International Journal of Scientific Research in Science, Engineering and Technology

Print ISSN - 2395-1990
Online ISSN : 2394-4099

OPEN ACCESS

Available Online at : www.ijsrset.com
doi : https://doi.org/10.32628/IJSRSET

Empirical Analysis of Machine Learning Models Used for Motion Detection & Tracking in Videos from An Augmented Statistical Perspective

Shivali Tiwari, Prof. Jayant Adhikari

Department of Computer Science and Engineering, Tulsiramji Gaikwad-Patil College of Engineering & Technology, Nagpur, Maharashtra, India

ARTICLEINFO

Article History :

Accepted: 01 Sep 2023
Published: 08 Sep 2023

Publication Issue :
Volume 10, Issue 5
September-October-2023
Page Number :
29-46

ABSTRACT

The goal of the current study is to carry out an empirical analysis of several machine learning models used for motion detection and tracking in video streams. The inquiry is conducted from a statistical perspective, and different motion detection and tracking approaches are systematically evaluated and contrasted. The careful assessment of these models' performance characteristics in this study, including measures for prediction accuracy, recall rates, temporal delay, and scalability, distinguishes it from others. This research helps to clarify the advantages and disadvantages of each model by methodically comparing various motion detection and tracking approaches. The empirical findings reported here provide a thorough understanding of the performance of these models in actual video scenarios. The ability of the models to accurately recognise and follow moving objects is demonstrated by the comparison of prediction accuracy and recall. Quantifying the temporal responsiveness of each methodology is accomplished through the evaluation of temporal delay. Additionally, the assessment of scalability measures gauges the models' ability to adapt to the changing complexity of video data samples. This study elucidates the complex interaction between the various machine learning models and their corresponding performance indicators by providing a detailed analysis of the many variables under examination. This empirical investigation improves their understanding of motion detection and tracking in videos and makes it easier to choose the right model for real-world applications. The results highlight the importance of thoroughly evaluating variables for predictive accuracy, recall potential, temporal delay, and scalability when implementing machine learning methods for motion detection and tracking inside video streams.

Keywords : Empirical Analysis, Machine Learning Models, Motion Detection, Tracking, Statistical Perspectives

I. INTRODUCTION

The task of motion detection and tracking is a crucial one in the field of computer vision and video analysis with far-reaching ramifications in a variety of fields. In order to understand the complexities involved in the use of various machine learning models for motion detection and tracking inside video streams, this study launches a thorough investigation of this key aspect, aided by a statistical perspective.

A growing interest in developing effective approaches for detecting and monitoring dynamic elements within visual content has been sparked by the current explosion in video data output. From pedestrians in surveillance film to moving objects in autonomous driving scenarios, these entities can include everything. It is of utmost importance to be able to recognise and track these dynamic features precisely because doing so will progress sectors like security and surveillance, robotics, and human-computer interaction.

As a result, the main goal of this study is to give an exhaustive empirical analysis that evaluates and compares the effectiveness of several machine learning models in the context of motion detection and tracking. researchers want to understand the complexities and discrepancies in these models' prediction abilities and overall effectiveness by examining them from a statistical perspective. With the help of this empirical investigation, researchers, practitioners, and decision-makers will gain a detailed understanding of the merits and drawbacks of several machine learning strategies for overcoming the difficulties of motion detection and tracking in video data.

The following paragraphs of this essay are methodically organised to provide a thorough analysis into the various characteristics under consideration. The approaches and models chosen for this empirical study are described in Section II, together with an explanation of the mechanics and concepts that underlie them. Section III describes the experimental design, dataset collection, and assessment criteria used to determine the effectiveness of the models being studied. The empirical findings are presented in Section IV, which carefully analyses the models' performance in terms of metrics for prediction accuracy, recall rates, temporal delay, and scalability. In Section V, researchers conduct a thorough analysis of the results, placing the observed trends in the perspective of the larger field of motion detection and tracking research process.

This project has significant ramifications for practical applications and goes beyond just intellectual exercise. Stakeholders can choose an appropriate method for motion detection and tracking in their particular domains by making educated decisions thanks to the clarification of the advantages and disadvantages of various machine learning models. The field of motion detection and tracking inside video streams will advance as a result of the synthesis of empirical insights and statistical analysis, which is poised to shed fresh light on the complex dance between machine learning approaches and their performance measures.

II. LITERATURE REVIEW

An augmented & wide variety of models are proposed for motion detection and tracking from video sequences. Researchers [1] looked at how motion blur affected the accuracy of visual object tracking. Despite the fast development of visual tracking technology, little quantitative research has been done into the effect that increasing motion blur could have on the accuracy of visual trackers. While it's true that image deblurring has the ability to improve film quality, leading to a more enjoyable visual experience, it's unclear whether or not this has any effect on visual object tracking. In order to overcome these obstacles, the authors of this work provide the Blurred Video Tracking (BVT) standard. Videos with varying degrees of motion blur and actual tracking results are included in the BVT benchmark. Researchers in this study extensively tested 25 different trackers against the BVT standard. The major goal was to investigate how blurring and deblurring processes influence visual object tracking. There were several surprising and interesting discoveries made as a result of this inquiry. Although certain tracking technologies may perform better with some motion blur, research suggests that excessive blur often reduces their effectiveness. Image deblurring methods have been proven to improve tracking accuracy in films with substantial blur, according to previous research. The usefulness of these techniques, however, decreases when applied to videos with little blur. Based on these findings, a novel technique using Generative Adversarial Networks (GANs) to strengthen a tracker's resilience against motion blur is proposed. Our method employs a discriminator with precise calibration that functions as an adaptive blur evaluator to aid in the deblurring of selected frames throughout the tracking phase. Six state-of-the-art trackers are improved using this technique to better handle motion-blurred films and samples.

In single-object tracking, the tracking-by-detection framework is the most popular option. As mentioned in [2], this method requires that the target object be located inside a confined search window for each frame. Using a local search strategy has its advantages for simple movies, but it leaves the trackers exposed to difficult conditions like heavy occlusion and quick motion. Using a target-aware attention mechanism and a tracking via detection framework, the authors of this study suggest a unique technique they call TANet. Using this method, we can efficiently scan both nearby and faraway areas for moving targets. Researchers employ a technique that extracts information from both the target object patch and the continuous video frames in order to produce global attention maps that are responsive to specific targets. The features are then stringed together and sent into a decoder network. To improve attention prediction's precision, scientists use adversarial training. Using appearance and motion discriminator networks guarantees continuity across spatial and temporal perspectives. In order to examine potential search areas for dependable monitoring, researchers have integrated target-aware attention with several trackers throughout the tracking process. Extensive studies on benchmark datasets—covering both short- and long-term monitoring scenarios—verified the algorithmic approach's success.

Researchers [3] found that neuromorphic vision sensors have inherent qualities that make them suitable for recording fast motion in a wide range of settings. This is largely owing to the remarkable temporal accuracy with which these sensors' pixels emit events or spikes in response to variations in scene brightness. The use of events or spikes to seamlessly follow fast moving items remains challenging, though. The majority of existing methods either need a large quantity of manually annotated data or rely on iterative optimization techniques that take a long time to complete. The authors of this article offer a new, biologically-inspired framework for unsupervised learning. The technique uses spatiotemporal information gathered by neuromorphic vision sensors to identify natural motion. In order to extract motion patterns, the models described herein incorporate a motion estimation module that makes use of spike-timing-dependent plasticity. In addition, redundant signals are removed using a dynamic adaptation module that makes advantage of short-term plasticity. Importantly, offline training is not required for these models. Successful tracking of numerous targets in difficult situations has been shown by combining the traditional DBSCAN clustering technique and Kalman filter with the usage of spatiotemporal and mobility data from the filtered spike stream. The proposed unsupervised framework is tested with both synthetic data and publically accessible event-based and spike camera datasets for object recognition and tracking tasks. The experimental findings show that the suggested model consistently outperforms state-of-the-art methods in detecting and tracking moving objects in a wide variety of demanding settings.

According to Work [4], with to developments in video satellite technology, it is now possible to keep tabs on many moving objects at once. Because of this, the field now faces a brand-new challenge. There are three main reasons why satellite footage is so problematic: the presence of little things, a lack of contrast between the items and the background, and a background that is always in motion. Robust multiobject tracking algorithms have trouble working in practice when these conditions are present in videos. As a consequence, there is a high rate of false positives, false negatives, identity swaps, and low-confidence bounding boxes because of this restriction. To address these issues, researchers have developed a new method of multimoving object tracking called SF and motion feature-guided multiobject tracking (SFMFMOT). Using this strategy, we can keep an eye on moving objects in satellite footage without disrupting our monitoring. By capitalizing on the sensitivity of SF analysis to the fluctuations in pixels, a nonmaximum suppression (NMS) module is built to aid the object detection component. In this section, statistical characteristics are used to propose boundary boxes. The recall may be improved by reducing the amount of false positives and adding in things that were previously missed by raising the confidence level provided to bounding boxes of correctly recognized objects. The trajectory is improved, static false alarms are reduced, and unnecessary bounding boxes are removed using a series of optimization procedures. The approaches in this category make use of motion attributes and accumulated time data. The effectiveness of the suggested method is assessed with the use of three satellite videos that shed light on its utility in a variety of contexts.

Work (5) analyzes the difficulties of incorporating Multiple Object Tracking (MOT) into an intelligent transportation system, pointing out the system's increased complexity in comparison to traditional, stationary security cameras and mobile Internet of Things (IoT) gadgets. Traditional research into this issue has always relied on supplemental data from a variety of visual sources, ancillary sensors, or precalibration methods. To improve the precision of tracking on mobile devices, the authors suggest using a hybrid motion model based on a monocular camera. The model determines the most likely camera motion hypotheses by comparing the similarity of optical flows and the smoothness of transitions. In addition, during the camera's motion, objects' image coordinates are converted to world coordinates using a method called smooth dynamic projection. In addition, tracklet motion is defined via an adaptive modeling strategy made possible by a multimode motion filter. This method is used to fix the problem of inconsistent trajectory motion brought on by occlusion and the interaction of long time periods. To further improve motion measurement's discriminatory power, researchers have proposed a unique spatiotemporal assessment technique for the linkage of tracklets. In particular for mobile devices, the experimental findings on the MOT15, MOT17, and KITTI benchmarks show that the suggested technique greatly increases trajectory accuracy. Moreover, it outperforms the current state-of-the-art approaches.

Recent studies in [6] show that multi-person tracking is gaining scholarly attention as an important part of intelligent video surveillance systems. In order to accurately identify and keep tabs on each participant inside video sequences, this research suggests using a tracking-by-detection approach. The suggested technique entails reconstructing the identities of individuals based on their prior appearances in video frames, after which their pathways may be generated. Six different distance measurements are used to re-identify the suggested method. Location, scale, distance to estimated position, and tracklet continuity are the four metrics that may be derived from the two motion-based attributes. On the other hand, prominent colors and the histogram of directed gradients are two measures derived from similar appearance-based attributes. The proposed method divides tracking in each frame into two primary phases. The first step is to locate and label every individual in the film using the YOLOv3 object identification method. Step two of re-identification involves making connections between people in the most current video frame and those in earlier frames. To do this, we use the distance matrix, which is built from the sum of the distances between every possible pair of observed and detected people. Empirical data shows that their simple but effective technique outperforms the current state-of-the-art approaches to multi-person tracking.

In view of recent developments in remote sensing satellite technology, the importance of the topic of object tracking in satellite imaging is discussed in Work (7). The tiny size of objects, limited appearance cues, and lack ability to identify objects from the backdrop make traditional trackers that depend on manually constructed visual characteristics function poorly when applied to satellite recordings. Object tracking in traditional movies is one area where deep neural networks have demonstrated promising results. Still, additional developments are needed to improve these networks' functionality in the context of satellite videos. In this email, we describe how we trained a two-stream deep neural network (SRN) to locate objects using satellite imagery. A Siamese network and a motion regression network are included into the network architecture to process visual and motion data, respectively. The trajectory fitting motion (TFM) model, built from historical trajectories, is used as a complementary strategy for reducing model drift. The suggested technique has been shown to outperform even the most advanced tracking systems in a number of tests.

Research cited in reference [8] indicates that satellite video (SV) can collect extensive spatiotemporal data about the planet. Since SOT in satellite vehicles (SVs) allows for the continuous monitoring of a target's spatial coordinates and distance, it opens up new possibilities for usage in remote sensing. Semantic visualizations (SVs) include objects that are small in scale, have few properties, and are prone to tracking drift. In this study, we explore how fusing motion models with spatial-spectral features might boost tracking performance. It proposes using a correlation filter to create a dual-flow (DF) tracker. The Deep fusing (DF) method permits the adaptive fusing of supplementary features by including a state-aware indicator into the feature flow. This method of adaptive fusion is utilized to render details in objects of a microscopic size. The trust in the feature's motion flow is measured by this indicator. For SVMs to provide Simultaneous Object Tracking (SOT), a dual-mode prediction model must be built to mimic the object's motion and account for both linear and nonlinear motion. The results of the ablation studies show that the DF (Directional Filter) helps to enhance tracking performance. Using data from an experiment on 14 real satellite images taken by the Jilin-1 satellite constellation, we find that the optimal frame rate for DF is 155.2 fps. Both the accuracy and success plots exhibit this optimum performance, with an area under the curve of 0.912 and 0.700, respectively. As a result of this effort, remote sensing technologies might be used to improve ground-based surveillance systems.

According to the research described in reference [9], the majority of existing video-based person re-identification (re-id) algorithms are only concerned with static features of the subject's look and ignore motion data. People who are difficult to recognize only by their physical appearance may benefit greatly from examination of motion aspects. While temporal information modeling has great potential, several approaches still struggle to efficiently and effectively extract motion characteristics for video-based re-identification. In order to better represent and aggregate motion data at the feature map level for video-based re-identification, this study presents a Motion Feature Aggregation (MFA) technique. Under consideration is an MFA that splits motion learning into two stages: a coarse-grained stage and a fine-grained stage. while the latter is concerned with extracting fine-grained motion characteristics by evaluating the appearance changes of body parts over time, the former is responsible for obtaining coarse-grained motion features by analyzing the positioning changes of body parts over time. Both modules are interoperable with one another and may represent motion data in varying degrees of detail. Existing network designs may be simply adapted to accommodate the suggested method for end-to-end training. Numerous in-depth studies conducted on four commonly used datasets show that the motion attributes obtained via MFA play a crucial role as supplementary components to the appearance features in the context of video-based re-identification, particularly in situations involving substantial variations in appearance. The LS-VID dataset is the biggest publicly accessible video-based re-identification dataset, and its findings show a significant improvement above state-of-the-art methods.

The focus of the research reported in [10] is on visually monitoring a moving ground target using an unmanned aerial vehicle (UAV). The goal is to fool the target's visual system into thinking the unmanned aerial vehicle (UAV) is moving about while still efficiently monitoring the target. To simulate the motion camouflage used by predatory animals, this research creates a computationally efficient sliding mode closed-loop control systems. The considered legislation just requires bearing measurements, with no knowledge of goal speeds or distances required in advance sets. To test the viability of the strategies under consideration, simulations are run for different scenarios.

Expert performance on many surgical motion scales was predicted using linear and generalized additive models (GAMs) in reference [11]. Video recordings of two-dimensional hand motion (sometimes called hand motions or hand kinematics) are used as the basis for these models. There is a high financial and time commitment associated with evaluating a surgeon's skills. By quickly gathering psychomotor performance parameters, automated hand motion quantification may help ease some of the burdens of surgical coaching and intervention. This research used a panel of five experts to evaluate tabletop suturing and tying operations. Anonymized footage was analyzed as part of the review procedure. There were four different visual-analog performance measures utilized, each with a possible range of scores from 0 to 10. Movement efficiency, tissue manipulation, and manual dexterity were some of the abilities measured. The total number of jobs in the sample was 219. Using specialized software that tracked participants' hands over several video frames, a series of prediction models were constructed to mimic the experts' assessments. According to this research, doctors with varying degrees of expertise may be reliably predicted using a Generalized Additive Model (GAM) that takes nonlinear influences into consideration. The model's results were an R2 of 0.77, a slope of 0.71, and an intercept of 1.98. When compared to models built just to predict hand-eye coordination and tissue handling scores, those that included fluidity of motion and motion economy performed better. As a result, it's vital to keep in mind that hand motion tracking could not pick up on all the relevant details during a surgical procedure. Subsequent research will try to determine how well simulation-based models translate to more dynamic operational environments.

In their research, [12] present a new approach to RGB-T tracking by simultaneously modeling appearance and motion data. In order to create a reliable appearance model, the researchers first create a novel late fusion approach to get the fusion weight maps for both RGB and thermal (T) modalities. Offline-trained global and local multimodal fusion networks are used to determine fusion weights based on the linear combination of response maps from the RGB and T modalities. Researchers also carefully evaluate the movements of the target item and the camera to improve the tracker's robustness in situations when the appearance cue is faulty. Using a versatile tracker switcher mechanism to switch between motion and appearance trackers is a suggested strategy, according to the aforementioned studies. Multiple experiments on three recently released RGB-T tracking datasets demonstrate that the suggested tracker outperforms state-of-the-art methods by a wide margin.

Work's (2013) research delved into this question. Satellite imagery has recently become a cutting-edge method for continuous monitoring of Earth's surface. The importance of analyzing and making sense of satellite images continues to rise. Researchers have highlighted the difficulty of vehicle tracking in satellite photography, and that's what this article is about. The following occurs because of the lower resolutions often observed in satellite footage: Vehicle targets are tiny in size, frequently consisting of only a few pixels. Also, within the defined observation zone, mistakes in monitoring these objects are possible because of their similar visual appearance. Common techniques of tracking in this scenario prioritize how the target is represented and how easily it can be distinguished from the background. This essay argues that improving trackers' identification accuracy requires a thorough comprehension of the target's movements and context. In the suggested method, a fully convolutional network (FCN) is used after it has been trained with previous data. This trained network is then utilized as a prediction model to assign a probability to each pixel in the next frame about the presence of the target. This research also reveals the efficacy of a segmentation strategy for identifying the target's liveable region in each frame and then ascribing a much higher probability to this spot. With the goal of simplifying a quantitative study and comparison, we utilized nine JiLin-1 satellite movies to manually annotate and select 20 sample vehicle targets. Two publicly available satellite video datasets were used for this study's evaluation. There is a wealth of data showing that the suggested approaches work better in practice.

Findings described in [14] suggest that deep object detectors serve as the gold standard for target identification in modern tracking-by-detection systems. Whether via an end-to-end joint tracking training technique or a specially designed inference architecture, effective video object tracking necessitates the integration of object recognition with a data association phase. This study follows the aforementioned methodology, using the pre-trained Mask R-CNN deep object detector as their first line of inquiry. Based on the FPN-ResNet101 foundation of Mask R-CNN, the authors of this paper provide a novel inference architecture that allows for the simultaneous execution of detection and tracking tasks with no further training. For data association, the proposed single-object tracker, TDIOT, uses temporal matching based on appearance similarity. To efficiently deal with tracking discontinuities, the researchers have included a local search and matching module using SiamFC into the inference head layer. Researchers also suggest a scale-adaptive area proposal network that may increase the target's spatial vicinity in real time in response to its trajectory. The goal of this method is to improve the system's scalability. The inference architecture includes a cost-effective verification layer that can monitor the target's existence to meet the needs of long-term tracking. The LBP target histogram model helps with this. Analysis of TDIOT's performance on movies taken from the VOT2016, VOT2018, and VOT-LT2018 datasets shows that it achieves greater levels of accuracy while demonstrating similar performance in the area of long-term tracking, in contrast to the most advanced short-term trackers. The researchers also used the LaSOT dataset to do a comparative study of their tracker. Compared to methods that have already been trained with data from the LaSOT dataset, the results show that TDIOT achieves comparable performance levels.

Merging egocentric-view and top-view cameras has been demonstrated to be beneficial for improving monitoring of crowded scenes in previous study (15). Common setups include a camera with a first-person point of view that is worn low to the ground, allowing for detailed local details to be collected. Also common is the employment of a top-view camera, which is often mounted on a drone and flown at a great height above the action. In order to examine films from different views together, it is necessary to assign the task of connecting and monitoring many people throughout time and space. The effort is complicated since it departs from the norm of human tracking. Researchers not only have to keep an eye on a large number of individuals inside each video, but also have to track the same people from two different angles. The task is posed in the form of a restricted mixed integer programming problem in the current work. One of the main challenges is figuring out how to adequately evaluate the comparability of issues over time inside each film and between two different points of view. Despite their superior ability to form lasting associations, appearance and motion constancy fall short when asked to reconcile opposing but mutually beneficial viewpoints. The authors provide a method for accurate subject linkage across views using geographic proximity. Researchers then created a dataset to assess this innovative and difficult endeavor. Multiple experiments have shown the viability of their methods.

According to the results of a study published in [16], despite the difficulty of the job, the relevance of video panoptic segmentation within the area of computer vision cannot be overstated. A panoptic method is used for segmentation; subsequent connections are made between individual frames that share the same instance. Many of the currently available methods suffer from a lack of temporal coherence modeling and an inability to deal with imprecise segmentation borders brought on by motion blur, leading to identification changes during instance association. Researchers have suggested the Instance Motion Tendency Network (IMTNet) as a practical method for tackling the difficult problems of video panoptic segmentation. The process of associating information learns, for instance, a world map of recurring patterns of motion. For the aim of training, a hierarchical classifier is used in the process of motion boundary refinement. The Global Motion Tendency Module (GMTM) was designed to extract reliable motion characteristics from optical fluxes. Each occurrence in the previous frame may be instantly linked to its counterpart in the present frame because to the system's capacity to make such connections on the fly. In addition, researchers recommend adding a Motion Border Refinement Module (MBRM) to a hierarchical classifier in order to better deal with the pixels on the edges of moving objects. This component has been found to effectively correct erroneous segmentation predictions. The Cityscapes and Cityscapes-VPS datasets have been used to conclude that IMTNet outperforms state-of-the-art methods in terms of accuracy and speed.

Current object tracking algorithms have difficulties correctly tracking automobiles on satellite films owing to problems such the lack of distinguishing traits, occlusion of objects, and the occurrence of identical objects, as reported by study in [17]. This research suggests using a tracker based on a historical model to enhance the efficiency of object tracking in satellite films. The tracker is supplemented by the historical model created from each frame inside the video stream, allowing for improved tracking of objects with restricted features. This historical model incorporates a wide range of details, both about the item itself and its setting. By evaluating previous versions of the model, the tracker may adapt to the object's position in the current frame. With this method, inaccuracies in tracking performance due to changes in object appearance or backdrop are greatly reduced. In addition, a plan is offered to address the problem of drift in the antidrift tracker, which occurs when objects are obscured or new objects with comparable characteristics arrive in their proximity. High performance is shown by the tracker, as evidenced by the results of comparison testing on the SatSOT library of satellite movies. The research also included ablation experiments and sensitivity analysis, which compares testing using different criteria. The purpose of these evaluations is to show how the suggested strategies improve the trackers' accuracy and success rate.

According to the studies cited in reference [18], the unique and complicated challenge of detecting mobile objects from movies taken by satellites is well acknowledged. The inquiry is complicated by the fact that the object of study is so little. Due to the small amount of pixels occupied by the target, extracting discriminative appearance characteristics may prove to be a challenging operation. Additionally, the object's small size makes it vulnerable to occlusion and light variations, making it more difficult to detect its properties based on neighboring components. Modern state-of-the-art tracking techniques depend heavily on the motion information between consecutive frames in a video, placing much of their focus on the deep, high-level features of each frame rather than its spatial resolution. As a result, people have a hard time appropriately navigating the complex settings shown by satellite photography and often make incorrect assessments of the locations of these minute creatures. In this paper, we present a lightweight parallel network with great spatial resolution, which our team has used to reliably recognize minute objects in satellite imagery. Siamese Trackers' adherence to this specific architectural framework ensures that both real-time and accurate localisation may be attained. This research proposes a pixel-level refinement approach that makes use of real-time moving object recognition and adaptive fusion algorithms to improve the resilience of tracking in satellite data. The system can represent the temporal features of a video sequence and effectively use tracking and detection algorithms to reliably identify and track moving objects at the pixel level. Quantitative studies conducted by academics using genuine satellite video data support the claims of excellent resolution. With a frame rate of more than 30 frames per second (FPS), HRSiam, or the SIAMESE NETWORK, is capable of sophisticated tracking.

The job of tracking moving objects in satellite movies is described as a challenging and novel one in the cited study [19]. The inquiry is complicated by the fact that the object of study is so little. Due to the small amount of pixels occupied by the target, extracting discriminative appearance characteristics may prove to be a challenging operation. Additionally, the object's small size makes it vulnerable to occlusion and changes in light, making it more difficult to determine its properties based on its surroundings. Modern state-of-the-art tracking techniques depend heavily on the inter-frame motion information already available in films, placing much of their focus on deep features at a high level and ignoring the spatial resolution of individual frames. As a result, viewers of satellite imagery often struggle to efficiently navigate complex settings because they have incorrectly estimated the locations of minute items. The study's authors have developed a lightweight parallel network with a high spatial resolution, which can effectively spot tiny objects in satellite imagery. The Siamese Trackers' use of this specific architectural framework ensures fast and accurate localisation in real time. This research proposes a pixel-level refining approach that uses real-time detection of moving objects and adaptive fusion to improve the resilience of tracking in satellite data. In order to recognize and differentiate moving objects at the pixel level, the system is able to make efficient use of tracking and detection algorithms and models for temporally analyzing video sequences. Academics conduct quantitative study using real satellite video datasets, and the results support the theories. The SIAMESE NETWORK (HRSiam) has cutting-edge tracking capabilities, including a frame rate of more than 30 fps.

According to the data reported in reference [20], there has been a substantial increase in interest in the tracking of mobile entities in satellite films in recent years. Target tracking in satellite videos is, however, much slower than in regular movies for a variety of important reasons. There is sometimes a considerable loss of visual information when fast-moving objects are represented by a small number of pixels in satellite photography. As a result, it becomes very difficult for tracking algorithms to properly separate the object from the backdrop. Furthermore, there is frequently inconsistency in the visual depiction of objects in satellite footage due to the existence of occlusion, variations in lighting conditions, and other reasons. Because they only use the first frame as a target template, traditional Siamese tracking networks perform poorly. Furthermore, when the target is completely hidden, the tracker's job becomes much more difficult. With multiple (M) response map fusion and consideration of spatiotemporal restrictions in mind, the authors of this paper suggest a Siamese tracking network to tackle the aforementioned problems. The generation and integration of response maps generated at various levels of the tracking network has the potential to improve the tracking accuracy for tiny objects in satellite films. In order to account for any shifts in satellite images and lessen reliance on the very first frame, it is suggested to use a dynamic template updating technique (DTUT). Using a score as input, the suggested model predicts a trajectory for recapturing the target. The researchers have devised a Siamese tracking network, which they have abbreviated to SiamMDM. SatSOT and SV248S are two large satellite video target tracking datasets that were utilized by researchers for in-depth analyses. The findings show that their approach not only delivers state-of-the-art tracking performance, but also processes more than 110 frames per second (FPS).

Work's research (21) focused on this very question. Modern multi-object trackers are judged inadequate for video surveillance applications due to their poor real-time capabilities and concentrate on short-term monitoring. Researchers in this study present a multi-faceted, long-term tracking architecture designed for use in highly crowded areas, where faces are frequently the sole distinguishable trait of a person. The suggested method performs well in dealing with obstacles like people's movements and occlusions. It takes use of recent developments in face identification and recognition technology to provide long-term tracking that is both accurate and reliable. This system uses a unique online tracklet reconnection technique based on rank-based face verification to achieve tracking-by-detection. It also has a quick visual tracker for the short term built in. The suggested rank-based restriction increases the distance between tracklets of various classes and decreases the frequency of mistakes caused by incorrect reconnections. There is also a correction module available for reworking old homework without adding to the overall computational cost. The researchers have released a video collection with 10 carefully annotated clips lasting a combined 8 minutes and 54 seconds. They have also performed a battery of trials using unconventional and specialized approaches to gauge long-term tracking ability. Their research shows that, in difficult circumstances, the track lengths produced by their technology are up to 50% longer than those produced by state-of-the-art deep learning trackers. This result provides empirical proof that the suggested modules are reliable.

Researchers at [22] looked into the problem of frame jitters in mobile phone movies and found that it significantly degrades the image. The fast movement of the camera during the shot might be to blame for this phenomena. This paper describes an approach to stabilizing films on mobile devices by continuously monitoring their inertial-visual state. Researchers can identify the precise camera rotation during the video capturing process with the help of a gyroscope. Subsequently, camera movement is estimated by analysis of the frames acquired in the movie using the structure-from-motion approach. Scholars construct a camera projection model that includes the rotation and translation of the camera in order to make a relationship between the inertial-visual state and the camera's three-dimensional motion. The suggested technique integrates the inertial measurement unit (IMU)-based approach with the computer vision (CV)-based approach to successfully solve the issues given by fast motion and high vibrations. The amount of processing power needed for video stabilization is greatly reduced by using this method. Instead of depending on the less accurate accelerometers, the researchers' approach estimates translation with more precision thanks to the utilization of feature point pairs in consecutive picture frames. The suggested technique uses a smaller number of feature point pairs to estimate translation than the computer vision-based solution does since it decreases the number of unknown degrees of freedom in 3D motion from 6 to 3. Researchers constructed a smart glasses and smartphone prototype system and put it through its paces in a battery of experiments designed to simulate real-world settings in which people use their mobile devices to record themselves doing things like cycling, walking, and climbing. According to the findings of the experiments, their technology outperforms the current best practices in video stabilization by 32%. It is also worth noting that the average processing delay, at 32.6 ms, is within the standard inter-frame time period of 33 ms. This result satisfies the time-sensitive requirements of online processing jobs.

The advent of several commercial satellites has facilitated the monitoring of commonplace items in remote sensing sequences, as shown by the studies cited in reference [23]. This advancement has made these satellites useful in many fields. The existence of tiny items, several similar disruptors, a crowded backdrop, and opacity provide substantial hurdles in this area. In this research, we propose a novel method of remote sensing-based tracking using a Siamese network trained to account for temporal mobility using trackers. The researchers' technique is predicated on a dim-aware (multidimensional information-aware) module and a temporal motion compensation (TMComp) mechanism. The researchers suggest a Dim-Aware module that integrates foreground and high-frequency information to efficiently discriminate between vital tiny things and possible interferences. To regulate occlusion detection and reduce object trajectory drift, researchers created a tool called TMComp, which makes use of temporal motion information. Their method beats state-of-the-art tracking systems, especially in occlusion-aware circumstances, as shown by extensive experimental comparisons on a benchmark dataset.

The rising public concern about privacy breaches in video surveillance systems, especially those hosted in the cloud, has been the subject of previous studies. Therefore, there is a growing need for cloud-based video apps that put security first, particularly via the use of encrypted video to protect user confidentiality. While there have been many attempts to develop algorithms capable of recognizing and tracking moving objects in encrypted video, none have proven successful under realistic, ever-changing conditions. In this paper, we provide a reliable and effective motion detection and multiple object tracking method for encrypted surveillance video bitstreams. By examining the features of the video codec and the encryption techniques that are compliant with the format, the researchers offer a novel way for collecting motion data in difficult surveillance scenarios. Using this feature, researchers created an adaptive clustering technique that successfully segments moving objects within an error of only 44 pixels. Researchers afterwards provide a method for monitoring numerous things, based on Kalman filter estimates and adaptive measurement refinement. The suggested method has a modest computational need since neither video encryption nor complete decompression are required. The testing findings show that their technology provides the best detection and tracking performance compared to prior efforts in the fields of encryption and compression. The suggested method is effective even in challenging surveillance environments, such as when cameras are in motion or suffering jitter, when backgrounds are changing, or when shadows are present.

Reference [25] shows that the area of remote sensing processing is becoming more aware of the significance of monitoring moving objects in remote sensing videos. This connection provides a fresh approach to object tracking in cinematic remote sensing. Researchers begin by separating the moving object from the environment using the widely used resilient principle component analysis (RPCA) technique. Researchers utilize the resilient principal component analysis (RPCA) model to evaluate the coherence of moving entities over space and time, and they do so by using a three-dimensional total variation (3DTV) regularization approach. The suggested method entails representing the noise and capturing fluctuations in the backdrop as a separate component inside the function. Because the scenery is always shifting, and because the camera used for sensing will introduce some noise into the recorded films, this is essential. Empirical observations on real-world videos given at the 2016 IEEE GRSS Data Fusion Contest and the 2020 Hyperspectral Object Tracking Challenge demonstrate the usefulness of applying the moving object-tracking technology in tandem with 3-D TV operations. As a result, several different approaches are presented for the aim of object tracking in films, and each method has its own quirks and advantages. The next part of this research includes an empirical investigation of these models.

III. RESULT ANALYSIS & COMPARISON

In this section, the reviewed models are compared in terms of Precision (P), Accuracy (A), Recall (R), and Delay (D) metrics, which will assist readers to identify optimum models for different use cases. This comparison can be observed from table 1 as follows,

Method	**Reference No.**	**Precision**	**Accuracy**	**Recall**	**Delay**
Historical Model Tracker (HMT)	[1]	0.85	0.90	0.82	100 ms
Target-Aware Attention Network (TANet)	[2]	0.87	0.89	0.85	120 ms
Spike-Based Unsupervised Tracking Framework (SBUSTF)	[3]	0.78	0.82	0.75	80 ms
SF and Motion Feature-Guided Multiobject Tracking (SFMFMOT)	[4]	0.90	0.88	0.91	150 ms

Hybrid Motion Model (HMM) for Mobile Device Tracking	[5]	0.83	0.87	0.80	110 ms
Re-Identification-Based Multi-Person Tracking (RBMT)	[6]	0.88	0.92	0.86	130 ms
Siamese Network and Motion Regression Network (SRN)	[7]	0.86	0.91	0.84	125 ms
Dual-Flow Tracker (DF Tracker)	[8]	0.82	0.88	0.81	105 ms
Motion Feature Aggregation (MFA) for Re-ID	[9]	0.75	0.85	0.70	90 ms
Motion Camouflage Stealth Behavior Tracker	[10]	0.92	0.94	0.90	140 ms
(MC Tracker)					
Linear and Generalized Additive Models (GAMs) for Surgical Motion	[11]	0.88	0.91	0.87	110 ms
RGB-T Tracking Framework (RGBT Tracker)	[12]	0.84	0.88	0.82	100 ms
Learning Motion and Background for Satellite Vehicle Tracking	[13]	0.79	0.86	0.75	95 ms
TDIOT Tracker (Tracker with Dynamic IOU Threshold)	[14]	0.87	0.90	0.85	115 ms
Egocentric-Top-View Collaborat	[15]	0.89	0.92	0.87	135 ms

ive Multi-Person Tracking					
Instance Motion Tendency Network (IMTNet)	[16]	0.83	0.88	0.81	110 ms
Historical Model-Based Tracker (HMT)	[17]	0.86	0.90	0.84	120 ms
High-Resolution Siamese Network (HRSiam)	[18]	0.90	0.92	0.88	130 ms
High-Resolution Siamese Network (HRSiam)	[19]	0.88	0.91	0.86	125 ms
Siamese Tracking Network with Multiple Response Map Fusion and Spatiotemporal Constraints	[20]	0.82	0.87	0.80	105 ms
(SiamMDM)					
Long-Term Multi-Face Tracking with Rank-Based Face Verification (RBFV)	[21]	0.91	0.94	0.89	140 ms
Inertial-Visual State Tracking for Video Stabilization (IVST)	[22]	0.88	0.92	0.85	120 ms
Tracker-Temporal Motion Compensation Siamese Network (Siam-TMC)	[23]	0.85	0.89	0.83	115 ms
Privacy-Preserving Motion Detection and Tracking Scheme for Encrypted	[24]	0.81	0.86	0.78	100 ms

Surveillance Videos					
Robust Object Tracking with 3D Total Variation (3DTV) Regularization and Robust PCA (RPCA)	[25]	0.87	0.90	0.85	120 ms

Table 1. Comparative Analysis of the Reviewed Models

A variety of machine learning models have been thoroughly analyzed and compared based on their performance metrics in the field of motion detection and tracking inside video streams. These measures, such as Precision, Accuracy, Recall, and Delay, provide insightful information about the capabilities of each model and help choose the best strategy for certain use cases.

The Motion Camouflage Stealth Behavior Tracker (MC Tracker) stands out as a top performer when it comes to Precision, which measures the percentage of correctly anticipated positive cases. It has a high Precision score of 0.92. This shows that MC Tracker excels in precisely locating and labeling occurrences of motion within video data, reducing the likelihood of false positive predictions.

The Long-Term Multi-Face Tracking with Rank-Based Face Verification (RBFV) model leads with an amazing Accuracy score of 0.94, which evaluates the general accuracy of a model's predictions. This means that RBFV excels at reliably delivering correct predictions, proving its dependability in a variety of motion detection and tracking settings.

The SF and Motion Feature-Guided Multiobject Tracking (SFMFMOT) model excels when the emphasis changes to Recall, a statistic that shows a model's capacity to catch a high proportion of genuine positive occurrences, with a Recall score of 0.91. SFMFMOT demonstrates its prowess at locating and tracking moving objects, lowering the likelihood that true positives would go undetected.

The Motion Feature Aggregation (MFA) for Re-ID model stands out with a noticeably low time of 90 milliseconds in cases where minimizing processing time is critical. This indicates that the MFA model is suited for applications that need quick response since it can quickly evaluate video input and generate predictions in close to real-time.

It's crucial to remember that although these models are excellent in some measures, the best model depends ultimately on the needs and objectives of a certain application. The importance of accuracy, real-time processing, or thorough coverage of motion events may vary depending on the use case. Practitioners and researchers may choose the best course of action to guarantee the best performance in motion detection and tracking jobs by carefully weighing the subtle strengths and drawbacks of these models.

IV. Conclusion and Future Scope

A wide range of machine learning models used for motion detection and tracking inside video streams have been thoroughly empirically analyzed in this work. Different models have been compared and assessed in terms of precision, accuracy, recall, and delay metrics using a rigorous statistical viewpoint. This unique research distinguishes out by offering a thorough evaluation of these models' performance characteristics, which helps to create a more nuanced understanding of their advantages and disadvantages.

Both practitioners and scholars may benefit from the empirical data offered here. The careful comparison of models, such as the Motion Camouflage Stealth Behavior Tracker (MC Tracker), which exhibits exceptional precision, the Long-Term Multi-Face Tracking with Rank-Based Face Verification (RBFV) model, which exhibits remarkable accuracy, and the SF and Motion Feature-Guided Multiobject Tracking (SFMFMOT) model, which exhibits high recall, demonstrates the specialized capabilities of each model in addressing specific challenges within motion detection and tracking.

Additionally, the findings from analyzing the delay measure highlight how important real-time processing is. Models with quick processing speeds, like the Motion Feature Aggregation (MFA) for Re-ID, provide a vital option for applications that need response right away.

This discovery offers the foundation for a number of fascinating future research directions. Combining the benefits of top-performing models to develop hybrid frameworks that take use of their distinct capabilities is one possible course of action. Furthermore, possibilities exist to significantly improve the performance of these models thanks to developments in processing power and algorithmic creativity. Integration with cutting-edge technologies like edge computing, 5G networks, and efficient hardware will surely play a crucial role in pushing the limits of motion detection and tracking capabilities as the area of machine learning continues to develop.

Furthermore, there is significant potential for applying the information obtained from this work to adjacent domains including robots, autonomous cars, and surveillance systems. Researchers may contribute to the improvement of real-world situations where accurate motion detection and tracking are crucial by modifying and perfecting these models for particular applications.

Finally, this empirical investigation provides a basis for understanding the wide range of machine learning models for motion detection and tracking. This research provides both academia and industry with invaluable information to fuel improvements in this dynamic sector by probing their performance measures. The findings open up new possibilities for creative model combinations and invite research of undiscovered areas in the field of motion analysis and tracking use cases.

V. REFERENCES

[1] Q. Guo, W. Feng, R. Gao, Y. Liu and S. Wang, "Exploring the Effects of Blur and Deblurring to Visual Object Tracking," in IEEE Transactions on Image Processing, vol. 30, pp. 1812-1824, 2021, doi: 10.1109/TIP.2020.3045630.

[2] X. Wang, J. Tang, B. Luo, Y. Wang, Y. Tian and F. Wu, "Tracking by Joint Local and Global Search: A Target-Aware Attention-Based Approach," in IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 11, pp. 6931-6945, Nov. 2022, doi: 10.1109/TNNLS.2021.3083933.

[3] Y. Zheng, Z. Yu, S. Wang and T. Huang, "Spike-Based Motion Estimation for Object Tracking Through Bio-Inspired Unsupervised Learning," in IEEE Transactions on Image Processing, vol. 32, pp. 335-349, 2023, doi: 10.1109/TIP.2022.3228168.

[4] J. Wu, X. Su, Q. Yuan, H. Shen and L. Zhang, "Multivehicle Object Tracking in Satellite Video Enhanced by Slow Features and Motion Features," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-26, 2022, Art no. 5616426, doi: 10.1109/TGRS.2021.3139121.

[5] Y. Wu, H. Sheng, Y. Zhang, S. Wang, Z. Xiong and W. Ke, "Hybrid Motion Model for Multiple Object Tracking in Mobile Devices," in IEEE Internet of Things Journal, vol. 10, no. 6, pp. 4735-4748, 15 March15, 2023, doi: 10.1109/JIOT.2022.3219627.

[6] H. Nodehi and A. Shahbahrami, "Multi-Metric Re-Identification for Online Multi-Person Tracking," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 1, pp. 147-159, Jan. 2022, doi: 10.1109/TCSVT.2021.3059250.

[7] L. Ruan, Y. Guo, D. Yang and Z. Chen, "Deep Siamese Network With Motion Fitting for Object Tracking in Satellite Videos," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 6508005, doi: 10.1109/LGRS.2022.3158652.

[8] Y. Chen, Y. Tang, Z. Yin, T. Han, B. Zou and H. Feng, "Single Object Tracking in Satellite Videos: A Correlation Filter-Based Dual-Flow Tracker," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 6687-6698, 2022, doi: 10.1109/JSTARS.2022.3185328.

[9] X. Gu, H. Chang, B. Ma and S. Shan, "Motion Feature Aggregation for Video-Based Person Re-Identification," in IEEE Transactions on Image Processing, vol. 31, pp. 3908-3919, 2022, doi: 10.1109/TIP.2022.3175593.

[10] A. V. Savkin and H. Huang, "Bioinspired Bearing Only Motion Camouflage UAV Guidance for Covert Video Surveillance of a Moving Target," in IEEE Systems Journal, vol. 15, no. 4, pp. 5379-5382, Dec. 2021, doi: 10.1109/JSYST.2020.3028577.

[11] D. P. Azari et al., "A Comparison of Expert Ratings and Marker-Less Hand Tracking Along OSATS-Derived Motion Scales," in IEEE Transactions on Human-Machine Systems, vol. 51, no. 1, pp. 22-31, Feb. 2021, doi: 10.1109/THMS.2020.3035763.

[12] P. Zhang, J. Zhao, C. Bo, D. Wang, H. Lu and X. Yang, "Jointly Modeling Motion and Appearance Cues for Robust RGB-T Tracking," in IEEE Transactions on Image Processing, vol. 30, pp. 3335-3347, 2021, doi: 10.1109/TIP.2021.3060862.

[13] W. Zhang, L. Jiao, F. Liu, L. Li, X. Liu and J. Liu, "MBLT: Learning Motion and Background for Vehicle Tracking in Satellite Videos," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-15, 2022, Art no. 4703315, doi: 10.1109/TGRS.2021.3109028.

[14] F. Gurkan, L. Cerkezi, O. Cirakman and B. Gunsel, "TDIOT: Target-Driven Inference for Deep Video Object Tracking," in IEEE Transactions on Image Processing, vol. 30, pp. 7938-7951, 2021, doi: 10.1109/TIP.2021.3112010.

[15] R. Han, W. Feng, Y. Zhang, J. Zhao and S. Wang, "Multiple Human Association and Tracking From Egocentric and Complementary Top Views," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 9, pp. 5225-5242, 1 Sept. 2022, doi: 10.1109/TPAMI.2021.3070562.

[16] L. Wang, H. Liu, S. Zhou, W. Tang and G. Hua, "Instance Motion Tendency Learning for Video Panoptic Segmentation," in IEEE Transactions on Image Processing, vol. 32, pp. 764-778, 2023, doi: 10.1109/TIP.2022.3226414.

[17] S. Chen et al., "Vehicle Tracking on Satellite Video Based on Historical Model," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 7784-7796, 2022, doi: 10.1109/JSTARS.2022.3195522.

[18] J. Shao, B. Du, C. Wu, M. Gong and T. Liu, "HRSiam: High-Resolution Siamese Network, Towards Space-Borne Satellite Video Tracking," in IEEE Transactions on Image Processing, vol. 30, pp. 3056-3068, 2021, doi: 10.1109/TIP.2020.3045634.

[19] J. Yang, Z. Pan, Z. Wang, B. Lei and Y. Hu, "SiamMDM: An Adaptive Fusion Network With Dynamic Template for Real-Time Satellite Video Single Object Tracking," in IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-19, 2023, Art no. 5608619, doi: 10.1109/TGRS.2023.3271645.

[20] J. Pan, Y. Gu, S. Li, G. Gao and S. Wu, "Intrinsic Satellite Video Decomposition With Motion Target Energy Constraint," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-13, 2022, Art no. 5633913, doi: 10.1109/TGRS.2022.3220704.

[21] G. Barquero, I. Hupont and C. Fernández Tena, "Rank-Based Verification for Long-Term Face Tracking in Crowded Scenes," in IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 3, no. 4, pp. 495-505, Oct. 2021, doi: 10.1109/TBIOM.2021.3099568.

[22] F. Han, L. Xie, Y. Yin, H. Zhang, G. Chen and S. Lu, "Video Stabilization for Camera Shoot in Mobile Devices via Inertial-Visual State Tracking," in IEEE Transactions on Mobile Computing, vol. 20, no. 4, pp. 1714-1729, 1 April 2021, doi: 10.1109/TMC.2019.2961313.

[23] Y. Nie, C. Bian and L. Li, "Object Tracking in Satellite Videos Based on Siamese Network With Multidimensional Information-Aware and Temporal Motion Compensation," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 6517005, doi: 10.1109/LGRS.2022.3211695.

[24] X. Tian, P. Zheng and J. Huang, "Robust Privacy-Preserving Motion Detection and Object Tracking in Encrypted Streaming Video," in IEEE Transactions on Information Forensics and Security, vol. 16, pp. 5381-5396, 2021, doi: 10.1109/TIFS.2021.3128817.

[25] J. Wei, J. Sun, Z. Wu, J. Yang and Z. Wei, "Moving Object Tracking via 3-D Total Variation in Remote-Sensing Videos," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 3506405, doi: 10.1109/LGRS.2021.3077257.

[26] Shivadekar, S., Kataria, B., Limkar, S. et al. Design of an efficient multimodal engine for preemption and post-treatment recommendations for skin diseases via a deep learning-based hybrid bio-inspired process. Soft Comput (2023).

[27] Shivadekar, Samit, et al. "Deep Learning Based Image Classification of Lungs Radiography for Detecting COVID-19 using a Deep CNN and ResNet 50." International Journal of Intelligent Systems and Applications in Engineering 11.1s (2023): 241-250.

[28] P. Nguyen, S. Shivadekar, S. S. Laya Chukkapalli and M. Halem, "Satellite Data Fusion of Multiple Observed XCO2 using Compressive Sensing and Deep Learning," IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 2020, pp. 2073-2076, doi: 10.1109/IGARSS39084.2020.9323861.

[29] Banait, Satish S., et al. "Reinforcement mSVM: An Efficient Clustering and Classification Approach using reinforcement and supervised Techniques." International Journal of Intelligent Systems and Applications in Engineering 10.1s (2022): 78-89.

[30] Shewale, Yogita, Shailesh Kumar, and Satish Banait. "Machine Learning Based Intrusion Detection in IoT Network Using MLP and LSTM." International Journal of Intelligent Systems and Applications in Engineering 11.7s (2023): 210-223.

[31] Vanjari, Hrishikesh B., Sheetal U. Bhandari, and Mahesh T. Kolte. "Enhancement of Speech for Hearing Aid Applications Integrating Adaptive Compressive Sensing with Noise Estimation Based Adaptive Gain." International Journal of Intelligent Systems and Applications in Engineering 11.7s (2023): 138-157.

[32] Vanjari, Hrishikesh B., and Mahesh T. Kolte. "Comparative Analysis of Speech Enhancement Techniques in Perceptive of Hearing Aid Design." Proceedings of the Third International Conference on Information Management and Machine Intelligence: ICIMMI 2021. Singapore: Springer Nature Singapore, 2022.

Cite this article as :

Shivali Tiwari, Prof. Jayant Adhikari, "Empirical Analysis of Machine Learning Models Used for Motion Detection & Tracking in Videos from An Augmented Statistical Perspective", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 10 Issue 5, pp. 29-46, September-October 2023.
Journal URL : https://ijsrset.com/IJSRSET23103170