# Content-oriented Multicamera Trajectory Forecasting Surveillance Network System

Xin Qi[1], Toshio Sato[1], Keping Yu[1], Zheng Wen[2], San Hlaing Myint[1], Yutaka Katsuyama[1],
Kazuhiko Tamesue[1], Kiyohito Tokuda[1], Takuro Sato[3]
[1]Global Information and Telecommunication Institute, Waseda University
[2]School of Fundamental Science and Engineering, Waseda University
[3]Research Institute for Science and Engineering, Waseda University

***Abstract*— To reduce safety violations in wide-area ranges, there is a need for highly functional multicamera surveillance systems. We introduce a multicamera trajectory forecasting surveillance network system based on a content-oriented suspicious object network system. This system uses multiple cameras in detection and recognition to track persons among different areas and is capable of retracking people. Each camera node has a processing unit and uses information-centric networking technology to build a content-oriented IoT network. We use field-recorded data to support the simulation, and the evaluation result indicates that our trajectory forecasting method is more efficient than conventional surveillance systems.**

***Keywords—ICN, IoT, content-oriented, surveillance network, trajectory forecasting***

## I. INTRODUCTION

In recent years, public security has become increasingly important. Traffic accidents, terrorist attacks, and deadly disease infections have severely affected the stability of society. To curb public safety and health risks, surveillance systems have been implemented in public areas. A surveillance system plays an essential role before, during, and after the onset of dangerous diseases and crimes.

A conventional surveillance network system uses a point-to-point connection between a central server and a camera or sensor nodes. Unprocessed video data is transmitted in this network. With the increase in camera nodes, the bandwidth capacity of the central server needs to increase at the same time as a simple solution. However, a higher data rate can lead to higher hardware investments and more complex processing coordination.

In this paper, we introduce a content-oriented multicamera surveillance network system with the help of trajectory forecasting [1]. There are several parts described in this paper. We use information-centric networking (ICN) to realize the basic network structure because of the efficiency of the content delivery mechanisms [2]. We also implement a multicamera scenario in the surveillance network to cover multiple areas and track individuals across different areas [3] [4]. Finally, we introduce a simple trajectory forecasting method to the system and conserve network and system resources.

The remainder of this paper is organized as follows. Section II presents preliminary studies, including related research and a larger picture of the overall project this paper is based on. In section III, we present our designed surveillance network system based on the ICN architecture, trajectory forecasting, and handover. Then, in section IV, we evaluate the system in a simulation with field recorded data. Finally, we conclude our work.

## II. RELATED WORKS

### A. Information-centric Networking

In an ICN network, the concept of addresses and hosts is weakened and replaced by data names. ICN packet samples are shown in Figure 1 [2]. A consumer sends an interest packet when requiring named data. An interest packet contains only the content name to be routed to the content producer. Once the interest arrives at the producer, the target data are returned from the producer back to the consumer via the same route. By default, all the router nodes in the data delivery path cache the data in their content store. When other consumers require the same data that match it in the content store, the router node can return the data directly. The named data are addressable, routable, authenticated and irrespective.

Previous ICN-related studies, such as NDN [5], named-node networking (3N) [6], and their simulators ndnSIM [7] and nnnSIM [4], were designed to improve ICN weak points, such as packet loss and delay performance in wireless environments.

| Interest packet | Data packet |
|---|---|
| Content Name | Content Name |
| Selector | Signature |
| Nonce | Signed Info |
| | Data |

Fig. 1. Information-centric Networking.

### B. Data delivery in surveillance network

A conventional surveillance system solution uses a TCP/IP-based video camera network [8], which can review video data by using indices such as timestamps and location information. A central server [9] is used to receive all of the video data and process the data. However, this architecture is insufficient for a large-scale video surveillance network. It uses a centralized principle that limits the scalability for multiple areas.

### C. *Suspicious Object Network System*

With the principle of "ensuring a sufficient security level without stopping the flow of people," we design a suspicious object network system (SONS) to recognize suspicious objects concealed by humans [7]. In the SONS network, we use many kinds of sensors to perform multiple screening processes for each person. When tracking a person across multiple areas, reducing network and computing resources is a challenging task. Figure 2 shows the topology of an SONS. When a person walks into the surveillance area and is detected by a visual light camera within 15 m, the screening process starts. The tracking application starts to generate the person's tracking information and tracks him/her. When the person reaches the 5 m range of the hybrid imager, the W-band active radar imager, the imager can recognize and identify if the person has a suspicious object hidden in his/her clothing. These processes are called the 1st screening process. It is completely managed with no human interaction or intervention. If the 1st screening process indicates suspicious objects hidden on a person, he/she will need to be handled by a secure person into the 2nd screening. The 2nd screening includes conventional and more accurate security checks to confirm whether there is any security risk.

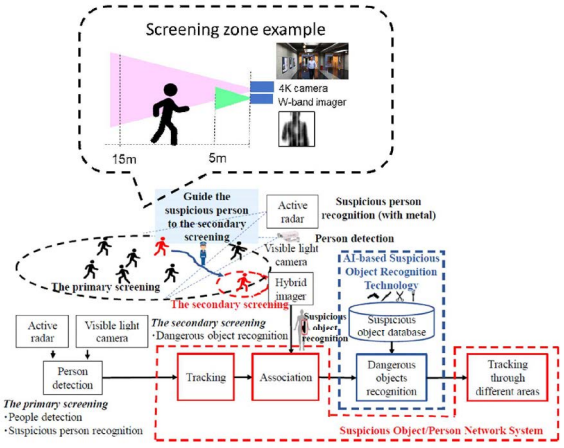



Fig. 2. Suspicious object network system.

## III. SYSTEM DESIGN

To build a properly working surveillance network system, necessary parts need to be designed. These include ICN-based naming and data delivery mechanisms, person trajectory-related data association and trajectory forecasting, and trajectory-based handover mechanisms.

### A. *Naming & Data Delivery*

To efficiently transmit named content in a suspicious object network system, a proper naming system needs to be designed.

The naming mechanism follows a hierarchical order while describing different layers of areas and nodes. An example is shown in Figure 3. The hierarchical order contains the area ID, person ID and data category. The area ID describes which area this content is generated and needs to be retrieved from. The person ID is the local person ID that indicates the ID assigned and used only in a certain area, as defined in the above area ID. The data category indicates different kinds of data related to the person, such as a person's face encoding data or location data.

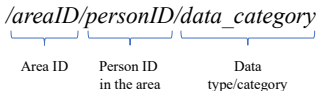



Fig. 3. Example of a content name.

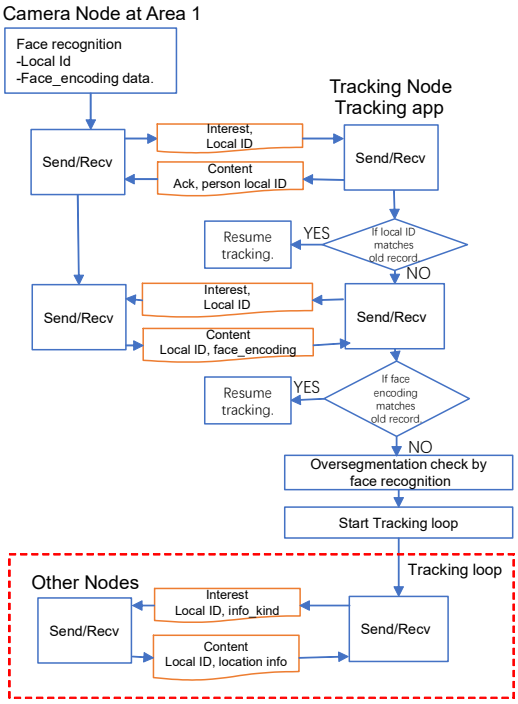



Fig. 4. Data delivery flowchart.

Data delivery for tracking has two basic procedures: start to track a new person and resume tracking a person roaming between areas. The flowchart of the new person tracking is described in Figure 4. When there is a new person detected by the camera node in an area, the camera node sends an interest packet to inform the tracking application. The tracking application then replies to an acknowledgment packet and checks if the person ID is in the record of that area. If not, the tracking application sends interest to ask for the person's face encoding data; otherwise, it resumes tracking for this person. The camera node sends the person's face encoding data to the tracking application, and the tracking application checks the face encoding record if it is already recorded from other areas. If not, this person is recognized as a new person in the whole area, and the tracking application starts to track this person; otherwise, it resumes tracking this person.

The basic procedures of resuming the tracking for a person who was being tracked in the same or another area can also be explained by the flowchart in Figure 4. When a tracked person

is lost in the tracking application, and the camera node reports a person with the same person ID, the tracking application resumes tracking this person. When a tracked person leaves an area and enters another area, the new area camera node reports a new person ID with a face encoding that can match another node's data. The tracking application confirms the face encoding match and resumes person tracking, as this person is roaming from one area to another.

### *B. Data Association & Trajectory Forecasting*

To properly track a person and further perform trajectory forecasting in a surveillance system, it is necessary to generate a valid trajectory for the person. This whole trajectory generating process includes person detection, ID number assignment, and data association. Then, with the proper person's trajectory, trajectory forecasting can take place. Based on the current face detection and face recognition performance, there might be some oversegmentation mistakes during multiple-person tracking. In this subsection, we introduce how to reduce oversegmentation data and generate unified trajectory data for each person.

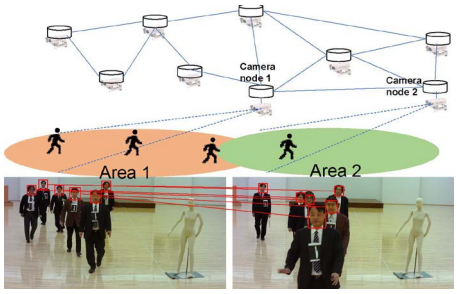



Fig. 5. Face recognition-based person tracking.

Global data association retracks a person who is roaming from one area to another, as shown in Figure 5. The steps are described as follows.

Local data association steps in an area:

a) Face detection for persons in the camera's FOV (Field of View).

b) Face ID number assignment for detected faces.

c) Estimate the location of the person.

d) Track the detected person's trajectory in the area.

When a person enters an area, a camera captures him/her in its FOV. The face detection application detects the person's face and generates this person's local information, including his/her local person ID. Then, the local tracking application can track the person's trajectory within the camera's FOV. During this process, there is a chance that the person is visually blocked from time to time, and the tracking application may lose track of the person. The ID application may assign a new ID to this person, and the new ID and its trajectory may be recognized as a separate person, oversegmenting the same person. To reduce oversegmentation and raise the success re-ID rate, we introduce the following method.

When oversegmentation occurs, face detection and ID application generate multiple info data for the same person. All the oversegmented data referring to the same person means sharing the face data signature. We choose to use a second-layer face recognition application to associate these data and reduce the oversegmentation before further influencing the system. The procedures can be described in the following and are shown in Figure 6, which shows the flowchart and examples of oversegmentation removal and data association between 2 areas. First, we create a table with all the individual info data in it, including the face encoding data. Then, we compare every face encoding with each other to obtain their similarity result. Based on the similarity value, we can determine which groups of faces/IDs refer to the same person and should be associated with one ID. After this series of processes, we can basically reduce most of the oversegmentations.

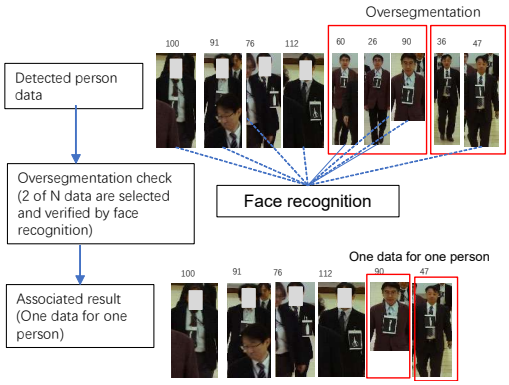



Fig. 6. Face recognition-based person tracking.

The location of a person is defined as the position on the floor plane $(x,y)$. The location $(x,y)$ is estimated by the foot position $(u_b,v_b)$, which is estimated by the face detection results [10]. The perspective transformation matrix A can be shown in an equation:

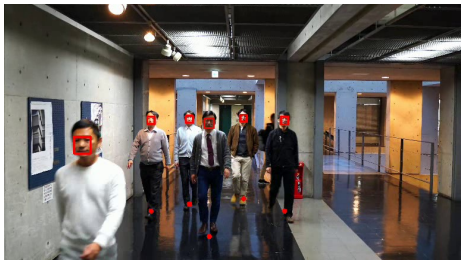
$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = A \begin{bmatrix} u_b \\ v_b \\ 1 \end{bmatrix} \quad (1)$$

which is calculated by four points $(x,y)$ on the floor and $(u,v)$ in the image plane [10]. An example of location is shown in Figure 7.

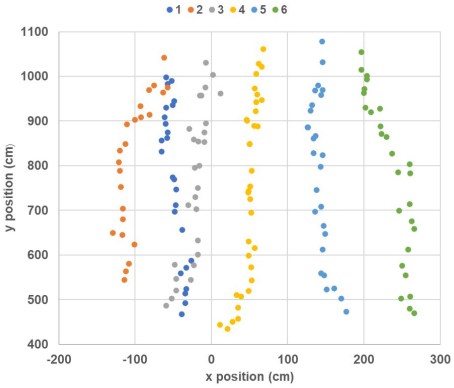
When a person leaves one area and moves into a new area, it is simple enough to request only person info data from all areas, but this is very inefficient and costs considerable bandwidth resources. It is important for the global tracking application to be proactive to know where the person reappears. Therefore, the application can try to request data from the next area at a minimum cost. To forecast the person's trajectory, we must first have information on the person's current trajectory. When he/she moves to a specific location in the camera's FOV, we can predict a handoff procedure to his/her next appearing place.

Based on the physical environment, we can tell where a person is going by his/her trajectory. For example, we can divide a surveillance area into an M*N grid, illustrated as a 4*4 grid in Figure 8. We can look at this forecasting as a categorization problem. If a person moves towards the blue region, he/she is most likely to go directly towards one possible direction, area 4. Otherwise, when a person goes towards the yellow regions in the corners, he/she is likely to move to 3 other connected areas. The blue and yellow regions are marked as trigger regions, which trigger the forecasting system when a person's trajectory leads there.

When trajectory forecasting is performed, possible person reappear area information is created. The handover process takes over.



(a) Surveillance video frame example.




(b) Estimated location and tracking for 6 persons.

Fig. 7. Example of estimation of location and tracking.

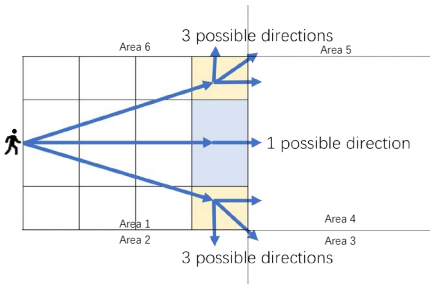



Fig. 8. Trigger region-based trajectory forecasting example.

## C. *Handover*

After a person is first detected and tracked in the tracking system, the person naturally continues his/her movement in the surveillance areas and eventually roams from one surveilled area to another. To regain the track of the same person with a minimum delay and overhead, we introduce the retracking handover.

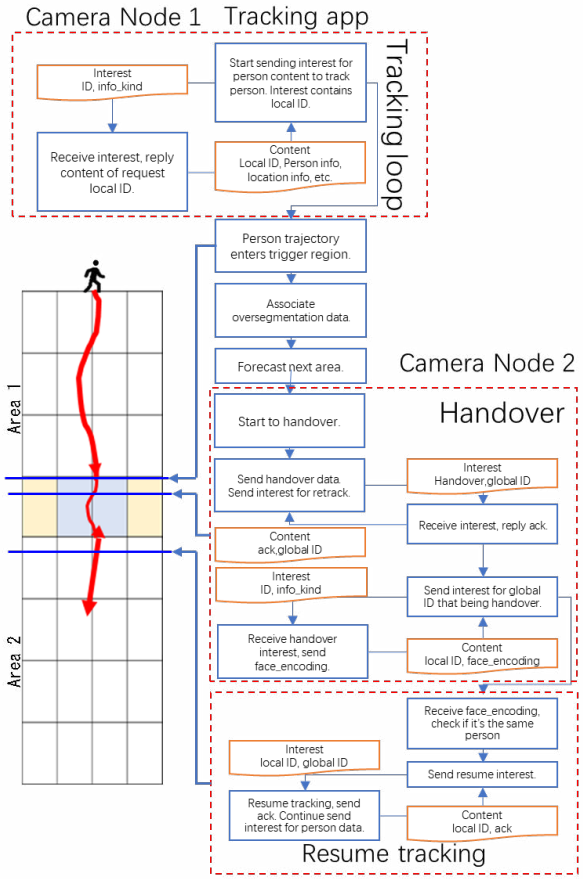



Fig. 9. Handover flowchart.

The handover flowchart is shown in Figure 9. Before a person roams to another area from the current area, we assume the tracking application has already been continuously tracking the person. In a conventional way, retracking a person only takes place after confirming that the person is in the new area. This method introduces delay time. With the help of forecasting the person's trajectory, we take a step ahead to try to retrack the person as soon as he enters the new area.

The retracking handover procedures and background designs are as follows. In this area, we transform the camera FOV to a reality axis. Based on the reality axis, the area is divided into multiple smaller regions. Some regions are close or connected to the exit of this area, and we make these regions trigger regions. As soon as the tracked person enters the trigger region, the tracking application is triggered to start the forecasted handover. First, we associate the person's trajectory in the area and use it and the trigger region's connected exit

information to forecast the next area result. After this process, the tracking application can start to inform the forecasted next area's camera node to report any new person as soon as detection occurs. During this process, tracking in the previous area is still ongoing until the person exits the area and disappears from the camera FOV. Before the next area camera detects the person, the tracking application sends the tracked person's information, including face encoding, to the camera node, so it can not only detect the person but also recognize the person as a retracked target. When the person is detected and recognized by the next area camera node, it informs the tracking application. The tracking application can now retrieve this person's information and retrack him/her once confirmed as the previously tracked person. Now, the handover finishes, and the tracking application continues to track the person.

## IV. EVALUATION

The objective of this paper is to describe the concept and simulation scenario of the content-oriented surveillance network system. The common TCP/IP-based surveillance network usually has multiple streams of video data delivered to a central processing node to process and generate needed content. The whole system is not aware of any persons in the area and cannot track or create the trajectory of a person in many applications.

### *A. Simulation Architecture*

To realize a content-oriented surveillance network with person tracking in multiple areas, we must first recognize and ID a person in an area where the person is first shown. In each camera node, there is a machine-learning-based face recognition module to recognize any appearing person faces. The faces are classified and registered.

Figure 10 shows the simulation structure and a serial person walking scenario. When a person walks from one area to another, he/she is captured in the camera's FOV, and when the person is in the trigger region, the handover process takes place. The raw data taken by the camera nodes in the real environment are used in this simulation. The processes of raw data processing, person detection and recognition, data association, data transmission and handover are all included in the simulation. The tracking application acts as the content consumer in the ICN network. The ICN network utilizes ndnSIM [6] to generate ICN throughput.

Figure 11 shows the visualized trajectory data used in the simulation. There are a total of 6 persons walking in 15 areas. During this time, they roam from one area to another area either in a vertical pattern or a horizontal pattern.

There are 3 kinds of handover interest packet casting methods used in the simulation, as shown in Figure 12. The most basic broadcasting method follows the ICN principle and broadcasts the interest packet to every node to find the content. The other 2 methods use unicast instead of broadcast. The nonforecasting unicast method unicasts interest to 5 physically connected areas. Finally, the forecasting method forecasts the most likely area based on the trigger region and sends interest to either 1 or 3 areas.

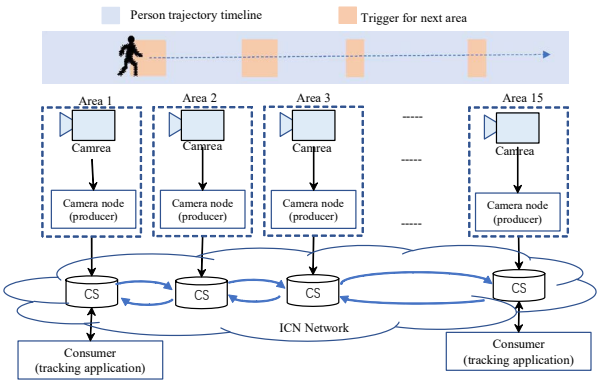



Fig. 10. Simulation structure.

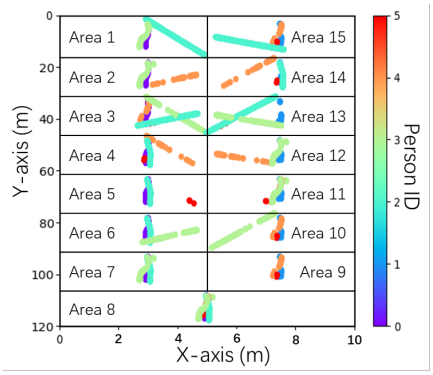



Fig. 11. Trajectory data in the simulation.

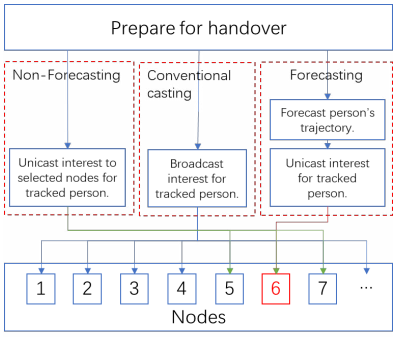



Fig. 12. Interest packet casting methods example.

### *B. Simulation Results*

The simulation provides three kinds of comparison results. The overall bandwidth consumption between the content-oriented and conventional surveillance networks, three interest casting methods based on interest bit rate comparison and their interest hit rate.

Figure 13 shows the overall bandwidth consumption of the conventional video data-based surveillance system and content-oriented surveillance system. In this comparison, the conventional video data consume stable but much more

bandwidth than the content-oriented system. This is because instead of delivering the raw video data to a content processing center, focusing on delivering useful data can greatly reduce network bandwidth consumption.

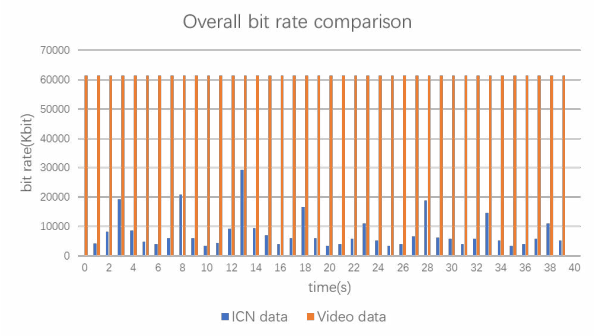



Fig. 13. Overall bitrate comparison.

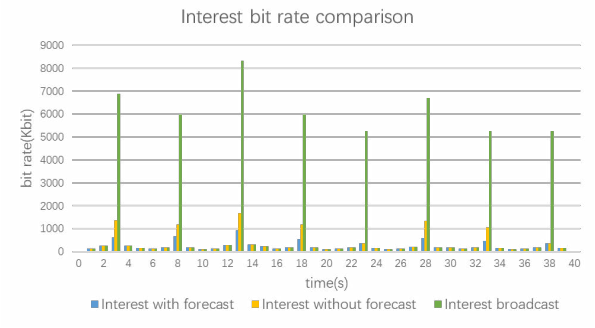



Fig. 14. Interest bit rate comparison.

Figure 14 shows three interest data bit rate comparisons: interest data with forecasting, interest data without forecasting and conventional interest data broadcasting. The differences are at the handover processes with different interest casting methods. Broadcasting interest data during handover costs the most bandwidth because interest data will be forwarded to all the nodes in the network. Unicasting interest data without forecasting costs less bandwidth, but it is still an inefficient method. Forecasting the most likely person reappearing area reduces the most bandwidth consumption. Based on the trigger region's definition, the forecasting method sends 1~3 interests, and the nonforecasting method sends 5 interests.

Table I shows the interest hit rate of each method based on our person trajectory data. When an interest fails to obtain a content, it will be discarded. The hit rate is calculated by dividing interest counts with actual person reappear counts. The three different interest casting methods provide three different interest hit rates. The lowest is interest broadcasting because by broadcasting to more camera nodes, the hit rate decreases. When unicasting interests without forecasting, the hit rate is 20.00% based on our trajectory data. With forecasting the possible person reappearing area to unicast the interest, the hit rate is maintained at a high level.

TABLE I. THE AVERAGE INTEREST HIT RATE

| | **Interest casting methods** | | |
|---|---|---|---|
| | ***Forecasting*** | ***No forecasting*** | ***Broadcast*** |
| Interest hit rate | 88.90% | 20.00% | 6.67% |

## V. CONCLUSION

This paper proposed a multicamera trajectory forecasting surveillance network system with evaluation to support its high interest hit rate and network traffic reduction. In the future work we will focus on a more adaptive trajectory forecasting method to improve forecasting success rate.

## ACKNOWLEDGMENT

This research has been supported by a research grant for expanding radio wave resources (JPJ000254) from the Ministry of Internal Affairs and Communications under contract for "Research and development of radar fundamental technology for advanced recognition of moving objects for security enhancement".

## REFERENCES

[1] Styles, O., Guha, T., Sanchez, V., & Kot, A. (2020). Multi-Camera Trajectory Forecasting: Pedestrian Trajectory Prediction in a Network of Cameras. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 1016-1017).

[2] Zhang, L., Afanasyev, A., Burke, J., Jacobson, V., Claffy, K. C., Crowley, P., ... & Zhang, B. (2014). Named data networking. ACM SIGCOMM Computer Communication Review, 44(3), 66-73.

[3] K. Yu, X. Qi, T. Sato, S. H. Myint, Z. Wen, Y. Katsuyama, K. Tokuda, W. Kameyama, and T. Sato, "Design and performance evaluation of an ai-based w-band suspicious object detection system for moving persons in the iot paradigm," IEEE Access, vol. 8, pp. 81 378–81 393, 2020.

[4] Qi, X., Wen, Z., Tsuda, T., Kameyama, W., Shibata, K., Katto, J., & Sato, T. (2016, December). Content oriented surveillance system based on information-centric network. In 2016 IEEE Globecom Workshops (GC Wkshps) (pp. 1-6). IEEE.

[5] Zhang, L., Afanasyev, A., Burke, J., Jacobson, V., Claffy, K. C., Crowley, P., ... & Zhang, B. (2014). Named data networking. ACM SIGCOMM Computer Communication Review, 44(3), 66-73.

[6] Lopez, J., & Sato, T. (2017). Seamless mobility in ICN for mobile consumers with mobile producers. IEICE Transactions on Communications, 2016EBP3435.

[7] Mastorakis, S., Afanasyev, A., & Zhang, L. (2017). On the evolution of ndnSIM: An open-source simulator for NDN experimentation. ACM SIGCOMM Computer Communication Review, 47(3), 19-33.

[8] M. Systems, "Xprotect vms 2020 r2, getting started guide – single computer installation," available: https://doc.milestonesys.com/.

[9] Wactlar, H. D., Kanade, T., Smith, M. A., & Stevens, S. M. (1996). Intelligent access to digital video: Informedia project. Computer, 29(5), 46-52.

[10] T. SATO et al., "Pedestrian Positioning in Surveillance Video using Anthropometric Properties for Effective Communication," 2020 23rd International Symposium on Wireless Personal Multimedia Communications (WPMC), 2020, pp. 1-6, doi: 10.1109/WPMC50192.2020.9309520.