A Mechanism on Video Tracking and Recognition with Fast Transmission in Wireless Communications

Jianjun Yang
University of North Georgia
3820 Mundy Mill Rd.,
Oakwood, GA, USA

Mingyuan Yan
University of North Georgia
82 College Circle
Dahlonega, GA, USA

Abi Salimi
University of North Georgia
82 College Circle
Dahlonega, GA, USA

Jason Porter
University of North Georgia
82 College Circle
Dahlonega, GA, USA

Ying Luo
Purdue University Northwest
2200 169th Street
Hammond, IN, USA

Ju Shen
University of Dayton
300 College Park
Dayton, OH, USA

ABSTRACT

In recent years, video-based object tracking and recognition have become critical research areas in computer vision and network communication systems. This paper proposes a novel method for video-based object tracking and recognition over fast network transmission using newly developed improved CAMSHIFT and Meanshift algorithms. The new mechanism is not only applies to single object tracking and recognition, but also to multiple objects and objects with motion. The proposed system also utilizes a wireless network to transmit video streams from remote devices to a central server where the tracking and recognition processes take place. The developed approach has several advantages, including the ability to track objects in real-time over wireless networks with low latency and high reliability. The proposed system can be applied in various applications, such as security surveillance, traffic monitoring, and human-computer interaction. In addition, an efficient and effective implementation is designed utilizing a wireless network to transmit video streams where the tracking and recognition processes take place.

Keywords

Tracking, Recognition, Transmission, Object tracking, Probability density function, Object detection, Video Tracking, Recognition, Wireless Sensor Network, Fast Transmission

1. INTRODUCTION

Video tracking & recognition as well as transmission in wireless environment are hot research topics nowadays[1] [2][3][4][5][6]. With the proliferation of smart devices and the internet of things (IoT), the demand for real-time object recognition and tracking has increased significantly. Object recognition and tracking have numerous applications, including security surveillance, autonomous vehicles, human-computer interaction, and industrial automation. In recent years, researchers have focused on developing algorithms and techniques to improve the accuracy and efficiency of object recognition and tracking systems.

Wireless networks have become an integral part of modern communication systems, enabling devices to communicate with each other wirelessly. The use of wireless networks in object recognition and tracking systems provides several benefits, such as flexibility, mobility, and ease of deployment. However, the limited bandwidth, latency, and reliability of wireless networks can also pose significant challenges to the development of real-time object recognition and tracking systems. To address these challenges, researchers have proposed various techniques and algorithms for object recognition and tracking on wireless networks. These techniques include feature-based methods, such as SIFT, SURF, and HOG[7], and model-based methods, such as CAMSHIFT and Meanshift[26][27]. In addition, researchers have proposed various machine learning algorithms, such as neural networks and support vector machines (SVM), for object recognition and tracking[28].

The proposed techniques and algorithms have shown promising results in various applications. However, the performance of these techniques is highly dependent on the network conditions, such as network bandwidth, packet loss, and latency. Therefore, it is essential to develop efficient and reliable object recognition and tracking systems that can operate in different network conditions. This paper presents a novel method for video-based object tracking and recognition over wireless networks using newly developed improved CAMSHIFT and Meanshift algorithms. The method utilizes the wireless network to transmit video streams from a remote camera to a central server, where the tracking and recognition processes take place. The system automatically adjusts the tracking parameters to account for changes in the object's position, size, and orientation in the video stream. The experimental results demonstrate that the developed new method achieves high tracking and recognition accuracy even under different network conditions.

The following sections dive deeper into the research on video tracking and recognition in wireless communications. Section 2 provides a comprehensive review of the relevant literature, setting the stage for the new scheme which is presented in Section 3. Then the article proceeds to evaluate the effectiveness of the new mechanism through simulated experiments in Section 4. Finally, Section 5 makes the conclusion by summarizing the key findings and highlighting opportunities for future research in this exciting area of study.

2. RELATED WORK

Numerous techniques have been developed for video processing, as evidenced by the existing literature [8, 9, 10, 11, 12]. One such approach was proposed by Q. Cai et al. [13, 14], who devised a mechanism to capture video in Wireless Mesh Networks. Their work was founded on the principles of adaptive learning and dynamic matching, which allowed for the identification of spatial and temporal changes.

J. Zhang et al. [15] proposed a wireless image sensor network-based system for wildlife monitoring. The architecture of the system consists of multiple image sensor nodes and a sink node. The authors provided a detailed design for a self-powered and rotatable wireless infrared image sensor node, along with supporting software design. The image sensors collect data within their range and transmit it to the sink node, which serves as a central processing station. The sink node aggregates and processes the data before transmitting it to the 3G network. The system is capable of providing accurate, automated, and remote monitoring in all weather conditions, making it an ideal solution for wildlife monitoring applications.

K. Gavrilyuk et al. [16] introduced an actor-transformer model that is capable of recognizing individual actions and group activities in videos. The model leverages both static and dynamic representations of actors obtained from a 2D pose network and 3D CNN. The authors investigate different techniques to integrate these representations and demonstrate their synergistic advantages. Through experiments on two public datasets for group activity recognition, the actor-transformers achieve state-of-the-art performance, outperforming the prior state of the art.

C. Zalluhoglu et al. [17] contribute to the field of collective activity recognition by addressing available datasets and introducing a new benchmark dataset, named ”Collective Sports (C-Sports),” which aims to recognize both collective activity and sports categories in a multitask manner. The authors evaluate state-of-the-art techniques and their multi-task variants on this dataset, achieving improved performance. However, there is still a need for improvement in collective activity recognition, especially in the generalization ability beyond previously seen sports categories. To evaluate this ability, the authors introduce a new evaluation protocol named ”unseen sports.” Overall, the C-Sports dataset should generate more interest in this research direction.

Y. Tang et al. [18] proposed a Semantics-Preserving Teacher-Student (SPTS) model for group activity recognition in videos that automatically identifies key people and discards misleading ones. This is achieved by allocating semantics-preserving attention to different people, a departure from conventional methods that aggregate features from individuals using pooling operations, which fails to fully explore contextual information. The SPTS network comprises a Teacher Network in the semantic domain and a Student Network in the vision domain. The Teacher Network determines group activity from individual actions, while the Student Network mimics the Teacher Network during learning. This method explores contextual information of different people and requires no extra labeled data, leading to superior performance compared to state-of-the-art techniques.

Several scholars have also conducted recognition for groups of people. For instance, B. Wang et al. [1] developed a model that utilizes high-order context to recognize groups of people. Ibrahim, G. et al. [2] designed a mechanism that can distinguish between individual and multiple people’s activities. Additionally, L. Wu et al. [3] created an algorithm that can recognize group activities in any scenario.

3. MECHANISM OF TRACKING & RECOGNITION AND TRANSMISSION

3.1 Objectives

The primary objective of this paper is to develop a robust and efficient system capable of accurately recognizing and tracking a moving object over time against a simple background while ensuring reliable and fast video transmission over a wireless network environment. To achieve this ambitious goal, this artilce is evaluating and implementing state-of-the-art tracking algorithms that suit the unique features of the object, optimizing their implementation to increase efficiency and accuracy. Specifically, the aim is to achieve 100% recognition and tracking accuracy for objects and gestures within a given scene.

In terms of the wireless network model in this article, it is committed to improving the compression ratio to 200, increasing the resolution to 352×288, and enhancing the transfer speed to a range of 10Kbps to 1.0Mbps. It will leverage the mobile TCP protocol to facilitate reliable video transfer over networks, ensuring that the transmitted video is delivered with minimal delay and packet loss. Furthermore, the article designed a user-friendly interface that features intuitive menus for loading and playing video files, as well as for tracking and transmitting the video. The focus on the user interface is essential to ensure that the system can be easily and efficiently used by operators with minimal technical knowledge, enabling the system to be widely used in various domains, including surveillance, sports, and healthcare.

3.2 The Method

To simplify the system design, this article makes the following assumptions about the video: there is a stationary background with a single hand moving back and forth continuously. The system design consists of four main components:

3.2.1 Object Detection and Tracking. The object detection and tracking implementation in this paper is based on the CAMSHIFT algorithm [26][27]. The CAMSHIFT algorithm is a popular face and colored object tracker that is known for its efficiency and ability to track objects accurately even in noisy environments. The CAMSHIFT algorithm is based on the MEAN SHIFT algorithm, which is used to find the mode of a probability density function. In the case of object tracking, the probability density function represents the color distribution of the object being tracked. The MEAN SHIFT algorithm works by iteratively shifting the center of a window towards the mode of the probability density function until convergence is reached. The CAMSHIFT algorithm improves upon the MEAN SHIFT algorithm by allowing the size and orientation of the window to adapt to the shape of the object being tracked. This is done by first applying the MEAN SHIFT algorithm to find the mode of the color distribution and then using the resulting window to calculate the moments of the object. The moments are then used to update the size, orientation, and center of the window for the next iteration, which allows the window to better fit the object being tracked.

(1)Back-projection Process
Given a color image and a color histogram, the image produced from the original color image by using the histogram as a look-up table is called back-projection image. If the histogram is a model density distribution, then the back projection image is a probability distribution of the model in the color image.

(2) Mean Shift Algorithm
CAMSHIFT detects the mode in the probability distribution image by applying mean shift while dynamically adjusting the parameters of the target distribution. In a single image, the process is iterated until convergence—an upper bound on the number of iterations is reached. The main steps of the Mean Shift algorithm are:

Algorithm 1 MeanShift Algorithm for Image Detection and Tracking

```
Require: $I$: Input image, $W$: Initial search window
Ensure: $p$: Peak in the probability density function of the object
    being tracked, $x$: Final position of the object
$p \leftarrow$ Probability density function of the object in $W$
repeat
    $x \leftarrow$ Center of mass of $p$ in $W$
    $W \leftarrow$ New window centered at $x$
    $p \leftarrow$ Probability density function of the object in $W$
until convergence
```

Note that this algorithm assumes that the object being tracked is represented by a probability density function p, and that the initial search window W is provided. The algorithm iteratively computes the center of mass of p in W, and then updates W to be centered at this new position. The process repeats until convergence, at which point the peak p and final position x of the object have been determined.

(3) Improved CAMSHIFT Algorithm
The mean shift algorithm is suitable for static probability distributions, but may not be effective for dynamic ones. In order to track a single target in a video sequence, CAMSHIFT adjusts the size of the search window for the next frame by using the 0th moment of the current frame's distribution. This approach enables the algorithm to anticipate object movement and quickly track the object in the next scene. By constraining the search area around the last known position of the target, computational efficiency can be improved significantly. This feedback mechanism establishes a loop where the detection result serves as input for the subsequent detection process.

This article implements the CAMSHIFT algorithm and extend the functionalities to track more than one object. The main procedures of the CAMSHIFT algorithm are as follows:

Algorithm 2 Improved CamShift Algorithm for Multiple Object Tracking

```
Require: $I$: Input video, $W_1, W_2, ..., W_n$: Initial search windows
    for $n$ objects, $T$: Maximum number of iterations
Ensure: $p_1, p_2, ..., p_n$: Peaks in the probability density function of
    the objects being tracked, $x_1, x_2, ..., x_n$: Final positions of the
    objects
$F \leftarrow$ Frames of $I$ split into images
$p_i \leftarrow$ Probability density function of the $i$-th object in $W_i$ for
    the first frame $F_1$
$x_i \leftarrow$ Center of mass of $p_i$ in $W_i$ for the first frame $F_1$
for $t = 2$ to $T$ do
    $W_i \leftarrow$ New window for the $i$-th object centered at $x_i$ for the
        current frame $F_t$
    $p_i \leftarrow$ Probability density function of the $i$-th object in $W_i$
        for the current frame $F_t$
    $x_i \leftarrow$ Center of mass of $p_i$ in $W_i$ for the current frame $F_t$
    $s_i \leftarrow$ Size of the $i$-th window for the current frame $F_t$
    $h_i \leftarrow$ Hue histogram of the $i$-th object in $W_i$ for the current
        frame $F_t$
    $backproj_i \leftarrow$ Backprojection of $h_i$ onto $I$ for the current
        frame $F_t$
```

The algorithm begins by splitting the input video I into individual frames F using a tool such as IrFanView. For each object i being tracked, the initial search window W_i is defined based on its position in the first frame F_1. The probability density function p_i of the object in the search window is calculated, and the center of mass x_i of the probability density function is determined using the MeanShift algorithm. For each subsequent frame F_t in the video sequence, the search window W_i for each object i is updated to be centered at the current position of the object x_i using the MeanShift algorithm. The probability density function p_i and the center of mass x_i are then recalculated for each object in the updated search window. The size s_i of the search window for each object and the hue histogram h_i of the object within the search window are also computed for each frame F_t. The hue histogram h_i is then used to generate a backprojection $backproj_i$ of the object onto the input image I for the current frame F_t. The CamShift algorithm uses the backprojection $backproj_i$ and the updated search window W_i to calculate the new position and size of the object for the current frame F_t. This is done by applying an iterative process to the backprojection $backproj_i$, which adjusts the size and orientation of the search window W_i until the peak in the probability density function is found. The resulting peak is used to determine the new position and size of the object for the current frame F_t. The CamShift algorithm is able to track multiple objects in a video sequence by repeating this process for each object being tracked.

3.2.2 *Video compression and transmission.* The proposed approach utilizes mean shift to detect the peak in the probability distribution image while simultaneously handling dynamic distributions by dynamically adjusting the search window size for the next frame based on the $0th$ moment of the current frame's distribution. This allows the new algorithm to efficiently track the object in the next scene by anticipating its movement. Moreover, this paper restricts the search area around the object's last known position, which results in significant computational savings. The process is iterated until convergence to an upper bound on the number of iterations is reached. In order to track the object in a video sequence, this article applys the detection algorithm to successive frames. This creates a feedback loop, where the result of one detection is used as input to the next detection process.

3.2.3 Interface design, system design and maintenance . To enhance the usability of the new system, this article designed a user-friendly interface by utilizing the uimenu function to generate intuitive main and submenus. To optimize the tracking process, it incorporated IrFanView, an efficient image processing tool, to split videos into multiple frames. This software tool also supports batch renaming and format conversion, making the video splitting process more convenient and streamlined. Once the individual frames are processed, this paper employs specific procedures to convert the batch of images back into a video format. These measures collectively contribute to a more efficient and user-friendly experience for the system users.

Algorithm 3 Video Processing and C Language Integration

```
Read the images into arrays
Convert the images into movie format by im2frame
Create a video from Matlab movie by movie2avi
To call C language in MATLAB environment:
    Writing the MEX-Function—the Interface to MATLAB
    Getting and Creating Data
    Calling Built-In Functions from a MEX-File
    Compiling, such as from Matlab interface, execute mex tracking.c
    Run the file in MATLAB interface
```

The algorithm first reads the input images and converts them into an array format. Next, the images are transformed into a movie format using the ”im2frame” function. Once the movie format is obtained, it is saved as a video using the ”movie2avi” function. To incorporate C language into the MATLAB environment, the algorithm follows a series of steps. These steps include writing the MEX-Function that serves as an interface to MATLAB, obtaining and creating the necessary data, and calling built-in functions from the MEX-File. Finally, the MEX-File is compiled, and the resulting file is executed in the MATLAB interface. Overall, the algorithm enables the conversion of image sequences to videos, while also incorporating C language functionality within the MATLAB environment.

4. EXPERIMENT AND DISCUSSION

In this section, the article presents the implementation of the proposed algorithm and the experimental results obtained through its application to various datasets. The article also performs a comprehensive evaluation of the algorithm to assess its effectiveness in tracking multiple objects in real-world scenarios. To this end, it analyze several performance metrics, including accuracy, robustness, and computational efficiency. Finally, the article explores the various potential applications and discuss specific scenarios where the new system can be deployed effectively, highlighting its benefits and limitations.

4.1 Object Detection and Tracking

The object detection and tracking approach in this article begins with a visual inspection to determine the size and location of the initial search window. Specifically, it sets the size of the initial search window to $W_{size} = [137, 158]$, and the initial location of the search window to $Location = [187, 341]$. To assess the performance of the new algorithm, the paper utilizes the probability distribution of skin and the corresponding frames from the video for analysis, as depicted in Figure 1. Specifically, it extracted frames 1, 29, 47, 79, 96, and 107 and created probability distribution images for each. The first row of images represents the original frames, while the second row corresponds to the skin distribution. The high intensity values in the probability distribution images indicate a higher probability of skin presence. It applied thresholding to the hue values to filter out the background and highlight the flesh color, simplifying the tracking algorithm. Additionally, the paper generated motion plots of the centroid of the search window for each frame. Based on the results obtained, it concludes that the program meets the project’s requirements. The red rectangle box encloses the target region during tracking. The corresponding tracking path, representing the 2D coordinates of the tracked hand over frames, is visualized in Figure 2.

In some cases, the hand target was not fully captured within the search window due to the scaling parameter, which required several iterations to optimize. Through extensive trials, the paper discovered that a scaling factor of 1.0 produced the most favorable results. Moreover, it determined that selecting an appropriate initial window size was crucial to avoid tracking the wrong object or background. To address this issue, it utilized hue thresholding to ensure proper tracking. Although the algorithm’s actual running time was not ideal, taking approximately 5 seconds to process a single frame, the article recognized the need to identify more efficient algorithms. This would further reduce the processing time while ensuring accuracy and effectiveness. Despite these challenges, the newly proposd algorithm represents a significant contribution to the field of computer vision and has potential for use in various applications.

Figure 3 illustrates the effectiveness of the improved Cam-shift algorithm in tracking multiple objects simultaneously in various video sequences. The four samples presented in the figure correspond to different scenarios, as indicated by the columns. Notably, the algorithm is able to track the target object accurately, even when it changes in size due to variations in camera perspective, as demonstrated in rows 2 to 4. For instance, the pedestrian walking in row 3 moves closer to the camera, resulting in a larger region of interest (ROI) in the image, yet the algorithm robustly tracks the object. Furthermore, the algorithm can track an object that was not captured in some frames, once it reappears in the scene. For instance, in sample 4, the coffee cup is accurately tracked in the first and last frames despite not being present in some of the intermediate frames. These results demonstrate the robustness and accuracy of the improved Cam-shift algorithm in tracking multiple objects in various scenarios.

4.2 Video compression and transmission

The developed approach to object detection and tracking relies on mean shift to detect peaks in the probability distribution image. By adjusting the search window size based on the 0^{th} moment of the current frame’s distribution, the algorithm is able to accurately track dynamic objects and anticipate object movement in the next scene. The search area can also be restricted around the last known position of the target, resulting in potentially large computational savings. The detection algorithm is then applied to successive frames of a video sequence to track the target, creating a feedback loop where the result of the detection is used as input for the next detection process.

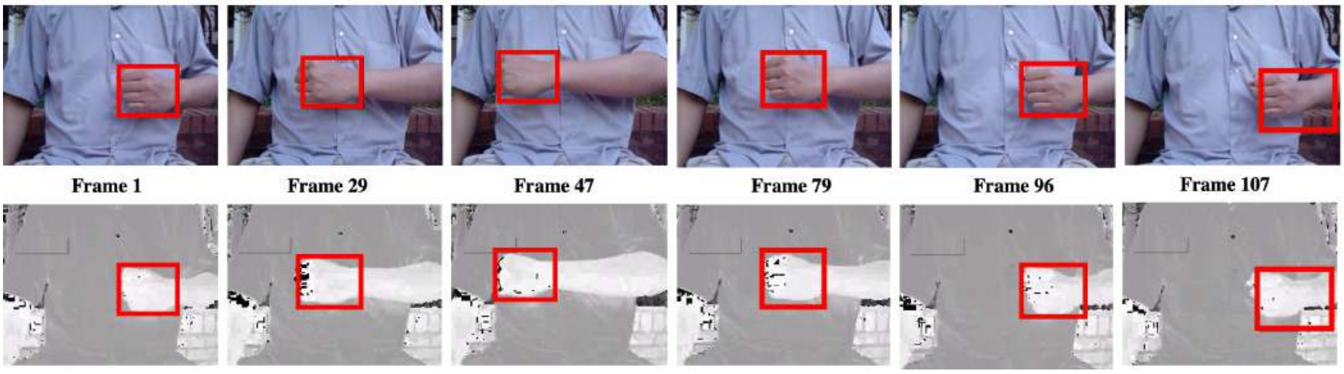


Fig. 1. The tracking result: the top row represents the original frames sampled from a video sequence; the bottom row represents the corresponding frames of skin probability distribution

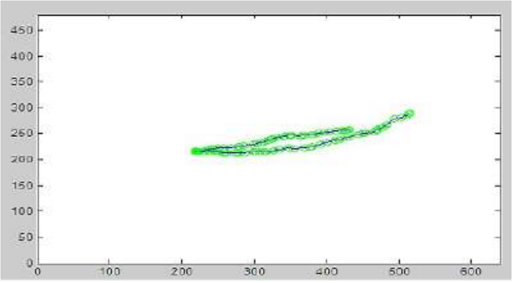


Fig. 2. Visualization of the tracking paths for a target object over time, demonstrating the effectiveness of the proposed tracking algorithm in accurately capturing the movement and trajectory of the targets.

To evaluate the effectiveness of the developed tracking and recognition algorithms, this paper gathered various videos and applied the new algorithms to them. After tracking the objects in the videos, it transmitted the resulting videos to the server using location-based forwarding instead of flooding. Specifically, it adopted the intermediate target-based geographic routing (ITGR) approached developed by Fei and Yang et al. [29], as illustrated in Figure 4. The network works by identifying intermediate target nodes that lie along the path between the source and destination nodes. These intermediate targets are selected based on their geographical proximity to the destination node, and the packet is forwarded from the source node to the intermediate targets until it reaches the final destination. This approach ensures that the packet follows a path that is both efficient and reliable, even in cases where the direct path between the source and destination is obstructed.

In addition to using the Mobile TCP protocol for reliable wireless transmission, the paper implemented a slow start phase that effectively controls the flow of data during transmission. By limiting the number of packets sent by the sender side through the use of a congestion window, it aimed to achieve a state of equilibrium in which the size of the congestion window increases exponentially over round trips. To ensure that the network is not overloaded, it determined the size of the congestion window based on the minimum of the congestion window and the receiver advertised window for flow control. As packets are transmitted, the receiver sends back ACKs to confirm receipt, and the congestion window is increased by one segment for each received ACK. Through this approach, the article was able to synchronize the tracking results among different computers, while ensuring reliable and efficient transmission. These results are illustrated in the right part of Figure 4.

5. APPLICATIONS

The proposed video-based tracking and recognition system using Camshift and wireless transmission can have various applications in areas such as security, surveillance, and robotics [19, 20]. In the field of security and surveillance, this system can be utilized for tracking people or objects in a crowded area, such as a shopping mall or train station [21]. It can also be applied in facial recognition for access control systems or classroom students monitoring purposes[22]. Moreover, it can be used in robotics for object tracking, which is particularly useful for autonomous robots or drones, where the ability to accurately track and follow moving targets is of utmost importance for a wide range of drone based applications such as aerial surveillance, search and rescue missions, precision agriculture.[23].

The seamless integration of a wireless network not only enhances the functionality of the system but also enables real-time applications in remote settings. For instance, in surveillance scenarios, the system can effectively track and transmit live video feed to a central command center where security personnel can monitor the situation and take immediate actions. Moreover, the system can be utilized in robotics to enable real-time tracking of moving objects or people, providing the robot with the ability to follow them or avoid obstacles in their path [24]. Furthermore, the system can be augmented with additional features, such as depth estimation or 3D mapping, to enable more complex tasks in augmented or virtual reality [25, 24]. Depth estimation provides distance information that can be useful for obstacle avoidance or navigation tasks. On the other hand, 3D mapping helps to build a more accurate model of the environment. With these additional features, the system can perform more advanced tasks and enable a wider range of applications.

6. ANALYSIS

The proposed mechanism in this paper is based on the CAMSHIFT algorithm. The CAMSHIFT algorithm is a popular face and colored object tracker that is known for its efficiency and ability to track objects accurately even in noisy environments. This

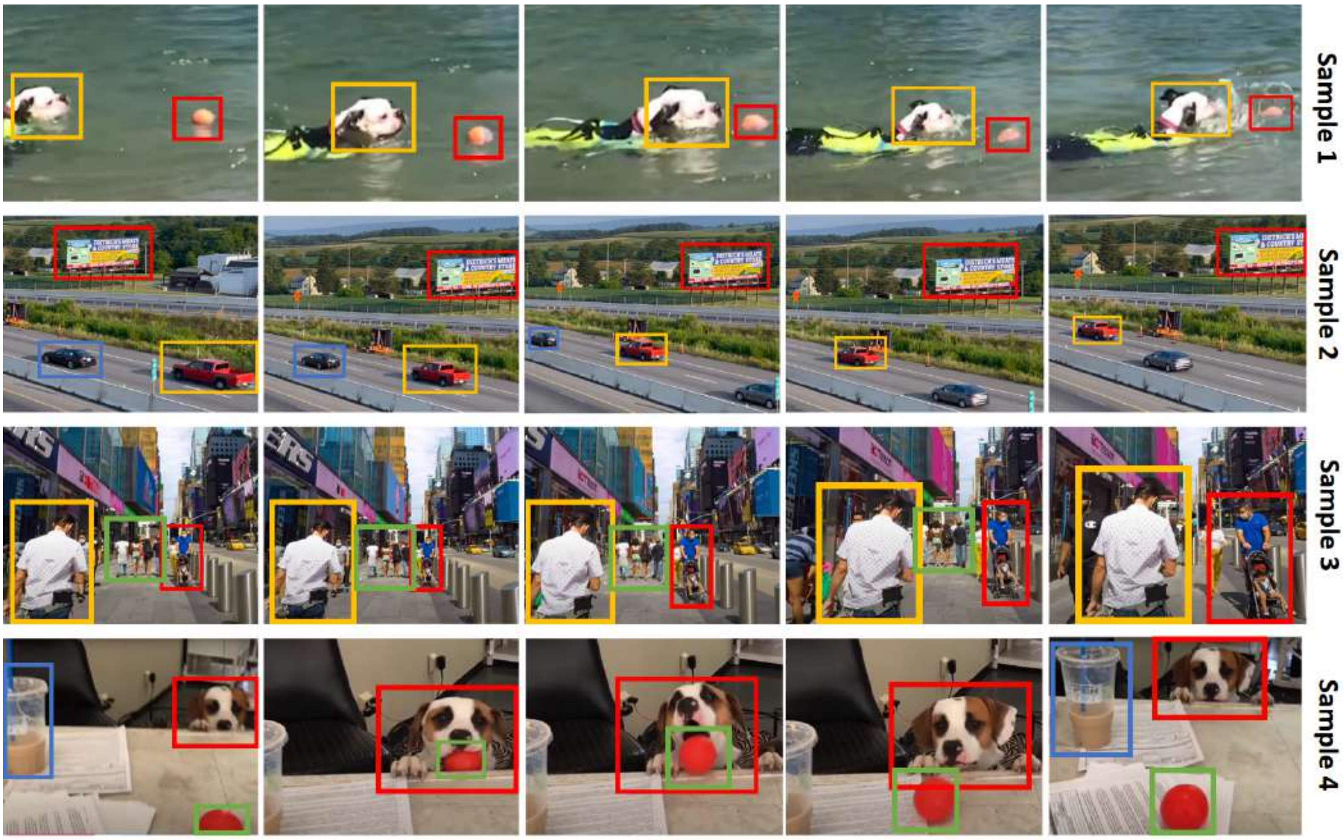
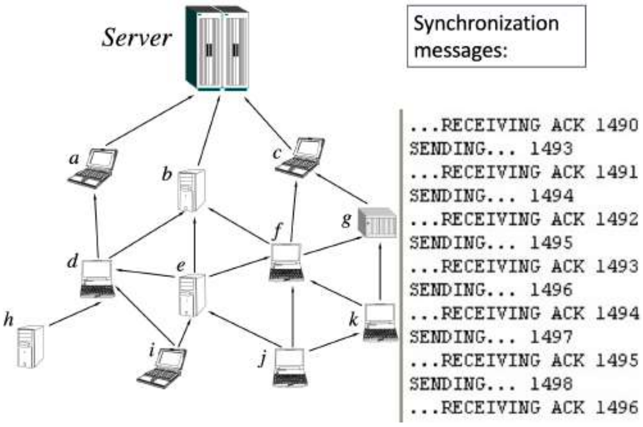


Fig. 3. More tracking results: multiple objects tracking in various scenarios.



```
...RECEIVING ACK 1490
SENDING... 1493
...RECEIVING ACK 1491
SENDING... 1494
...RECEIVING ACK 1492
SENDING... 1495
...RECEIVING ACK 1493
SENDING... 1496
...RECEIVING ACK 1494
SENDING... 1497
...RECEIVING ACK 1495
SENDING... 1498
...RECEIVING ACK 1496
```

Fig. 4. Synchronization of tracking results among different computers using Mobile TCP protocol and slow start algorithm for reliable wireless transmission

article implemented the CAMSHIFT algorithm and extended the functionalities to track more than one object. This article also improved CAMSHIFT Algorithm. The mean shift algorithm is suitable for static probability distributions, but may not be effective for dynamic ones. In order to track a single target in a video sequence, CAMSHIFT adjusts the size of the search window for the next frame by using the 0th moment of the current frame's distribution. This approach enables the algorithm to anticipate object movement and quickly track the object in the next scene.

However, the proposed approach does not automatically determine the initial search window instead of relying on human inspection. Also, it could be improved to handle the color-based segmentation performance by considering other spatial features besides the hue value. In summary, the paper has successfully integrated the two main parts into a whole tracking and transmission system, and it has designed a user interface. While it has realized most of the functionalities, there is still room for improvement in terms of an intelligent, more user-friendly human interface.

7. CONCLUSION AND FUTURE WORK

This study successfully developed a novel mechanism that combines Image Processing and Wireless Networking to achieve Video Based Tracking and Recognition via Wireless Transmission. The improved CAMSHIFT algorithm has demonstrated remarkable robustness in tracking objects. This computationally efficient face and colored object tracker works based on color or hue, and it enhances the MEAN SHIFT by accounting for dynamic probability distributions. The CAMSHIFT algorithm excels in tracking flesh-colored objects that move quickly, and it self-corrects. However, parameters such as thresholds and search window scaling factors must be carefully selected to track the correct object. Future work should incorporate into larger, more complex modules that provide more robust tracking.

8. REFERENCES

[1] Wang, M., Ni, B. and Yang, X., 2017. Recurrent modeling of interaction context for collective activity recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3048-3056)

[2] Ibrahim, M.S. and Mori, G., 2018. Hierarchical relational networks for group activity recognition and retrieval. In Proceedings of the European conference on computer vision (ECCV) (pp. 721-736).

[3] Wu, J., Wang, L., Wang, L., Guo, J. and Wu, G., 2019. Learning actor relation graphs for group activity recognition. In Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (pp. 9964-9974).

[4] Shu, X., Zhang, L., Sun, Y. and Tang, J., 2020. Host–parasite: Graph LSTM-in-LSTM for group activity recognition. IEEE transactions on neural networks and learning systems, 32(2), pp.663-674.

[5] Ehsanpour, M., Abedin, A., Saleh, F., Shi, J., Reid, I. and Rezatofighi, H., 2020. Joint learning of social groups, individuals action and sub-group activities in videos. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16 (pp. 177-195). Springer International Publishing.

[6] Vahora, S.A. and Chauhan, N.C., 2019. Deep neural network model for group activity recognition using contextual relationship. Engineering Science and Technology, an International Journal, 22(1), pp.47-54.

[7] Srinivas, D. and Hanumaji, K., 2019. Analysis of various image feature extraction methods against noisy image: SIFT, SURF and HOG. J Eng Sci, 10(2), pp.32-36.

[8] Wu, L.F., Wang, Q., Jian, M., Qiao, Y. and Zhao, B.X., 2021. A comprehensive review of group activity recognition in videos. International Journal of Automation and Computing, 18, pp.334-350.

[9] Liu, Z., Abbas, A., Jing, B.Y. and Gao, X., 2012. WaVPeak: picking NMR peaks through wavelet-based smoothing and volume-based filtering. Bioinformatics, 28(7), pp.914-920.

[10] Shen, J. and Cheung, S.C.S., 2013. Layer depth denoising and completion for structured-light rgb-d cameras. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1187-1194).

[11] Shen, J., Raghunathan, A., Sen-ching, S.C. and Patel, R., 2011, July. Automatic content generation for video self modeling. In 2011 IEEE International Conference on Multimedia and Expo (pp. 1-6). IEEE.

[12] Yang, J., Hua, K., Wang, Y., Wang, W., Wang, H. and Shen, J., 2014, April. Automatic objects removal for scene completion. In 2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) (pp. 553-558). IEEE.

[13] Cai, Q., Yin, Y. and Man, H., 2013, July. Dspm: Dynamic structure preserving map for action recognition. In 2013 IEEE international conference on multimedia and expo (ICME) (pp. 1-6). IEEE.

[14] Cai, Q., Yin, Y. and Man, H., 2013, September. Learning spatio-temporal dependencies for action recognition. In 2013 IEEE International Conference on Image Processing (pp. 3740-3744). IEEE.

[15] Zhang, J., Luo, X., Chen, C., Liu, Z. and Cao, S., 2014. A wildlife monitoring system based on wireless image sensor networks. Sensors and Transducers, 180(10), p.104.

[16] Gavrilyuk, K., Sanford, R., Javan, M. and Snoek, C.G., 2020. Actor-transformers for group activity recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 839-848).

[17] Zalluhoglu, C. and Ikizler-Cinbis, N., 2020. Collective Sports: A multi-task dataset for collective activity recognition. Image and Vision Computing, 94, p.103870.

[18] Tang, Y., Wang, Z., Li, P., Lu, J., Yang, M. and Zhou, J., 2018, October. Mining semantics-preserving attention for group activity recognition. In Proceedings of the 26th ACM international conference on Multimedia (pp. 1283-1291).

[19] Han, Z., Zhang, R., Wen, L., Xie, X. and Li, Z., 2016, December. Moving object tracking method based on improved camshift algorithm. In 2016 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII) (pp. 91-95). IEEE.

[20] Angadi, S. and Nandyal, S., 2020. A review on object detection and tracking in video surveillance. International Journal of Advanced Research in Engineering and Technology, 11(9).

[21] Liu, J., Sridharan, S. and Fookes, C., 2016. Recent advances in camera planning for large area surveillance: A comprehensive review. ACM Computing Surveys (CSUR), 49(1), pp.1-37.

[22] Harahap, M., Manurung, A., Prakoso, A. and Tambunan, M.F., 2019, July. Face tracking with camshift algorithm for detecting student movement in a class. In Journal of Physics: Conference Series (Vol. 1230, No. 1, p. 012018). IOP Publishing.

[23] Kanellakis, C. and Nikolakopoulos, G., 2017. Survey on computer vision for UAVs: Current developments and trends. Journal of Intelligent and Robotic Systems, 87, pp.141-168.

[24] Kim, J.S., Kim, M.G. and Pan, S.B., 2021. A study on implementation of real-time intelligent video surveillance system based on embedded module. EURASIP Journal on Image and Video Processing, 2021(1), pp.1-22.

[25] Lee, J. and Park, S.Y., 2021. PLF-VINS: Real-time monocular visual-inertial SLAM with point-line fusion and parallel-line fusion. IEEE Robotics and Automation Letters, 6(4), pp.7033-7040.

[26] Fran, A.R.J., 2004, July. CAMSHIFT Tracker Design Experiments with Intel OpenCV and SAI. In International Mass Spectrometry Conference.

[27] Du, S., Xu, H. and Li, T., 2020. Implementation of Camshift Target Tracking Algorithm Based on Hybrid Filtering and Multifeature Fusion. Journal of Sensors, 2020, pp.1-13.

[28] Almansour, N.A., Syed, H.F., Khayat, N.R., Altheeb, R.K., Juri, R.E., Alhiyafi, J., Alrashed, S. and Olatunji, S.O., 2019. Neural network and support vector machine for the prediction of chronic kidney disease: A comparative study. Computers in biology and medicine, 109, pp.101-111.

[29] Fei, Z., Yang, J. and Lu, H., 2015. Improving routing efficiency through intermediate target based geographic routing. Digital Communications and Networks, 1(3), pp.204-212.