Evaluation of Person Search System Using Multidirectional Cameras and a Drone Based on Active Inference

Eito Kosuga[1], Shin'ichi Arakawa[1], and Masayuki Murata[1]

Abstract— **In recent years, there has been an increasing expectation to understand and utilize information from physical space. However, the vast amount of information, including historical data, makes it inevitable that some information will be missing in the virtual space. Active inference, based on the free energy principle, is a control framework that captures the uncertainty and ambiguity of information, allowing for the observation of objects and inference of one's own actions. In active inference, an agent probabilistically infers the state of the environment by combining prior beliefs about the state with observational results. Then, the agent estimates the state change for each action as if it is selected and finally selects the most appropriate action. This enables the estimation of the state of the environment and the control of actions to reduce uncertainty within a single framework, even when information is uncertain. In this paper, to address the missing information in virtual space, we aim to appropriately engage sensing in the physical space by applying active inference to the search for individuals using multidirectional cameras and drones. The target of the search continues to move, making continuous sensing difficult and leading to missing information about the target. Therefore, when a multidirectional camera cannot sense the target, it estimates the target's position and controls the capture range of other cameras and the movement direction of drones to more efficiently capture the target. Simulation results show that the time to capture the target is reduced compared to cases where active inference is not used.**

Index Terms—Active Inference, Person Search, Drone, Multidirectional Camera, Free Energy Principle

I. INTRODUCTION

In recent years, there has been an increasing expectation to understand and utilize information from the physical space through sensing. One of the examples is the Cyber-Physical System (CPS). CPS collects data from the physical space using sensing devices, analyzes the big data in the virtual space, and feeds back the results to control and services in the physical space [1]. CPS is utilized in various fields such as data analysis in healthcare and monitoring conditions in factories and farms [2]. Furthermore, by conducting sensing at the city or regional level and analyzing and feeding back the vast amount of collected data, CPS can be applied to more human-centric use cases such as traffic information estimation, lost item search, and the search for missing persons, wandering elderly, or wandering pets [3].

In CPS, it is crucial to accurately grasp the situation in the physical space. However, the number of sensing devices is limited, and it is impossible to sense information from every location in the physical space. To address such missing information, methods such as interpolation and supplementation of missing information are considered. Interpolation of missing information involves estimating the missing information using past data in the CPS [4], [5]. In contrast, supplementation involves conducting additional sensing in the physical space using actuators such as robots, drones, and multidirectional cameras. However, due to the vastness of the physical space, it is inevitable to narrow down the sensing targets for supplementation. Therefore, it is necessary to select appropriate sensing targets while considering the expected information gain, such as the reduction of uncertainty.

Active inference is a control framework that captures the uncertainty of information and infers the observation of objects and one's own actions [6]. In active inference, an agent holds a generative model that generates prior beliefs about the state of the environment. By combining the prior beliefs about the state of the environment output from the generative model with actual observational results, the agent estimates the posterior beliefs about the state of the environment. Furthermore, based on the estimated state, the agent understands the changes in the state due to its own actions, enabling the selection of optimal actions to reduce uncertainty in line with the beliefs about the state of the external world.

In this paper, we evaluate the effectiveness of applying active inference to control sensors and actuators in the physical space while considering the missing information in CPS. As a use case, we assume the search for missing persons or wandering elderly in a city and individually control the observation directions of multidirectional cameras and the movement directions of drones using active inference. By using active inference, it becomes possible to capture the uncertainty of observing the target with adjacent cameras or drones when a camera does or does not observe the target, leading to a more efficient use of actuators. Additionally, we introduce uncertainty in the accuracy of person identification by cameras and drones. We demonstrate that active inference achieves person search while considering the possibility of not identifying a person with a certain probability.

II. RELATED WORK

A. Free Energy Principle

This section explains active inference based on the free energy principle [6], [7]. The free energy principle, proposed by Friston, is an information theory of the brain. It posits

[1]Eito KOSUGA, Shin'ichi Arakawa and Masayuki Murata are with Graduate School of Information Science and Technology, The University of Osaka, Osaka, Japan, e-mail: {e-kosuga, arakawa, murata}@ist.osaka-u.ac.jp

that the perception and learning of living organisms are determined by minimizing a cost function called variational free energy. As a result, organisms, represented as agents, adapt to their environment.

At each time step t, the brain estimates the state of the environment s_t from the observations o_t generated by the environment. Additionally, the agent performs actions that affect the environment. These actions are determined by a policy π, which is a sequence of actions chosen by the agent. The estimation of the state s_t is achieved by minimizing the variational free energy F_t as shown in Equation 1.

$$\begin{aligned} F_t(\pi) &= \mathbb{E}_{Q(s_t|\pi)}[\log Q(s_t|\pi) - \log P(o_t, s_t|\pi)] \\ &= \mathbb{E}_{Q(s_t|\pi)}[\log Q(s_t|\pi) - \log P(o_t|s_t)P(s_t|s_{t-1},\pi)Q(s_{t-1})] \end{aligned} \tag{1}$$

Equation 1 represents variational Bayesian inference, which derives the approximate posterior belief $Q(s_t|\pi)$ to estimate the state s_t. Here, $P(o_t|s_t)$ is the likelihood of the observation o_t given the state s_t, and $P(s_t|s_{t-1},\pi)$ is the state transition distribution. By minimizing the expected free energy shown in the equation, the inference of the policy π for determining actions is performed.

$$\begin{aligned} G(\pi,t) &= \mathbb{E}_{Q(s_t,o_t|\pi)}[\log Q(s_t|\pi) - \log P(o_t, s_t|\pi)] \\ &\simeq -\mathbb{E}_{Q(s_t,o_t|\pi)}[D_{KL}[Q(s_t|o_t,\pi)||Q(s_t|\pi)]] \\ &\quad - \mathbb{E}_{Q(s_t,o_t|\pi)}[\log p(o_t)] \end{aligned} \tag{2}$$

The first term is called information gain. It represents the degree of change in the predicted approximate posterior belief when an observation is obtained. A higher value indicates a higher potential to reduce the uncertainty of the state. Therefore, this term is used for selecting exploration policies by the agent. The second term is called prior preference. It is necessary to define the prior preference as the agent's objective. The prior preference is a term for setting rewards for observations. Thus, this term is used for selecting policies to achieve the agent's objectives. These two terms enable the handling of both exploration for information and the achievement of objectives.

B. Maintaining the Integrity of the Specifications

The study by [8] proposes a mapping method using active inference with a mobile robot. The goal is to understand the map of a warehouse and the route to the target location using images from a camera mounted on the robot. The robot is considered an agent, and its generative model is implemented as a hierarchical model, with the upper model handling navigation and the lower model handling mapping.

In navigation, the start point, goal point, and the agent's pose at the beginning of navigation are given as initial states. In the lower model, no initial state is provided, and the map of the warehouse is created through exploration. However, the state transition probabilities under policy π and the generation probabilities of o_t under state s_t are provided through reinforcement learning by running the robot in the warehouse beforehand.

The warehouse used in the study has many visually similar locations, making the environment and the robot's position ambiguous. Therefore, without using active inference, the map could not be created in a single exploration and was completed by combining the results of three explorations. In contrast, with active inference, the robot infers its position while exploring, allowing the map to be constructed in a single exploration. Additionally, using the created map, active inference successfully derived the shortest path from the start point to the goal point.

However, this study controls only a single sensor (camera) and a single actuator (robot). Coordinating sensors and actuators could enable the application to a wider range of use cases.

III. Person Search System Using Active Inference

In this paper, we apply active inference to a problem of person search using multidirectional cameras and drones. We evaluate the effectiveness of analyzing information in the virtual space and efficiently controlling multidirectional cameras and drones in the physical space, considering the uncertainty in the virtual space of CPS. Below, we describe the structure of the person search system.

A. System Overview

The free energy principle allows the generative model to represent the state of the external world as beliefs. The generative model and beliefs can be updated by minimizing the variational free energy from the obtained observations. Using these expressed beliefs, active inference can take actions to reduce the uncertainty of those beliefs. Therefore, in CPS, it is expected to represent the uncertainty caused by missing or aggregated information in the virtual space as beliefs and realize actions and observations that reduce that uncertainty.

In person search, it is difficult to continuously sense the target, and uncertainty arises due to missing information about the target's location in the virtual space. By utilizing the free energy principle, it is possible to represent the target's location as beliefs based on past sensing results. Thus, active inference can plan actions to reduce the uncertainty of the target and appropriately control multidirectional cameras and drones in the physical space to advance the person search.

The overview of the person search system using active inference with multidirectional cameras and drones is shown in Figure 1. We suppose a city where multidirectional cameras and drones are placed as the environment. When a person search request is made, the system uses the video from the cameras and drones, along with the information aggregated in the virtual space of CPS, to search for the target moving through the city. In active inference, the agent is considered to have a generative model on a server that constructs the virtual space, along with cameras and drones.

First, the cameras and drones observe the presence or absence of the target at their observation positions and send this information to the server. The server then infers the target's location by calculating the variational free energy

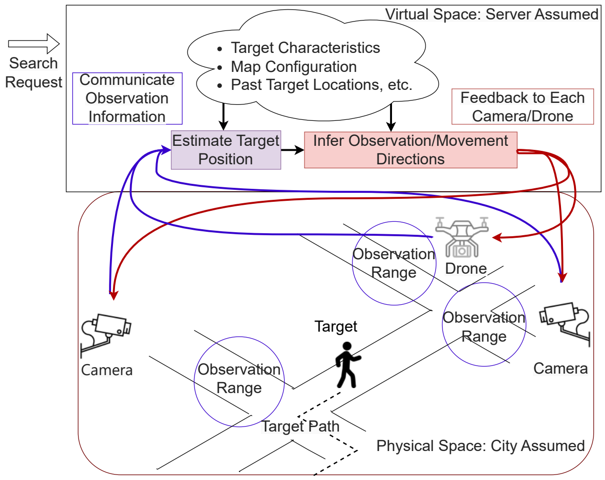

Fig. 1. System Overview Diagram

based on prior beliefs about the target's location, the obtained observations, and the policy. The posterior estimate at this time becomes the prior probability for the target's location at the next step. Furthermore, by calculating the expected free energy from the posterior estimate of the target's location, the policy for changing the observation direction of each camera and the position of the drones is inferred.

Finally, based on the inferred policy, the observation direction of each camera and the movement direction of the drones are determined, and commands are sent from the server to the multidirectional cameras and drones. The multidirectional cameras change their observation direction, and the drones move and then observe at the next time step. By repeating this process, active inference can realize action plans that reduce the uncertainty about the target's location, advancing the person search.

B. System Details

This section explains the details of the person search system. To develop and evaluate the proposed system, we make the following assumptions:

- No Individual Identification: The system detects the presence or absence of a person but does not distinguish between individuals.
- Predefined Map: The map is pre-prepared and stored on the server, providing known movement paths for the target.
- Movement Tendencies: The target follows probabilistic movement tendencies, such as a preference for going straight at intersections, but does not adapt to environmental factors dynamically.
- Camera Recognition Errors: Cameras may fail to detect the target with a certain probability, introducing observation uncertainty, which is accounted for in the active inference framework.

To perform person search by identifying individuals through images, the task known as Person Re-Identification (Person Re-Id) is crucial. This task involves linking a person captured by one camera to the same person captured by another camera [9]. Research on Person Re-Id has predominantly focused on processing images using deep learning. Methods have been proposed to achieve Person Re-Id in environments like the remote surveillance systems assumed in this study [10].

People moving through the city have movement tendencies based on their characteristics such as destination, age, and gender [11]. This includes preferences for wider roads over narrow ones, brighter roads over darker ones, and roads with better visibility over those with poor visibility. This paper simplifies these tendencies as the probability of going straight at intersections. These assumptions do not affect the core objective of this study, which is to evaluate the effectiveness of active inference in controlling multidirectional cameras and drones for efficient target search.

The map is represented as a graph $GR(N,V)$, where intersections are nodes and paths between intersections are links. The connection matrix C is defined, where the element $P_{i,j}$ represents the number of paths between intersections i, j. Cameras are installed at some intersections, denoted as $N_c \subset N$, and the connection matrix between cameras C_c is derived from C and N_c.

Each camera i at time t has a direction $a_t^{(i)}$ $(1 \leq i \leq N_c)$. Drones can move up to two links per time step, with actions $a_t^{(drone)}$ including stopping or moving up, down, left, or right. The agent infers the camera direction vector as $a_t = (a_t^{(1)}, \ldots, a_t^{(N_c)}, a_t^{(drone)})$.

The state s_t consists of the target's position, the directions of the cameras, and the drone's position. Each camera i at time t observes $o_t^{(i)}$, indicating the presence or absence of the target. Similarly, the drone captures observations $o_t^{(drone)}$. Thus, the observation vector is represented as $o_t = (o_t^{(1)}, \ldots, o_t^{(N_c)}, o_t^{(drone)})$.

The target moves probabilistically based on a **straight probability**, which determines the likelihood of continuing straight at an intersection. If the straight probability is close to $\frac{1}{3}$, movement is nearly random, while a value closer to 1 indicates more deterministic movement. The likelihood $P(o_t|s_t)$ represents the probability of obtaining an observation o given state s, primarily indicating the presence of the target in the field of view of cameras and drones.

Next, we explain an active inference method for person search problems. The estimation $Q(s_t|\pi)$ of the state s_t, including the target's position, is derived by minimizing the variational free energy F_t based on Equation 3.

$$
\begin{aligned}
F_t(\pi) &= \mathbb{E}_{Q(s_t|\pi)}[\log Q(s_t|\pi) - \log \hat{P}(o_t|s_t)P(s_t|s_{t-1},\pi)Q(s_{t-1})] \\
&= \mathbb{E}_{Q(s_t|\pi)}[\log Q(s_t|\pi) \\
&\quad - \log \alpha P(o_t|s_t)P(s_t|s_{t-1},\pi)Q(s_{t-1})] \\
&= \mathbb{E}_{Q(s_t|\pi)}[\log Q(s_t|\pi) - \log P(o_t|s_t)Q(s_{t-1}) \\
&\quad - \log \alpha P(s_t|s_{t-1},\pi)]
\end{aligned} \tag{3}
$$

Here, $\hat{P}(o_t|s_t)$ represents the likelihood of observing the

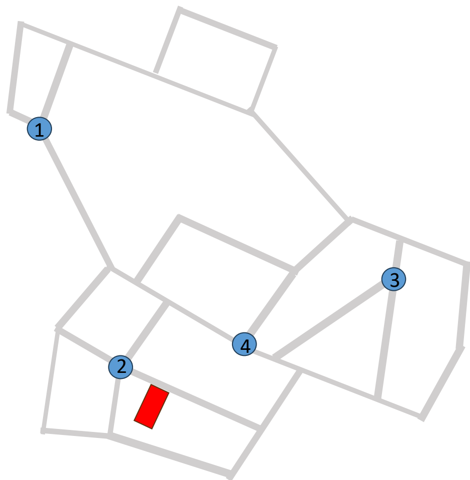

Fig. 2. Example of the Map Used

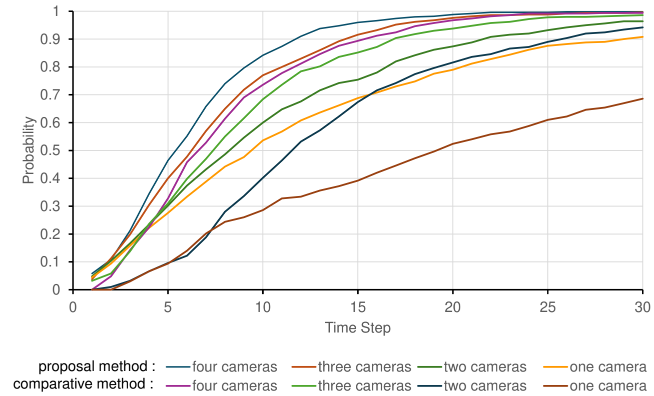

Fig. 3. Capture Time Steps by Number of Cameras (Straight Probability 0.33)

target, incorporating recognition errors, modeled as $\hat{P}(o|s) = \alpha P(o|s)$ with error parameter α. Since $0 < \alpha < 1$, observation errors increase the uncertainty in the posterior distribution $Q(s_t|\pi)$, but the model accounts for this uncertainty.

Next, to plan the future changes in the observation directions of the multidirectional cameras π, the expected free energy $G(\pi,t)$ is calculated based on Equation 4 based on the posterior belief $Q(s_t)$, the target's transition state $P(s_t|s_{t-1})$, the observation likelihood $P(o_t|s_t)$, and the prior preference $p(o_t)$.

$$\begin{aligned} G(\pi,t) = &-\mathbb{E}_{Q(o_t|\pi)}[D_{KL}[Q(s_t|o_t,\pi)||Q(s_t|\pi)]] \\ &-\mathbb{E}_{Q(o_t|\pi)}[\log p(o_t)] \end{aligned} \tag{4}$$

By selecting actions that minimize $G(\pi,t)$, the system efficiently reduces uncertainty in the target's position, making it easier to capture the target. The updated posterior belief $Q(s_t)$ is used as the prior $P(s_t)$ for the next step, iterating this process to conduct person search.

IV. Evaluation

A. Physical environments

An example of the map used in this paper is shown in Figure 2. Figure 2 is created based on the Suita Campus of Osaka University [12]. The gray areas represent roads, indicating where the target can move. The cameras are numbered in the order they are added in Figure 2, with a blue circle labeled “1” indicating one camera, “1” and “2” indicating two cameras, and so on. The drones start their search from the red square shown in Figure 2.

Through observation, the presence or absence of the target is determined. When the target is present within the observation range, the target's position is obtained as the observation value. If the target is not present within the observation range, the observation value is treated as NULL.

B. Simulation Settings

In this paper, we evaluate whether active inference can efficiently capture the target by controlling multidirectional cameras and drones while estimating the target's movement rules. This simulation assesses the time taken to capture the target with a drone from an unknown initial position. The rationale for using drones is that, even if the target is captured by cameras or other sensors, it is easy to lose track of the target's subsequent movements. In contrast, capturing the target with a drone allows continuous tracking, independent of the sensor placement along the target's path. This makes it easier to protect wandering elderly or lost children.

For evaluation, we use both the proposed method and a comparative method. The comparative method involves capturing the target from the nearest candidate position, which is a potential location where the target might be. This method maintains possible branching positions as candidates each time the target enters an intersection. By prioritizing the capture of the candidate with the highest probability, it is expected to capture the target in fewer attempts. However, since many candidates may have the same probability, the nearest candidate is captured first.

The simulation starts the target from a random position and evaluates the time steps taken to capture the target with a drone, repeating this process 500 times. The simulation evaluates different patterns of the target's straight probability at intersections to verify if the target's straight probability can be incorporated into the action plan. Specifically, the straight probabilities are 0.33, assuming random movement, and 0.9, assuming rule-based movement. The number of cameras is changed from 1 to 4 to verify the time taken to capture the target by integrating information from cameras.

Additionally, to evaluate the observation error of the cameras, the probability of successful observation is varied. This probability, defined as the observation success rate, indicates the likelihood of obtaining information about the target's presence within the observable range. The observation success rates are varied in four patterns: 0.33, 0.66, 0.9, and 1.0. The proposed method and the comparative method are evaluated with four cameras and two patterns of the target's straight probability (0.33 and 0.9).

C. Results

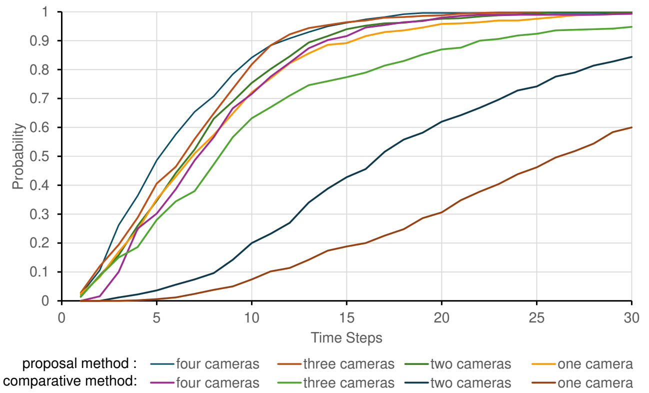

Fig. 4. Capture Time Steps by Number of Cameras (Straight Probability 0.9)

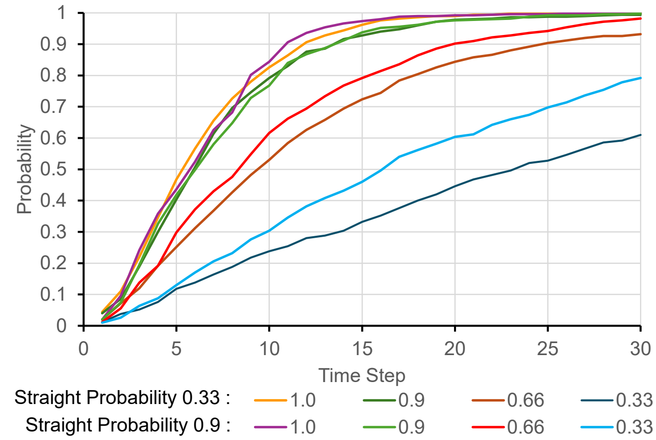

Fig. 5. Capture Time Steps by Observation Success Rate

1) Evaluation Based on the Number of Cameras and Target's Straight Probability: Figure 3 shows the cumulative frequency distribution of the time steps to capture the target when the straight probability at intersections is 0.33 (random assumption). The horizontal axis represents the time steps x, and the vertical axis represents the cumulative frequency distribution of the probability of capturing the target within x time steps. The proposed method shows a decrease in the number of time steps to capture the target as the number of cameras increases.

Notably, there is a significant change from two to three cameras, with the time steps decreasing from 10.894 to 7.776. This suggests that the proposed method efficiently integrates information from each camera to capture the target. However, the change from 3 to 4 cameras is minimal, with the time steps decreasing only slightly from 7.776 to 6.756. The comparative method performs worse, because the probability of the presence of the target for each candidate is not known precisely, so the search is actually performed sequentially from the nearest candidate location, including candidates with low probability.

Similarly, Figure 4 shows the cumulative frequency distribution of the capture time steps when the straight probability is 0.9 (assumed tendency). In this case, there is little change in the number of cameras on the graph, and the average time does not vary significantly. When the target has a tendency, the proposed method estimates the target's position quite accurately, resulting in less variation with the number of cameras.

As with the straight probability of 0.33, the time steps for 3 cameras are 7.18, and for 4 cameras, they are 6.674, showing little difference. This suggests that with a sufficient number of cameras, the information obtained does not differ much. Therefore, in a simplified target tracking scenario using the map based on Suita Campus (Figure 2), having about 3 cameras is sufficient for exploration regardless of the target's tendency.

In the comparative method, the time steps for 4 cameras are almost equivalent to the time steps for 1 camera using active inference, with values of 8.524 and 8.576, respectively. When the straight probability is 0.9 (indicating a tendency), the time to capture the target is longer compared to when the straight probability is 0.33. This is because the comparative method cannot utilize the target's straight probability tendency, leading to more cases where it searches locations with low probabilities of the target's presence. Additionally, a higher straight probability means the target is less likely to loop around the same location, increasing the difficulty of drone-based search.

2) Evaluation Based on Observation Success Rate: Figure 5 shows the cumulative frequency distribution of the time steps to capture the target when varying the observation success rate. The horizontal axis represents the time steps x, and the vertical axis represents the cumulative frequency distribution of the probability of capturing the target within x time steps. When the observation success rate is high (0.9, 1.0), there is little difference in the time steps to capture the target regardless of the straight probability. However, when the observation success rate is lower (0.33, 0.66), the time steps are shorter when the straight probability is 0.9. Specifically, when the observation success rate is 0.33, the average time steps are 31.45 for a straight probability of 0.33, compared to 20.854 for a straight probability of 0.9. This indicates that when there are significant observation errors, utilizing the target's tendencies allows the proposed method to capture the target more efficiently.

3) Evaluation Based on Information Entropy: From the results in IV-C.1 and IV-C.2, it is evident that the proposed method, by controlling multidirectional cameras and drones, can efficiently capture the target. This is particularly effective when the target's tendencies can be utilized. However, this paper aims to address the uncertainty caused by missing information in the virtual space by appropriately engaging sensing in the physical space. Therefore, it is necessary to confirm that the actions generated by the proposed method actually reduce information uncertainty.

To this end, we evaluate the reduction in information entropy, or mutual information, through repeated observations and actions from the start of the search to the capture of the target. Mutual information is expressed as $I(X;Y) =$

TABLE I
AVERAGE MUTUAL INFORMATION PER ACTION/OBSERVATION

Straight Probability	0.33	0.9
Proposed Method	0.9960	0.8251
Comparative Method	0.9785	0.7582

$H(X) - H(X|Y)$. Here, $H(X)$ is the uncertainty of a random variable, and $H(X|Y)$ is the remaining uncertainty in X after knowing Y. Thus, $I(X;Y)$ represents the reduction in X's uncertainty by knowing Y. In this scenario, it indicates how much the uncertainty in the target's position X is reduced by obtaining observation Y. Therefore, the higher the mutual information, the higher the value of the action, indicating that actions are being generated to reduce uncertainty.

We evaluate the proposed and comparative methods with four cameras and two straight probability patterns (0.33 and 0.9), repeating the search process 500 times. The average mutual information reduced per observation and action is calculated by dividing the total mutual information by the number of time steps taken for the search.

Table I shows the average mutual information reduced per action/observation just before capturing the target. The results indicate that for both the assumed straight probability of 0.9 (with a tendency) and 0.33 (without a tendency), the proposed method achieved a higher reduction in mutual information per action/observation. This means that each action/observation in the proposed method has a higher value, demonstrating that the proposed method generates actions that more effectively reduce uncertainty.

V. CONCLUSIONS

This paper presented an active inference for the method to collect information in physical space by efficiently using sensors and actuators. As a use case, we evaluated a system using active inference to search for a person in a city using multidirectional cameras and drones. The results showed that the target of the search system could be captured more efficiently using the proposed method compared to not using it, indicating that the system could appropriately judge the missing information. The effectiveness of the system was also demonstrated in the presence of observation errors.

In this paper, the movement tendencies of the target were simplified in a stochastic manner, but more realistic behaviors should be handled. Our next step is to use human flow data or mobility models, and also to show the effectiveness of our method under more complex scenarios such as the case where persons should be tracked. While this paper focuses on person search, our method has a possibility to apply other scenarios requiring information gathering in urban areas, such as traffic information estimation and incident/accident detection. Examining other use cases and generalizing the framework of the active inference method is also left our future research topic.

ACKNOWLEDGMENT

This paper is based on results obtained from a project, JPNP22007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

REFERENCES

[1] S. Zanero, "Cyber-physical systems," *Computer*, vol. 50, no. 4, pp. 14–16, Apr. 2017.

[2] D. G. Pivoto, L. F. de Almeida, R. da Rosa Righi, J. J. Rodrigues, A. B. Lugli, and A. M. Alberti, "Cyber-physical systems architectures for industrial internet of things applications in industry 4.0: A literature review," *Journal of Manufacturing Systems*, vol. 58, pp. 176–192, Jan. 2021.

[3] Cabinet Office, Government of Japan, "Society 5.0," https://www8.cao.go.jp/cstp/society5_0/index.html, 2024, (Accessed: 2025-01-25).

[4] X. Tang, K. Yang, H. Wang, J. Wu, Y. Qin, W. Yu, and D. Cao, "Prediction-uncertainty-aware decision-making for autonomous vehicles," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 4, pp. 849–862, July 2022.

[5] Z. Pang, J. Li, P. Tokmakov, D. Chen, S. Zagoruyko, and Y.-X. Wang, "Standing between past and future: Spatio-temporal modeling for multi-camera 3d multi-object tracking," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2023, pp. 17 928–17 938.

[6] K. Friston, F. Rigoli, D. Ognibene, C. Mathys, T. Fitzgerald, and G. Pezzulo, "Active inference and epistemic value," *Cognitive Neuroscience*, vol. 6, no. 4, pp. 187–214, May 2015.

[7] R. Smith, K. J. Friston, and C. J. Whyte, "A step-by-step tutorial on active inference and its application to empirical data," *Journal of Mathematical Psychology*, vol. 107, pp. 1–60, Apr. 2022.

[8] O. Çatal, T. Verbelen, T. Van de Maele, B. Dhoedt, and A. Safron, "Robot navigation as hierarchical active inference," *Neural Networks*, vol. 142, pp. 192–204, May 2021.

[9] M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. Hoi, "Deep learning for person re-identification: A survey and outlook," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 6, pp. 2872–2893, Jan. 2021.

[10] C. Wu, W. Ge, A. Wu, and X. Chang, "Camera-conditioned stable feature generation for isolated camera supervised person re-identification," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, June 2022, pp. 20 238–20 248.

[11] H. Zhang, Y. Liu, C. Wang, R. Fu, Q. Sun, and Z. Li, "Research on a pedestrian crossing intention recognition model based on natural observation data," *Sensors*, vol. 20, no. 6, p. 1776, Mar. 2020.

[12] The University of Osaka, "View – detailed Suita campus map," https://www.osaka-u.ac.jp/ja/access/files/campus_map_suita_A4_202002.pdf/@@download/file, 2020, (Accessed: 2025-01-23).