OPEN ACCESS
Global Journal of Research in Engineering & Computer Sciences
ISSN: 2583-2727 (Online)
Volume 04 | Issue 01 | Jan.-Feb. | 2024
Journal homepage: https://gjrpublication.com/gjrecs/
Review Article

Intelligent Eyes for Smart Security Environments

*Prof. Aravind Naik[1], Dr. Sandeep Bhat[2], Prof. Daya Naik[3]

[1]Assistant Professor, Department of Computer Science and Engineering, Srinivas Institute of Technology, Mangaluru, Karnataka, India
[2]Professor, Department of Computer Science and Engineering, Srinivas Institute of Technology, Mangaluru, Karnataka, India
[3]Assistant Professor, Department of AIML, Srinivas Institute of Technology, Mangaluru, Karnataka, India
DOI: 10.5281/zenodo.10578303 **Submission Date: 19 Nov. 2023 | Published Date: 29 Jan. 2024**

***Corresponding author: Prof. Aravind Naik**
Assistant Professor, Department of Computer Science and Engineering, Srinivas Institute of Technology, Mangaluru, Karnataka,India

Abstract
Human motion detection and prediction have become pivotal fields in computer vision and artificial intelligence, revolutionizing applications ranging from surveillance and robotics to healthcare and interactive technologies. The ability to perceive and anticipate human movements is crucial for developing intelligent systems that can seamlessly interact with the physical world and enhance overall user experience. Predicting environmental human motion involves forecasting future positions, poses, or actions of individuals based on their past movements. In this comprehensive review, various approaches and techniques for human motion detection and prediction are explored. The methods used, major challenges and considerations for different approaches are highlighted. Although various methods are used in the detection and prediction of human motion, neural networks and deep learning have transformed the landscape of human motion detection and prediction, enabling the development of accurate and versatile models. With ongoing research, the integration of emerging architectures, attention mechanisms, and multi-modal approaches is expected to further enhance the capabilities of these models across a wide range of applications.

Keywords: Human motion detection, human motion prediction, human robot interaction, Computer vision, deep learning for motion.

1.INTRODUCTION
In a world increasingly characterized by technological advancements, the accurate detection and prediction of human motion play a fundamental role in addressing numerous challenges. Human motion detection and prediction is an advancing field at the intersection of computer vision, machine learning, and robotics, aimed at understanding, modelling, and forecasting the movements of individuals. As the demand for intelligent systems continues to rise, the ability to perceive and anticipate human motion is becoming increasingly essential for applications ranging from surveillance and security to human-computer interaction and virtual reality [1].

Human motion detection involves the identification and localization of individuals within a visual scene, often in the context of video surveillance or real-time monitoring systems. On the other hand, motion prediction extends this capability by forecasting the future trajectories, poses, or actions of individuals, contributing to enhanced situational awareness and proactive decision-making. With the advent of deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs)[7], as well as advancements in sensor technologies, researchers and practitioners are exploring innovative approaches to address challenges in accuracy, real-time processing, and adaptability to diverse and dynamic environments.

Several algorithms have been employed for human motion detection and prediction, covering a spectrum of traditional computer vision methods to sophisticated machine learning and deep learning approaches. The traditional technique Background Subtraction where the background of a scene is modelled, and moving objects are identified by subtracting the background from the current frame.

1.1 Frame Differencing
The simplest form of background subtraction involves subtracting the current frame from the previous frame. The advantage is that it is computationally inexpensive and easy to implement, but it is sensitive to noise and lighting changes and may not handle gradual variations in illumination well.[2]

1.2 Temporal Averaging (Running Average)
This approach computes an average image of the video frames over time and subtract it from each frame. It is effective in handling gradual changes in the background, but can be sensitive to sudden changes that is it may not adapt well to dynamic scenes.

1.3 Gaussian Mixture Models (GMM)
The basic approach models the pixel intensities as a mixture of several Gaussian distributions. The advantage is that it is adaptable to varying lighting conditions and robust to noise. The GMMs can struggle with complex scenes and may require tuning parameters.

1.4 Codebook Models
In this approach we represent the background using a set of codewords, each associated with a pixel value. The advantage is that it is efficient and works well with stationary backgrounds but may find difficulty with dynamic scenes and requires careful parameter tuning.

1.5 Eigen-backgrounds
The basic Idea here represents the background as a linear combination of eigenvectors obtained through Principal Component Analysis (PCA). This approach is very effective in modelling complex backgrounds. The approach is sensitive to changes in lighting and may require periodic updates.

1.6 Adaptive Background Models
It dynamically updates the background model over time, adapting to changes. It may pose limitations when there are sudden changes and the parameters need to be carefully adjusted.

1.7 Statistical Methods with Thresholding
This approach uses statistical measures like mean and standard deviation to identify foreground pixels based on a threshold. It is simple and computationally efficient, but sensitive to noise and may require adaptive thresholding.

1.8 Deep Learning-Based Approaches
These approaches Utilize Convolutional Neural Networks (CNNs) or other deep architectures to learn features for foreground/background discrimination. They have the advantage to handle complex scenes and learn hierarchical representations. These approaches require substantial computational resources and training data.

1.9 Hybrid Approaches
Hybrid approach combines multiple background subtraction techniques to improve overall performance. This approach can leverage the strengths of different methods, but with increased complexity and parameter tuning challenges.

1.10 Change Detection in Frequency Domain
In this approach the frequency content of the video frames is analysed to detect changes in the scene. This approach is robust to lighting changes and effective in certain scenarios. It may not perform well in complex scenes and is computationally demanding.

The choice of algorithm depends on the specific characteristics of the video data and the requirements of the application. It's often necessary to experiment and fine-tune parameters based on the nature of the scene being analysed. Additionally, real-world scenarios might benefit from a combination of algorithms to enhance accuracy and robustness.

2. Optical Flow Technique

Optical flow is a technique used in computer vision to estimate the motion of objects or surfaces in a sequence of images or video frames. It provides a dense motion field by estimating the apparent motion of pixels between consecutive frames. Optical flow is particularly useful for motion detection, tracking, and understanding the dynamics of a scene. Optical flow assumes that the intensity of a pixel does not change significantly between consecutive frames. This is known as the brightness constancy assumption.[3]

2.1 Lucas-Kanade Method:

The Lucas-Kanade method is a widely used technique for estimating optical flow. It assumes that the motion is essentially constant in a local neighborhood. It solves the optical flow equation for a small window around each pixel using least squares.

2.2 Horn-Schunck Method:

The Horn-Schunck method is another approach for optical flow estimation. It considers the entire image and enforces global smoothness constraints on the estimated flow. It formulates optical flow as a global optimization problem.

The optical flow poses challenges like Aperture problem, Ambiguity in regions with repetitive patterns, occlusions or when there are abrupt changes in scene. The computational cost of optical flow can be high especially in dense motion estimation scenarios. Optimization methods may require iterative processing. Also, the Parameter tuning like selection of window size, regularization parameters, and handling of discontinuities require careful tuning for optimal results.

Optical flow is a powerful tool for motion detection, and its applications range from video surveillance and object tracking to robotics and autonomous navigation. While traditional methods like Lucas-Kanade and Horn-Schunck are still relevant, recent advancements in deep learning have led to the development of deep optical flow methods that can handle more complex scenarios.

3. Methodologies of Motion Detection

3.1. Blob Analysis

Blob analysis is commonly employed in motion detection applications to identify and track moving objects within a video stream or sequence of images. The process involves detecting connected components (blobs) in consecutive frames and analyzing their properties to infer motion.

Blob analysis is sensitive to noise, and small variations in pixel values can lead to false positives. Proper pre-processing and filtering are crucial. Blob analysis may need to adapt to changes in lighting conditions, sudden scene changes, or variations in the background. Real-time applications require efficient implementations of blob analysis to keep up with the video frame rate.

3.2. Skeletonization and Joint Detection

Skeletonization and joint detection are crucial components of human motion detection systems, particularly in computer vision applications related to action recognition, gesture analysis, and human-computer interaction. Skeletonization is a process that reduces the shape of an object to its essential geometric structure, often represented as a set of connected lines or points. In human motion detection, this structure represents the human skeleton. Skeletonization may face challenges in accurately capturing complex poses, dealing with occlusions, and maintaining consistency across frames in dynamic scenes.

Joint detection involves identifying key joints or key-points on the human skeleton, such as the head, shoulders, elbows, wrists, hips, knees, and ankles. These joints serve as critical reference points for characterizing human poses and movements. Heat Map Regression, Part based models and deep learning approaches are common methods used in joint detection. Challenges in joint detection include dealing with occlusions, variations in poses, and ensuring accurate localization of key points under different lighting conditions and background complexities.

Skeletonization and joint detection play complementary roles in human motion detection. Skeletonization simplifies the representation of the human body, while joint detection provides detailed information about key anatomical points. Integrating these approaches enables the understanding and analysis of complex human movements in diverse applications.

3.3. Hidden Markov Models (HMMs)

Hidden Markov Models (HMMs) are widely used in the field of human motion estimation due to their ability to model temporal dependencies in sequential data. In the context of human motion, HMMs can be employed to represent and recognize different motion patterns, making them valuable in various applications.Human motion is inherently temporal, and actions or gestures involve a sequence of movements over time. HMMs are well-suited for modelling such temporal dependencies.

HMMs can be trained to recognize a variety of motion patterns, making them suitable for applications with diverse human activities. HMMs can handle missing or noisy observations, making them robust in real-world scenarios where complete and accurate data may not always be available. In real-time applications, the computational complexity of HMMs may be a consideration. Efficient algorithms and model optimizations may be necessary. HMMs may face challenges when dealing with occlusions, variations in motion speed, or changes in the environment.

3.4. Support Vector Machines (SVMs)

Support Vector Machines (SVMs) have been employed in various aspects of human motion detection and prediction due to their effectiveness in handling complex, high-dimensional data and nonlinear relationships. SVMs are a type of supervised machine learning algorithm that can be applied to both classification and regression tasks.

SVM performance can be affected by the dimensionality of feature vectors. Proper feature selection or dimensionality reduction techniques are crucial. SVMs may struggle with modelling complex temporal dynamics in human motion sequences. Hybrid approaches or additional temporal modelling techniques may be required.

3.5. Neural networks and Deep Learning

The usage of Neural Networks (NNs) and Deep Learning (DL) in human motion detection and prediction has witnessed significant advancements.CNNs are extensively used for spatial feature extraction from image sequences, capturing hierarchical patterns in individual frames. Pre-trained CNNs, such as ResNet and VGG, are commonly used as feature extractors. 3D CNNs directly model spatiotemporal features from video sequences, allowing for end-to-end learning of motion patterns over time.

Incorporating two streams - one for spatial (RGB frames) and one for temporal (optical flow) information improves the understanding of both appearance and motion dynamics, enhancing action recognition accuracy.RNNs, particularly Long Short-Term Memory (LSTM) networks, are employed for recognizing temporal dependencies in sequential data, making them suitable for dynamic gesture recognition. Attention mechanisms, like those in Transformer architectures, can selectively focus on relevant frames, improving the model's ability to attend to critical temporal features. With ongoing research, the integration of emerging architectures, attention mechanisms, and multi-modal approaches is expected to further enhance the capabilities of these models across a wide range of applications.

3.6 Depth based Approaches

The depth-based approaches in human motion detection and prediction leverage depth information obtained from depth sensors, such as Time-of-Flight (ToF) cameras or structured light sensors, to capture the three-dimensional structure of the scene. These approaches are valuable for overcoming challenges related to occlusions, lighting conditions, and variations in appearance that can be present in RGB-based methods.

Depth data enables the tracking of individuals in three-dimensional space, allowing for more accurate trajectory analysis and behaviour understanding. Unusual behaviours or deviations from typical motion patterns can be detected more reliably using depth information, aiding in anomaly detection.Depth sensors may have limitations in terms of range, accuracy, and susceptibility to environmental factors like reflections or interference. Depth data can be noisy, and occlusions may lead to missing information. Robust algorithms are needed to handle such challenges in real-world scenarios.

Depth-based approaches play a crucial role in human motion detection and prediction, offering enhanced capabilities in scenarios where RGB-based methods may face challenges. As technology continues to advance, integrating depth sensing with other modalities and addressing real-world challenges will further expand the applicability of these approaches.

3.7 Kalman Filters

Kalman Filters are widely used in human motion detection and prediction to estimate the state of an object, such as the position, velocity, and possibly other parameters, based on noisy measurements obtained over time. Kalman Filters excel in scenarios where there is uncertainty in the measurements and a need for real-time or dynamic tracking. Kalman Filters are known for their efficiency and low computational cost, making them suitable for real-time applications where low latency is crucial, such as interactive systems. The accuracy of Kalman Filters relies on the correctness of the dynamic system model. Deviations from the assumed model can impact the accuracy of predictions. Kalman Filters are sensitive to outliers and may not perform well in the presence of abrupt changes or outliers in the measurements. Robust variants or outlier rejection mechanisms may be necessary.

3.8 Long Short Term Memory (LSTM) Networks

Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), are well-suited for human motion detection and prediction tasks due to their ability to capture temporal dependencies in sequential data. LSTMs excel in handling long-range dependencies and can effectively model the dynamics of human motion sequences. As with any deep learning approach, careful consideration of data pre-processing, model architecture design, and training parameters is essential to enhance the capabilities of LSTMs.[13]

3.9 Ensemble methods

Ensemble methods combine predictions from multiple base models to make a final prediction. The diversity among base models is crucial for the success of ensembles.

Ensemble methods have the benefits of improved generalization, increased stability, and enhanced performance compared to individual models. Different types of ensemble methods such as Bagging, Boosting, Random forests, stacking, voting classifiers are used for obtaining better accuracy and other performance parameters.

3.10 Generative Adversarial Networks (GANs)

GANs offer a versatile set of tools for generating realistic and diverse human motion sequences. Whether used for data augmentation, anomaly detection, or predictive modelling, the application of GANs in human motion analysis continues to evolve. As research progresses, addressing challenges related to training stability and ethical considerations will be crucial for maximizing the benefits of GANs in motion detection and prediction.[12]

4.CONCLUSION

The realm of human motion detection and prediction stands at the forefront of technological advancements, contributing significantly to the evolution of intelligent surveillance systems. The fusion of computer vision, machine learning, and predictive analytics has enabled us to decipher the intricacies of human movement, enhancing security measures and redefining surveillance paradigms. The robust algorithms for real-time motion detection to sophisticated predictive models anticipating future actions, the field continues to witness remarkable progress. These advancements have implications far beyond security, extending into areas such as human-computer interaction, robotics, and healthcare. However, challenges persist, including the need for ethical considerations, addressing biases, and ensuring robustness across diverse scenarios. As researchers and practitioners push the boundaries of innovation, the future promises a convergence of cutting-edge technologies, fostering a new era where the analysis and prediction of human motion play a pivotal role in shaping intelligent and responsive surveillance ecosystems.

REFERENCES

1. Sharma, Lavanya. "Human detection and tracking using background subtraction in visual surveillance."Towards Smart World (2020): 317-328.1.
2. Monnet, Mittal, Paragios and Visvanathan Ramesh, "Background modeling and subtraction of dynamic scenes," Proceedings Ninth IEEE International Conference on Computer Vision, Nice, France, 2003, pp. 1305-1312 vol.2, doi: 10.1109/ICCV.2003.1238641.
3. Ranjan, Anurag, Javier Romero, and Michael J. Black. "Learning human optical flow."arXiv preprint arXiv:1806.05666(2018).
4. K. Kale, S. Pawar and P. Dhulekar, "Moving object tracking using optical flow and motion vector estimation,"*2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, Noida, India, 2015, pp. 1-6, doi: 10.1109/ICRITO.2015.7359323.
5. Afsar, Palwasha, Paulo Cortez, and Henrique Santos. "Automatic human trajectory destination prediction from video."Expert Systems with Applications110 (2018): 41-51.
6. Win, Sandar, and Thin Lai Lai Thein. "Real-time human motion detection, tracking and activity recognition with skeletal model." 2020 IEEE Conference on Computer Applications (ICCA). IEEE, 2020.
7. Martinez, Julieta, Michael J. Black, and Javier Romero. "On human motion prediction using recurrent neural networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
8. Cao, Dongwei, et al. "Human motion recognition using support vector machines." Computer Vision and Image Understanding 113.10 (2009): 1064-1075.
9. Butepage, Judith, et al. "Deep representation learning for human motion prediction and classification. "Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
10. Rudenko, Andrey, et al. "Human motion trajectory prediction: A survey."The International Journal of Robotics Research 39.8 (2020): 895-935.

11. Gui, Liang-Yan, et al. "Teaching robots to predict human motion."2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018.
12. Hassan, Muhammad Ahmed, et al. "Predicting humans future motion trajectories in video streams using generative adversarial network."Multimedia Tools and Applications (2021): 1-23.
13. Rossi, Luca, et al. "Human trajectory prediction and generation using LSTM models and GANs."Pattern Recognition120 (2021): 108136.

CITATION
Aravind N., Sandeep B., & Daya N. (2024). Intelligent Eyes for Smart Security Environments. In Global Journal of Research in Engineering & Computer Sciences (Vol. 4, Number 1, pp. 40–45). https://doi.org/10.5281/zenodo.10578303