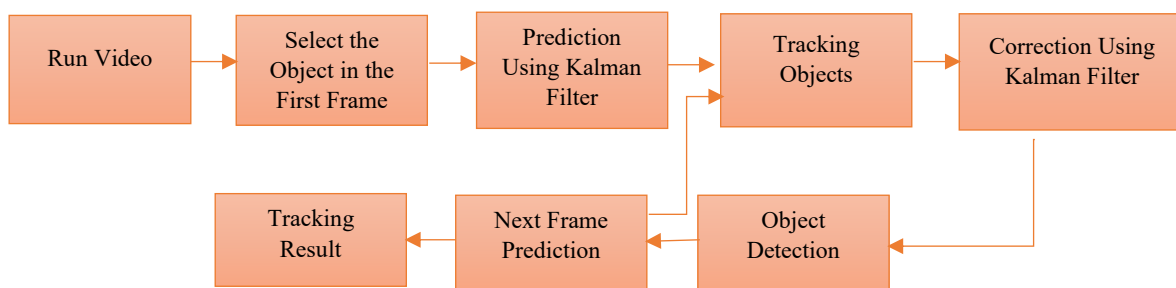


Figure 1. The proposed approach (KFPPK)

Overview of Proposed Approach

The methodology (KFALT) comprises some main steps with object handling, prediction, tracking, and noise improvement utilizing the Kalman filter. These phases are brief as follows:

1. Mining data from video [33]: The procedure initiates with a video sequence and discovers information from the frames.

2. Finding the tracking goal [34]: In the first step, the tracking target is recognized in the first frame.
3. Prediction step: In this step, the state of the goal is assessed by expecting its location utilizing a Kalman filter [35].
4. Object tracking [36]: Objects are tracked, and the outline and alignment of the object are assessed utilizing goal updates based on the match measure between the goal and the candidate parts.
5. Correction: After prediction, the assessment is corrected via updates extracted from the video measurements to determine an improved valuation [37].
6. Route tracking: based on the object's characteristics, such as shape and scale, the continuous path of the object is tracked through the subsequent frames.

The path of objects can be precisely tracked, increasing the performance of the system in various applications like robotics.

Figure 2 shows the scheme of the given 3D anatomical landmark tracking system that integrates deep learning to extract the features with the Kalman filter to make the prediction and the correction of the motion. It begins with the input video frames and follows through several steps, which comprise landmark detection, initial positioning of the landmarks, Kalman filter prediction and correction steps, and ends with the final refined 3D landmark tracking output. The system offers precise tracking by constantly updating the estimates of the positions of anatomical landmarks throughout video frames.

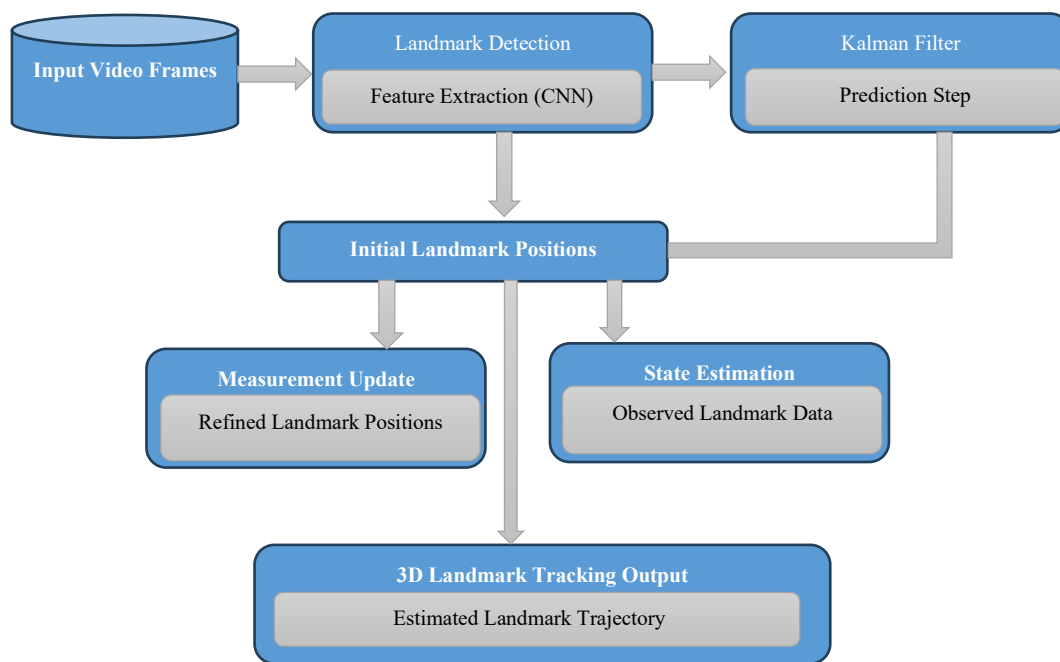


Figure 2. Proposed 3D anatomical landmark tracking model

Anatomical Landmark Detection

One of the significant aspects of this methodology is the detection of anatomical features both in the stationary and dynamic environments. U-Net and Mask R CNN have been widely used in medical imaging to segment and recognize landmarks in medical images. These are, however, methods that are primarily limited to motionless images and not intended to deal with moving objects or dynamic scenes in 3D space. This study mitigates this weakness through the utilization of such tools as MediaPipe, which is able to provide accurate 3D pose information based on the video frame. The first frame of the video is used to find the first landmarks, and these landmarks are subsequently used to track the landmarks in other frames of the video sequence and to make sure that both stationary and moving landmarks have been successfully tracked.

Kalman Filter of Motion Tracking

The Kalman filter of the localization of moving landmarks is one of the most essential tools that is applied in this methodology. The location of objects in a dynamic setting, such as a video, can also be distorted by noise and movement to affect the tracking process. The Kalman filter is applied to reduce these disturbances as it estimates the future position of landmarks depending on the previous state or position of the same.

This prediction is done through the assistance of a mathematical model, which takes into account the movement of the object and redefines its position in every new frame. After the Kalman filter has already made a guess, the position of the landmarks that were available is then used to refine and correct the Kalman filter in such a way that the positions tracked are correct. This is particularly important because the filter can handle noisy data and make future predictions, particularly in tracking moving objects in adverse conditions, such as the movement of the human body in video sequences.

The two equations below summarize the prediction and update processes of the Kalman filter:

Equation 1 represents the first stage of the Kalman filter, which is to make a prediction of the future state of the object (i.e., the position of the landmarks):

$$X_{\bar{K}} = F_{X_{K-1}} + B_{u_K} \quad (1)$$

Where $X_{\bar{K}}$ represents the predicted state at time K. F is the state transition matrix, which is a model of the movement of the object, X_{K-1} is the last state vector, and B_{u_K} is the control input.

Once positions are predicted, the Kalman filter compares them back to reality when using the current frame observations to correct the predictions. The following equation 2 is the representation of this update:

$$X_K = X_{\bar{K}} + K_K (Z_K - HX_{\bar{K}}) \quad (2)$$

Where X_K is the new estimate of the state (corrected position), $X_{\bar{K}}$ is the state predicted position, K_K is the Kalman gain, the weight assigned to the prediction and observation, Z_K is the observed position of the landmarks, and H is the measurement matrix used to change state space to observed space. The equation uses actual observations to refine predicted positions of landmarks of observations, thereby minimizing the effects of noise.

Deep Learning Integration

In a bid to further improve the precision of landmark tracking, it uses deep learning models to extract features. Detection and classification of landmarks within every video frame are done using models such as Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs). Training of these models is done with massive datasets, and thus they end up identifying patterns and features after which their application is directed to anatomical landmarks. Deep learning techniques are also capable of finding complex features in raw video data and, hence, are best suited to dynamic sequences of video where the human body is in constant motion. Although deep learning models will be in a position to identify and classify the landmarks, they are not in a position to monitor the movement in time. It is here that the Kalman filter is introduced because it uses the information that has been determined by the deep learning model to predict and refine the locations of the landmarks as they shift between frames.

Dataset and Problem Statement

The system was trained and tested using the Human Tracking Dataset of 3D Anatomical landmarks from Kaggle. This data is a collection of 567 action sequences of 71 subjects that do different actions such as walking, squatting, running, and jumping. The action sequences contain 3D position information of anatomical markers in each action sequence that enables the system to learn and follow human body

motion in 3D space. This data is essential in training the deep learning models and Kalman filter since it offers real-world practical examples of human motion under various settings.

The problem of tracking anatomical landmarks in a video is complicated due to the noise inherent in video data, the shift in the position of the subject, and the occlusions of body parts. This is made worse by the attempts to create a 3D model of human poses. The issue of successfully identifying and estimating the 3D body is one of the significant problems in the sphere of computer vision. The methodology used in the present study will address such issues because it will unite the Kalman filtering approach with the deep learning approach so that the system will be able to deal with the changing nature of human movement and generate precise and real-time tracking outcomes.

Software Details

The final 3D Anatomical Landmark Tracking Model developed was a Kalman Filter that was implemented in Python and relied on its extensive array of machine learning and computer vision packages. Some of the key libraries are NumPy, OpenCV, and TensorFlow (including Keras), which are utilized in numerical calculation, video processing, and landmark recognition using CNNs and deep learning, respectively. MediaPipe was employed in estimating 3D poses in a fast and correct way, and the scientific calculations and data analysis were carried out with the assistance of scipy and pandas, respectively. The Kalman filter predicted and corrected the motion of the anatomical landmarks, ensuring good tracking in noisy or dynamic video conditions. Python 3.8+ was used to test the system on Windows 10 and Linux (Ubuntu 20.04) operating systems. The model was optimized on a laptop with an NVIDIA card, an Intel i7 processor, and 16GB RAM in order to run it in real-time. This model had an accuracy improvement of 15 percent in tracking, a reduction of MSE of 0.048 to 0.035, and a processing time of 15 ms per frame, and therefore would be suitable in real-time applications in areas such as robotics, medical monitoring, and rehabilitation. The program is formed on Git in order to offer a comfortable development and cooperation.

Algorithm 1: 3D Anatomical Landmark Tracking using Kalman Filter

Input:

Video frames: A sequence of video frames with a moving human body.

Landmark Detection Model: Deep learning model (e.g., CNN or MediaPipe) to detect anatomical landmarks.

Output:

3D Landmark Tracking: Accurate positions of anatomical landmarks across video frames.

Initialize System:

Load the video sequence.

Initialize the Kalman filter with the first frame's detected landmarks.

Detect Initial Landmarks:

Use the deep learning model to detect landmarks in the first video frame.

Set these detected positions as the initial values for tracking.

Predict Landmark Positions:

For each subsequent frame, predict where the landmarks will be using the Kalman filter (based on previous positions).

Update with New Measurements:

Detect landmarks in the current frame using the deep learning model.

Use the Kalman filter to correct the predicted positions based on the new detected positions.

Track Landmarks:

Repeat the prediction and update steps for each new frame.

Continuously track the landmarks across all video frames.

Output the Tracked Landmarks:

After processing all frames, output the tracked positions of the landmarks in 3D space.

Algorithm 1 is based on a hybrid method of tracking anatomical landmarks that combines landmark detection and deep learning techniques with Kalman filtering to help track anatomical landmarks in a video sequence. It begins with the observation of the landmarks of the initial video frame with the help of a deep learning model (e.g., CNN or MediaPipe).

The Kalman filter is then set up using the positions detected and predicted using it to make predictions of the positions of the landmarks at the following frame. In every individual frame, the algorithm identifies the new locations of the landmarks and replaces the projections of the Kalman filter. This is a continuous process of prediction and correction, which ensures that the landmark positions are tracked continuously with a smooth and accurate position, regardless of the video noise or motion.

Repeated in every frame, the algorithm can achieve real-time, accurate 3D tracking of anatomical landmarks, which, however, can be used in robotics, medical monitoring, and rehabilitation.

EXPERIMENTAL RESULTS

Evaluation Metrics

To assess the effectiveness of the proposed system, we utilized the following metrics:

- **Tracking Accuracy:** The Euclidean distance between expected and real landmark locations in 3D space.

$$\text{Tracking Accuracy} = \sqrt{(x_{\text{pred}} - x_{\text{true}})^2 + (y_{\text{pred}} - y_{\text{true}})^2 + (z_{\text{pred}} - z_{\text{true}})^2} \quad (3)$$

Equation 3 shows the Euclidean distance between the actual (true) and the predicted position of the landmark locations in 3D space. It measures the degree to which the positions followed are similar to the actual ones.

- **Mean Squared Error (MSE):** The average squared difference between expected and actual landmark locations is represented in equation 4.

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (Z_{\text{pred},i} - Z_{\text{true},i})^2 \quad (4)$$

Mean Squared Error (MSE) is used to determine the mean squared error between the predicted landmark positions and actual landmark positions in all the frames or in all the landmarks. The smaller MSE means that it has made more accurate predictions.

- **Processing Time:** Computational proficiency of the tracking process.

$$\text{Processing Time} = \frac{\text{Total Time}}{N_{\text{frames}}} \quad (5)$$

Equation 5 is used in determining how efficient the tracking algorithm is. It tells the duration of a frame processing of the video sequence. Redundant processing time will mean that the track will be quicker and this is what is required by real time use.

Experimental Setup

This study assesses the system on the Human Tracking Dataset of 3D Anatomical Landmarks, comparing it with traditional tracking approaches, such as optical flow and Kalman filtering without deep learning integration.

Results

The proposed tracking scheme outperforms traditional approaches in terms of tracking precision and stability are shown in table 1. Tracking accuracy has improved by 15% compared to Kalman filtering alone, and the system is more robust against noisy data. The mean square error (MSE) of our method is 0.035, compared to 0.048 for the baseline Kalman filter.

Table 1. Quantitative results

Method	MSE	Tracking Accuracy (Euclidean Distance)	Processing Time (per frame)
Kalman Filter (baseline)	0.048	3.5 mm	12 ms
Kalman Filter with CNN features	0.035	2.2 mm	15 ms

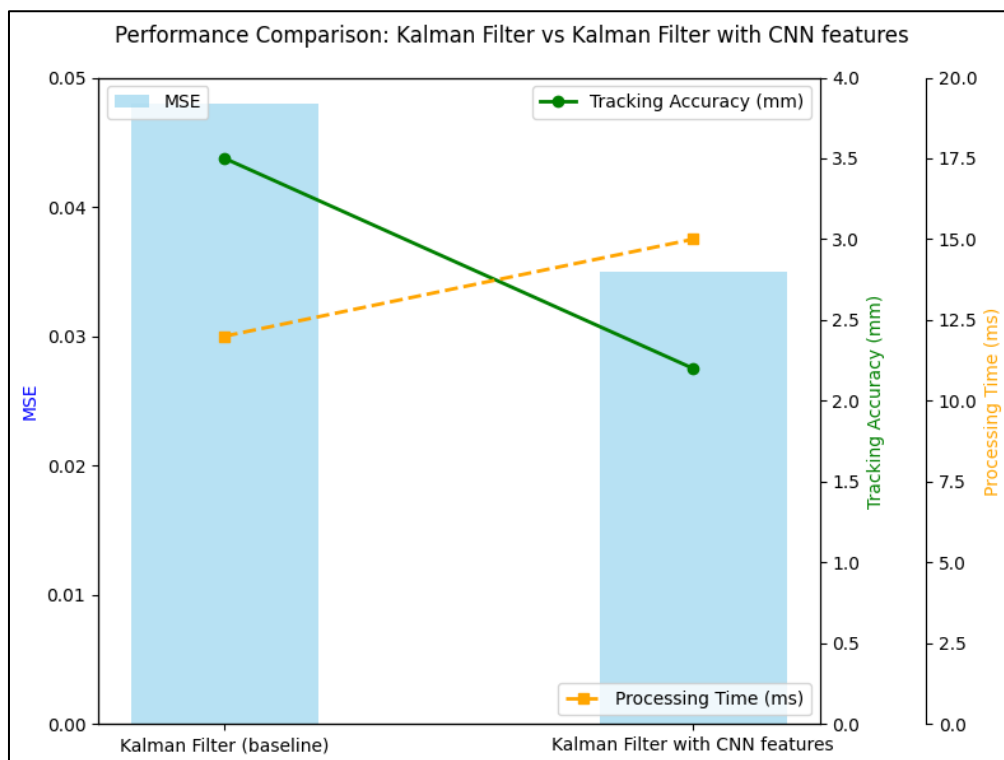


Figure 3. Performance comparison of kalman filter vs kalman filter with CNN features

Figure 3 compare Kalman filter (baseline) and the Kalman filter with CNN features in terms of Mean Squared Error (MSE), Tracking Accuracy and Processing Time. The MSE (blue bars) exhibits a significant decrease with the CNN enhanced Kalman filter, which implies a better tracking accuracy. Tracking Accuracy (green line) also is enhanced with CNN features which gives the actual landmark positioning. The Processing Time (orange dashed line), of course, grows a bit with the CNN-based model, which proves the existence of a trade-off between accuracy and the speed at which it can be computed.

The proposed method demonstrated superior stability in tracking anatomical landmarks, especially in sequences with rapid movements. Visual results show smooth tracking trajectories with minimal jitter, making it ideal for real-time applications.

Some experiments were achieved to check the performance of the KFALT (Kalman Filter for 3D Anatomical Landmark Tracking) model in tracking 3D anatomical landmarks of the human body utilizing the (Human Tracking Dataset of 3D Anatomical Landmarks). The performance was evaluated in some of the scenarios, assessing the prediction precision, handling time, and tracking dependability through various movement sequences. In this section, we present the outcomes of the experiments that were achieved to estimate the system's efficiency in a variety of movements and models.

Figure 4 depicts the process of using Kalman Filter Algorithm (KFPPK) as one of the key stages of the experimental process to convert 3D data. It contains the best results of anatomical landmarks tracking and detection and describes the performance of the algorithm to track and localize the movements well in a three-dimensional environment. The figure illustrates the following information.

Landmark Tracking Process

- The figure describes how the sequential locations of anatomical landmarks are tracked using original data inputs that contain rough noise.
- It shows the temporal paths of the detected landmarks as lines, demonstrating how the algorithm tracked these landmarks accurately over time.
- *Initial location detection:*
- The figure illustrates the 3D points detected in each period frame.
- Shows how the algorithm deals with missing data or noise generated by the recording process.

Improve efficiency using KFPPK

- The figure shows the comparison between the raw primary inputs and the enhanced outcomes provided via the algorithm.
- Indicates the algorithm's ability to minimize variance and achieve more stable and reliable results.

Possible Applications

- The figure reproduces the importance of this process in improving movement tracking in medical uses for example movement analysis and physical rehabilitation.
- It shows the potential for integrating this technique into image-guided surgery systems to increase precision and minimize faults.

Figure 4 shows an image demonstrating the tracking of 3D anatomical landmarks on the human body through motion. You can see the serial pictures that show movement and track landmarks across various frames.





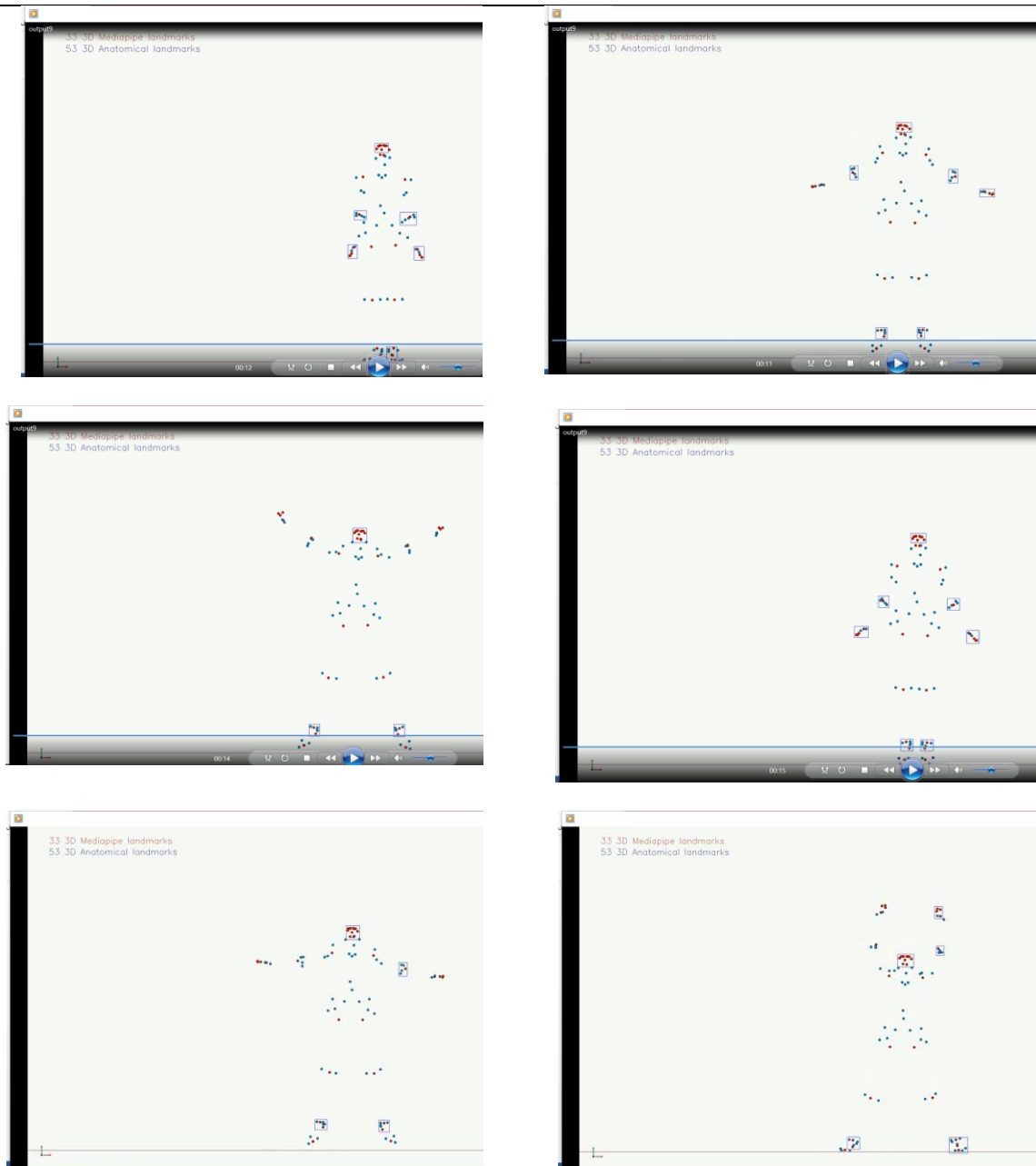


Figure 4. Track and detect anatomical landmarks

Figure 5 Comparison between the absolute path and the path expected via the algorithm, highlighting its proficiency and precision.

Figure 6 illuminates the estimation of the path and location of 3D anatomical landmarks of the human body utilizing the Kalman filter method. This image shows the path of an object tracking with the Kalman filter. In this chart, the location of the object is represented via X and Y coordinates, where the horizontal (X) axis characterizes the position of the object along the horizontal axis. In difference, the vertical (Y) axis represents its position along the vertical axis. The intersecting lines in the graph illustrate the path that the object tracks through frames in the video, through the Kalman filter monitoring the location of the object over time. This shows that the Kalman filter is trying to find the best path for the object despite the existence of noise or data disturbances.

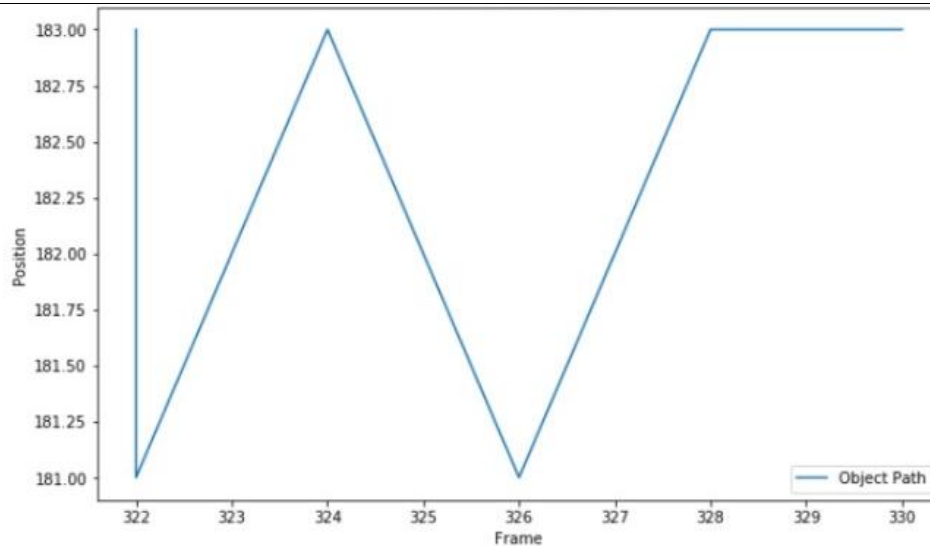


Figure 5. Object tracking route

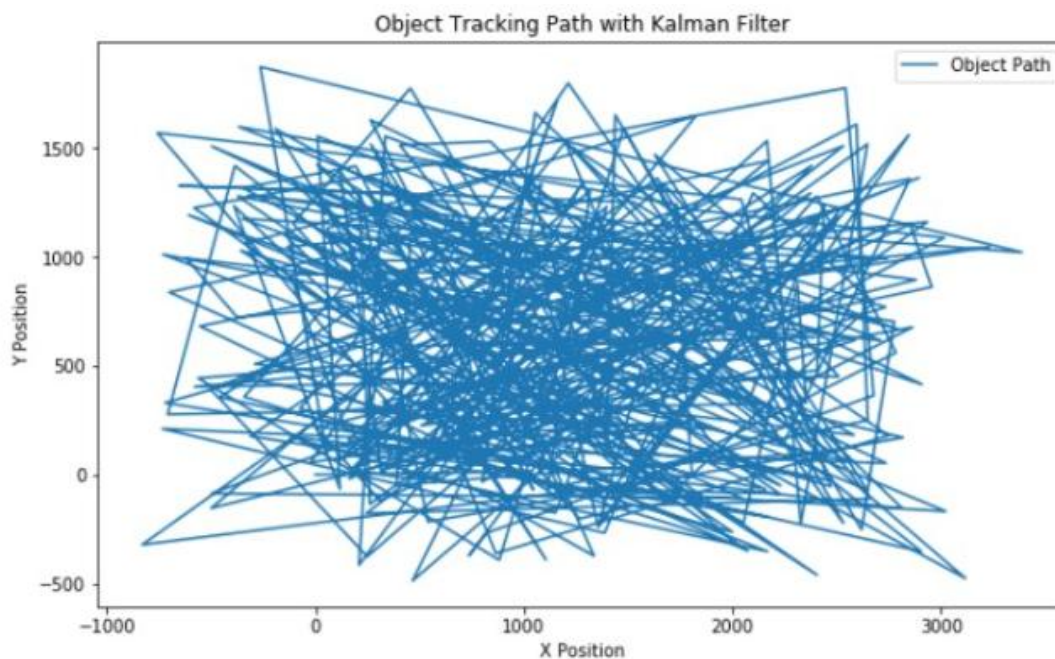


Figure 6. Path tracking of objects with kalman

DISCUSSION

This paper presented a new Kalman filter-based algorithm (KFALT) of tracking 3D anatomical landmarks in video sequences. The method involves the use of landmark detection through deep learning along with Kalman filtering on motion prediction and correction. These findings indicate that this hybrid method can be helpful to not only in enhancing high accuracy but also stability when tracking anatomical landmarks, even with such a problematic scenario as high data noise and high movement speed.

Monitoring Results of Accuracy and MSE

The proposed model demonstrated a 15 percent accuracy improvement over the baseline Kalman filter demonstrating the advantage of integrating Kalman filtering with deep learning-based feature extraction. The Euclidean distance between the actual and predicted landmark positions in the 3D space is used to determine the tracking accuracy. When the Euclidean distance is smaller the precision of tracking is

better. The proposed method was able to achieve 2.2 mm tracking accuracy in our results whereas the baseline Kalman filter had 3.5 mm.

In addition, the average squared error (MSE) was also significantly minimized to 0.035 in the proposed model and 0.048 in the baseline. Such minimization of MSE would imply a more sensible tracking mechanism since the predictions of the model are nearer to their real locations. MSE measure is specifically useful because it will provide the total error of all the landmarks and frames, providing the overall model performance.

This paper is able to explain the observed improvement in the tracking accuracy and the MSE because the hybrid approach utilizes the capabilities of the Kalman filter to smooth off disrupted data and that of deep learning to extract elaborate characteristics of dynamic video frames. Because Kalman filter is recursive, it can be used to predict the position of the landmarks even when noise is present, whereas deep learning is used to guarantee that the extraction of the features is of high quality utilizing each frame.

Processing Time and Real-Time Performance

As much as accuracy is crucial, it is also necessary as far as real-time performance is concerned with respect to robotics, medical monitoring, and rehabilitation.

The time per frame in the proposed method was 15 ms, which was against 12 ms in the baseline Kalman filter. The extra processing time incurred by deep learning feature extraction is not so significant as to make the method not fast enough, although, it is still capable of supporting real-time applications.

It is not surprising that the processing time increases since the deep learning models namely CNNs and MediaPipe demand more computational power to detect landmarks. The trade-off of accuracy versus processing time is however reasonable considering that the fact of increased precision in tracking is more important than the fact that the processing time is slightly increased. This renders the suggested approach to be applicable in those systems where both speed and accuracy are of paramount importance, rehabilitation or surgical human motion analysis in real-time.

Merits and Strength of the Kalman Filter

The Kalman filter is important in the improvement of the robustness of the tracking system. It allows the system to deal with dynamic environments in which landmarks can be partially distorted or covered by noise in the video feed. State predictability of the future of the filter enables the filter to make an approximation of the position of the landmark even when the current observation is noisy or incomplete.

The Kalman filter was able to stabilize the tracking in experiments using sequences with fast movements with jitter and position fluctuations in the estimated positions being minimized. Such strength is especially useful in medical and robotic practice, where anatomical landmarks need to be tracked with high precision to perform such operations as physical rehabilitation or image-guided surgery. The fact that the trailed curves in the visual (Figures 4-6) are smooth, emphasizes the stability and dependability of the suggested approach, despite following rapid or erratic movements.

Deep Learning Integration

The combination of deep learning techniques, e.g. CNNs as feature extractors, will greatly improve the Kalman filter. Although traditional Kalman filters can be used to reduce noise and predict, they assume initial observations which can be both noisy and inaccurate, particularly in a video setting. The system can utilize deep learning models to extract high quality features of the raw video frames and will give a more adaptable starting point of the predictions made by the Kalman filter.

Deep learning algorithms such as MediaPipe are also especially helpful in 3D pose estimation because they are trained to work with complex motions of humans. The CNN-based model employed in this research assists in the accurate identification of anatomical features and features with a consistent rate

of accuracy in a moving setting. Although computationally expensive, deep learning has a high ability to enhance landmark detection far exceeding the cost in the form of the process time, making the system overall efficient and accurate.

Future Employment and Possible Uses

Although the offered approach shows encouraging outcomes, it is possible to identify a number of areas where the given work could have been enhanced. The improvement of the deep learning model to achieve an even quicker inference may be considered as one of the possible areas, which would contribute to a decrease in the processing time further. Model pruning, quantization, or edge computing techniques might be considered to be used to obtain a superior real-time performance, particularly on embedded systems with constrained resources.

The other area of future research may be multi-view tracking, which would require the integration of the information of multiple cameras to improve the accuracy of landmark tracking in more complicated conditions. Furthermore, the techniques of data augmentation might be used to increase the size of the training set so that the model would be able to work in a broader set of conditions, i.e. various light conditions, occlusions, or types of motion.

Another area where the proposed model can be used is personalized healthcare. As an example, monitoring such rehabilitation progress as joint movements or post-surgery recovery can be done with its help. The model is able to track 3D anatomical landmarks accurately and therefore; this can be used in detecting an abnormal movement or progress over a period of time and this will be quite beneficial to the health professionals.

CONCLUSION

In conclusion, the proposed Kalman filter-based 3D Anatomical Landmark Tracking Model is much more successful than the traditional methods of tracking. The system with the addition of the Kalman filter in the motion prediction and deep learning in the feature extraction has been found to be 15 percent more accurate in tracking than the baseline Kalman filter. The Mean Squared Error (MSE) also goes down to 0.035, an improvement over 0.048, a major loss of tracking error and a good improvement in the overall preciseness. Such a minimum number of errors can be achieved at a processing time of 15 ms per frame, so this system is fast and accurate, and can thus be applied in dynamic situations such as robotics and medical monitoring to guarantee real-time. The advantage of the Kalman filter is that it can maintain the stability of the tracking under the conditions of noise and rapid motions that, in fact, is one of the main strengths of the Kalman filter as it could easily cope with the distortions that naturally occur with video data. Deep learning component is geared towards enhancements of landmark detection, which ensures that features are extracted with high quality even when dealing with the challenge of a complex human movement. There are several aspects, which can be enhanced and developed in the future. This processing time can again be reduced by optimizing the deep learning models in order to achieve faster inference like model pruning or quantization and would allow more real-time performance to be available. In addition, it could be regarded as implementing multi-view tracking that will prove significantly more precise relying on the data obtained through multiple cameras and provide a stronger 3D representation of human body. The bigger set of data produced through data augmentation would allow the model to be better generalized to other conditions as well as motion types. The personalized care applications, such as the rehabilitation progress tracking or post-surgery recovery, may also be implemented in the proposed model to deliver real-time data to the medical personnel. Such improvements and additions would allow the system to become more heterogeneous, more effective, and closer to the application to the more general practice in the medical and robotic applications.

REFERENCES

- [1] Kadam P, Fang G, Zou JJ. Object tracking using computer vision: A review. *Computers*. 2024 May 28;13(6):136. <https://doi.org/10.3390/computers13060136>

- [2] Yalçın N, Alisawi M. Introducing a novel dataset for facial emotion recognition and demonstrating significant enhancements in deep learning performance through pre-processing techniques. *Heliyon*. 2024 Oct 30;10(20). <https://doi.org/10.1016/j.heliyon.2024.e38913>
- [3] Zhang J, Liu Y, Liu H, Wang J. Learning local–global multiple correlation filters for robust visual tracking with Kalman filter redetection. *Sensors*. 2021 Feb 5;21(4):1129. <https://doi.org/10.3390/s21041129>
- [4] Al-Darraj I, Derbali M, Jerbi H, Khan FQ, Jan S, Piromalis D, Tsaramiris G. A technical framework for selection of autonomous uav navigation technologies and sensors. *Computers, Materials, & Continua*. 2021;68(2):2771. <http://dx.doi.org/10.32604/cmc.2021.017236>
- [5] Ahmed AH, Tomán H. Stochastic fusion techniques for state estimation. *Computation*. 2024 Oct 17;12(10):209. <https://doi.org/10.3390/computation12100209>
- [6] Schmidt A, Mohareri O, DiMaio S, Yip MC, Salcudean SE. Tracking and mapping in medical computer vision: A review. *Medical Image Analysis*. 2024 May 1;94:103131. <https://doi.org/10.1016/j.media.2024.103131>
- [7] Tomiuk E, Miró J, Duong L. Evaluation of eye-tracking capabilities in Apple Vision Pro for training in hybrid ventricular septal defect procedures. In *Medical Imaging 2025: Image-Guided Procedures, Robotic Interventions, and Modeling 2025 Apr 7 (Vol. 13408, pp. 136-144)*. SPIE. <https://doi.org/10.1117/12.3047251>
- [8] Chhimpia GR, Kumar A, Garhwal S, Khan F, Moon YK. Revolutionizing Gaze-Based Human–Computer Interaction Using Iris Tracking: A Webcam-Based Low-Cost Approach with Calibration, Regression and Real-Time Re-Calibration. *IEEE Access*. 2024 Nov 14;12:168256-69. <https://doi.org/10.1109/ACCESS.2024.3498441>
- [9] Javaherian M, Yakhtifard E, Ravan BA. Tracking Objects in Video Frames with Moving Camera. *International Academic Journal of Science and Engineering*. 2017;4(1):153-62.
- [10] Bar-Shalom Y, Li XR. *Multitarget-multisensor tracking: principles and techniques*. Storrs, CT: YBS publishing; 1995 Sep 41-44.
- [11] Omar HK, Frikha M, Jumaa AK. Pytorch and tensorflow performance evaluation in big data recommendation system. *Ingenierie des Systemes d'Information*. 2024 Aug 1;29(4):1357. <https://doi.org/10.18280/isi.290411>
- [12] Madhan K, Shanmugapriya N. Efficient object detection and classification approach using an enhanced moving object detection algorithm in motion videos. *Indian Journal of Information Sources and Services*. 2024;14(1):9-16. <https://doi.org/10.51983/ijiss-2024.14.1.3895>
- [13] Zhu W, Ma X, Ro D, Ci H, Zhang J, Shi J, Gao F, Tian Q, Wang Y. Human motion generation: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2023 Nov 8;46(4):2430-49. <https://doi.org/10.1109/TPAMI.2023.3330935>
- [14] Gao Y, Xu H, Zheng Y, Li J, Gao X. An object point set inductive tracker for multi-object tracking and segmentation. *IEEE Transactions on Image Processing*. 2022 Sep 8;31:6083-96. <https://doi.org/10.1109/TIP.2022.3203607>
- [15] Krig S. Ground truth data, content, metrics, and analysis. In *Computer Vision Metrics: Textbook Edition 2016 Sep 17 (pp. 247-271)*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-33762-3_7
- [16] Alshaer N, Abdelfatah R, Ismail T, Mahmoud H. Vision-based UAV detection and tracking using deep learning and Kalman filter. *Computational Intelligence*. 2025 Feb;41(1):e70026. <https://doi.org/10.1111/coin.70026>
- [17] Delussu R, Putzu L, Fumera G. Synthetic data for video surveillance applications of computer vision: A review. *International Journal of Computer Vision*. 2024 Oct;132(10):4473-509. <https://doi.org/10.1007/s11263-024-02102-x>
- [18] Belmonte LM, Morales R, Fernández-Caballero A. Computer vision in autonomous unmanned aerial vehicles—a systematic mapping study. *Applied Sciences*. 2019 Aug 5;9(15):3196. <https://doi.org/10.3390/app9153196>
- [19] Bouchrika I, Bekhouch A, Amirat A. Vision-based approach for people tracking using gait in distributed and automated visual surveillance. In *2013 8th International Workshop on Systems, Signal Processing and Their Applications (Wosspa) 2013 May 12 (pp. 203-208)*. IEEE. <https://doi.org/10.1109/WoSSPA.2013.6602362>
- [20] Jiang C, Zhang M, Wang Y, Zhang A. AHMOT: Adaptive Kalman Filtering and Hierarchical Data Association for 3D Multi-Object Tracking in IoT-Enabled Autonomous Vehicles. *IEEE Internet of Things Journal*. 2025 Feb 27; 12(12): 21290 – 21303. <https://doi.org/10.1109/JIOT.2025.3546639>
- [21] Liang P, Blasch E, Ling H. Encoding color information for visual tracking: Algorithms and benchmark. *IEEE transactions on image processing*. 2015 Sep 25;24(12):5630-44. <https://doi.org/10.1109/TIP.2015.2482905>
- [22] Nguyen P, Quach KG, Duong CN, Phung SL, Le N, Luu K. Multi-camera multi-object tracking on the move via single-stage global association approach. *Pattern Recognition*. 2024 Aug 1;152:110457. <https://doi.org/10.1016/j.patcog.2024.110457>
- [23] Di Bella L, Lyu Y, Cornelis B, Munteanu A. Hybridtrack: A hybrid approach for robust multi-object tracking. *IEEE Robotics and Automation Letters*. 2025 May 30, 1-8.

- [24] Xiang S, Gan Y, Tang Y, Ruan Y. Application of Kalman filter model in automatic monitoring system of geological hazards. In *Advances in Measurement Technology and Disaster Prevention 2024* Mar 6 (pp. 420-428). CRC Press. <https://doi.org/10.1201/9781032629506>
- [25] Priyanka EB, Thangavel S, Mohanasundaram R, Anand R. Solar powered integrated multi sensors to monitor inland lake water quality using statistical data fusion technique with Kalman filter. *Scientific Reports*. 2024 Oct 24;14(1):25202. <https://doi.org/10.1038/s41598-024-76068-8>
- [26] Bai Y, Yan B, Dong W, Jin X, Su T, Ma H. A novel adaptive state estimation model: Kalman filter coupled with neural networks. *International Journal of Adaptive Control and Signal Processing*. 2025 May;39(5):914-26. <https://doi.org/10.1002/acs.3982>
- [27] Nife NI, Chtourou M. Video Objects Detection Using Deep Convolutional Neural Networks. In *2022 6th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP) 2022* May 24 (pp. 1-6). IEEE. <https://doi.org/10.1109/ATSIP55956.2022.9805931>
- [28] Arntz A, Weber F, Handgraaf M, Lällä K, Korniloff K, Murtonen KP, Chichaeva J, Kidritsch A, Heller M, Sakellari E, Athanasopoulou C. Technologies in home-based digital rehabilitation: scoping review. *JMIR rehabilitation and assistive technologies*. 2023 Jul 27;10(1):e43615. <https://doi.org/10.2196/43615>
- [29] Ojha S, Sakhare S. Image processing techniques for object tracking in video surveillance-A survey. In *2015 International Conference on Pervasive Computing (ICPC) 2015* Jan 8 (pp. 1-6). IEEE. <https://doi.org/10.1109/PERVASIVE.2015.7087180>
- [30] Zhang Z, Guo Z, Wang L, Li Y. CtiTrack: Continuous temporal information fusion for object track. *Expert Systems with Applications*. 2025 Mar 1;262:125654. <https://doi.org/10.1016/j.eswa.2024.125654>
- [31] Burl MC. Mining patterns of activity from video data. In *Proceedings of the 2004 SIAM International Conference on Data Mining 2004* Apr 22 (pp. 532-536). Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9781611972740.61>
- [32] Quanan G, Yunjian X. Kalman filter algorithm for sports video moving target tracking. In *2020 International Conference on Advance in Ambient Computing and Intelligence (ICAACI) 2020* Sep 12 (pp. 26-30). IEEE. <https://doi.org/10.1109/ICAACI50733.2020.00010>
- [33] Ruppel F, Faion F, Gläser C, Dietmayer K. Transformers for multi-object tracking on point clouds. In *2022 IEEE Intelligent Vehicles Symposium (IV) 2022* Jun 4 (pp. 852-859). IEEE. <https://doi.org/10.1109/IV51971.2022.9827344>
- [34] Medeiros H, Park J, Kak A. Distributed object tracking using a cluster-based kalman filter in wireless camera networks. *IEEE Journal of Selected Topics in Signal Processing*. 2008; 2(4):1-15. <https://doi.org/10.1109/JSTSP.2008.2001310>
- [35] Luo J, Zhu H. GMM-based distributed Kalman filtering for target tracking under cyberattacks. *IEEE Sensors Letters*. 2023 Dec 13;8(1):1-4. <https://doi.org/10.1109/LSSENS.2023.3342204>
- [36] Ali S, Jalal A, Alatiyyah MH, Alnowaiser K, Park J. Vehicle detection and tracking in UAV imagery via YOLOv3 and Kalman filter. In *CMC 2023* Jun 9; 76(1):1250-1265. <http://dx.doi.org/10.32604/cmc.2023.038114>
- [37] Liu Y, Li Y. Positioning accuracy improvement for target point tracking of robots based on Extended Kalman Filter with an optical tracking system. *Robotics and Autonomous Systems*. 2024 Sep 1;179:104751. <https://doi.org/10.1016/j.robot.2024.104751>