IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 April 2022, accepted 26 May 2022, date of publication 29 June 2022, date of current version 8 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3187027

Optimal Viewpoint Selection by Indoor Drone Using PSO and Gaussian Process With Photographic Composition Based on KL Divergence

TAISEI YOKOMATSU AND KOSUKE SEKIYAMA
Department of Mechatronics Engineering, Graduate School of Science and Engineering, Meijo University, Nagoya 468-8502, Japan
Corresponding author: Kosuke Sekiyama (sekiyama@meijo-u.ac.jp)

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT This study processes an autonomous indoor drone photographer that searches for and selects a heuristic optimal viewpoint to obtain a well-composed photograph of a group of subjects. The subjects on the drone's camera screen are represented by a Gaussian mixture model. When there are four or more subjects, they are represented by a Gaussian mixture model with clustering by variational Bayes. The Kullback–Leibler divergence is evaluated between the Gaussian mixture model and a user-defined reference composition, and it is defined as the composition evaluation value. The reference composition is pre-set by the user based on the basic composition rules, such as the three-section method. The drone searches for a viewpoint in a 3D space to optimize the composition evaluation value using particle swarm optimization (PSO). A Gaussian process is used to facilitate the PSO search. This enables the drone to significantly reduce the search time and successfully capture a photograph with a well-balanced composition.

INDEX TERMS Autonomous drone, robot photographer, real-time search, Gaussian process, particle swarm optimization.

I. INTRODUCTION

One of the typical applications of the drones is to capture photographs or video from the air. Industrial drones have been used for broadcasting sports events or inspecting construction sites, such as bridges, which are difficult for humans to access. In addition, photography drones are being increasingly used for hobbies owing to a reduction in their size and price. Drones have made it possible to obtain high-quality aerial images at a low cost [1]. In addition, recent progress in Artificial Intelligence (AI) techniques has allowed for autonomous flight, which is used to identify and track specific people and objects using a camera mounted on a drone [2]. However, current photography drones still require a skilled operator to select the best shooting viewpoint and time from an aesthetic perspective. Autonomous photographer will be applicable to wide range of scenes such sports broadcast and filming as well as hobby use. In this work, we do not argue what makes a photograph aesthetic but evaluate the balance of composition in a photograph displayed on a camera screen. An example of a well-balanced composition has been discussed in a previous work on the robot photographers. However, it is limited to the evaluation function corresponding to *the rule of thirds*, which is the conventional composition rule [3].

Therefore, in this paper, we propose an optimal viewpoint selection method for an indoor photography drone with the aim of capturing an image with a well-balanced composition in the 3D space. As GPS is not available for the indoor drone, visual SLAM is employed for the position control. A photograph of subjects is obtained, and its composition is evaluated using a 2D Gaussian mixture model.

The associate editor coordinating the review of this manuscript and approving it for publication was Laxmisha Rai.

The variational Bayes method is used to cluster of the group of subjects as a Gaussian mixture model. The mean vectors and the covariance matrices of each mixture element are updated according to the change in the position and viewpoint of the drone. The reference composition is also a 2D Gaussian mixture model; however, its layout can be defined according to a user's preferences to obtain a user-specific photo.

In our previous work, we have developed a ground mobile robot photographer with a similar scheme, where the Kullback–Leibler (KL) divergence is employed as a composition evaluation function based on an image-based 2D Gaussian mixture model and that of the user-defined reference composition [4]. The robot explores a shooting viewpoint such that the KL divergence is minimized. In this study, this approach is applied to the drone photographer, but significant extension is required.

To the best of the authors' knowledge, the conventional work on robot photographers is limited to the ground robots [5]–[8]. In this study, the use of a drone allows us to obtain photographs from more viewpoints compared to previous studies. However, the problem with this method is that considerable time is required to evaluate the composition a photograph obtained in a 3D search space.

A number of methods have been studied for efficiently finding the optimum viewpoint. In this study, particle swarm optimization (PSO) [9] is employed, which is one of a metaheuristic algorithm. PSO is derived from the behavior of a flock of birds finding food [10], [11], and it is a multipoint distributed search algorithm that converges rapidly. However, the merit of PSO diminishes when a single drone explores a space. Optimization methods that employ Gaussian processes have been used in various fields [12]. The combination of PSO and a Gaussian process is an effective approach [13]. A Gaussian process is a nonlinear regression model that estimates an unsearched area using PSO. Hence, it is expected to predict the photographic composition of an entire 3D space using a small number of observations and reduce exploration time.

II. PROPOSED METHOD

The proposed system is as shown in Fig.1. The system consists of 5 process modules.

In the *Image Processing* module, an image is acquired by a drone camera. A subject in the image is detected via object recognition using You Only Look Once (v3 Darknet). The camera information acquired by the image-processing module is used for self-position estimation and composition evaluation.

The *Position Estimation* module estimates the position of the drone using visual SLAM. An Augmented Reality (AR) marker is used to convert a coordinates to the horizontal coordinate system.

The *Composition Evaluation* module acquires the coordinates and size of a subject detected on the camera screen and evaluates the degree of similarity between the photographic composition shown on the drone camera and

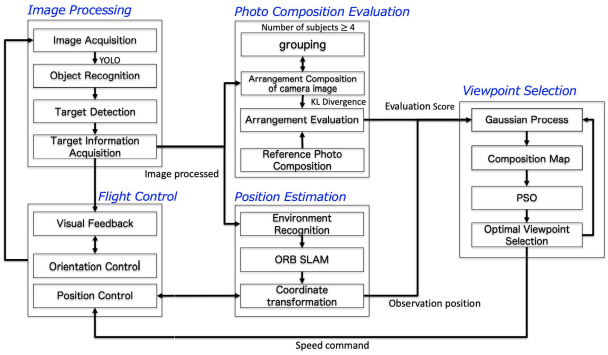


FIGURE 1. System architecture of the drone photographer.

the predefined reference composition based on the KL divergence.

The *Viewpoint Selection* module implements a search and optimal viewpoint selection by combining PSO and a Gaussian process. The Gaussian process estimates the composition function value given by the KL divergence in an unsearched area; thus, a smooth evaluation field is created. PSO is applied in this area.

The *Flight Control* module is used for position control, and it is based on ORB-SLAM. After a subject is detected by the image-processing module, visual feedback control is performed so that the subject does not move outside the frame of the camera screen.

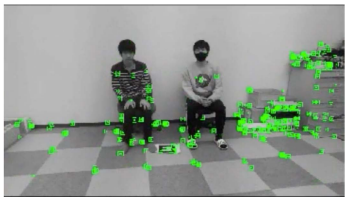


FIGURE 2. Self-position estimation by ORB-SLAM.

III. POSITIONING OF INDOOR DRONE USING ORB-SLAM

A. SELECTION OF SLAM

Self-position estimation in a non-GPS environment is necessary for positioning an indoor drone [14]. An indoor mapping system based on geographic information system is available; however, it is expensive. As an alternative approach, we adopt ORB-SLAM [15], which is based on the feature points in visual SLAM. Visual SLAM is a method to detect the landmarks from camera images and grasp the drone position and surrounding 3D position information. Since the distance scale is not identified only with a monocular camera mounted on the drone, scaling the map with a known object size such as an AR marker is required to improve the accuracy of the map. An example of ORB-SLAM is shown in Fig.2. ORB-SLAM has the highest accuracy among visual SLAM techniques, and it can be operated in real time. Furthermore, it is and open source and easy to implement.

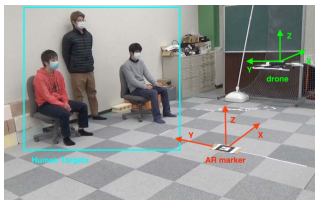


FIGURE 3. Matching coordinate system of the drone.

B. TRANSFORMATION OF SLAM COORDINATES

When visual SLAM is applied to the drone camera, the height from the ground is unknown because the position and orientation estimated by ORB-SLAM is in the local coordinate system. To address this, the local coordinate system of the camera viewpoint is transformed to the global coordinate system with the floor surface as the origin, which is marked by an AR marker. The AR marker is also used to adjust the scale of the SLAM coordinate system to the physical space. Fig.3 shows the position and the coordinate system of the AR marker setting for the experiment.

Let the drone's position in the local coordinate system be ${}^{local}\boldsymbol{p} \in \mathbb{R}^3$, and that in the global coordinate system be ${}^{global}\boldsymbol{p} \in \mathbb{R}^3$. The coordinates systems are right handed, where the z-axis represents the vertical upward direction, as shown in the Fig.3. The rotation matrix for convertiong the drone's local coordinate system to the global coordinate system, ${}^{global}_{local}\boldsymbol{R} \in \mathbb{R}^{3\times3}$, is expressed below, where (ϕ, θ, ψ) are the angles of rotation of the drone around the X, Y, and Z axes, respectively.

$$ {}^{global}_{local}\boldsymbol{R} = \begin{bmatrix} c_\theta c_\psi & s_\phi s_\theta c_\psi - s_\psi c_\phi & s_\phi s_\psi + s_\theta c_\phi c_\psi \\ s_\psi c_\theta & c_\phi c_\psi + s_\phi s_\theta s_\psi & s_\theta s_\psi c_\phi - s_\phi c_\psi \\ -s_\theta & s_\phi c_\theta & c_\phi c_\theta \end{bmatrix} \quad (1) $$

The global coordinate system is expressed as follows using (1) and the translation vector, ${}^{global}_{local}\boldsymbol{t} \in \mathbb{R}^3$:

$$ {}^{global}\boldsymbol{p} = {}^{global}_{local}\boldsymbol{R}\,{}^{local}\boldsymbol{p} + {}^{global}_{local}\boldsymbol{t} \quad (2) $$

This a coordinate transformation is updated sequentially, and the drone is controlled using the correct position and orientation in the global coordinate system.

IV. COMPOSITION EVALUATION USING KL DIVERGENCE

A. COMPOSITION EVALUATION FUNCTION

A subject in the drone camera is represented by a 2D Gaussian distribution in the pixel coordinates of the image. The arrangement of multiple subjects can be expressed as a 2D mixed Gaussian distribution. The photographic composition is evaluated by estimating the similarity between this 2D mixed Gaussian distribution and the 2D mixed Gaussian distribution with the reference composition using KL divergence. The drone searches for the viewpoint in the 3D space to minimize the KL divergence.

Suppose the number of human subjects is L. The center coordinates of the subject l are expressed by a vector, $\boldsymbol{\mu}_l = (x_l, y_l)^T$. The width and height of the area are expressed as a variance–covariance matrix, $\boldsymbol{\Sigma}_l \in \mathbb{R}^{2\times2}$. The 2D mixed Gaussian distribution, $P(\boldsymbol{a}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, of the L subjects at an arbitrary position, $\boldsymbol{a} = (x, y)^T$, on the camera screen is defined below, where the weight of the subject l is π_l and $\sum_{l=1}^{L}\pi_l = 1, 0 \le \pi_l \le 1$. Let the set of the mean vector and the variance–covariance matrix be $\boldsymbol{\Sigma} = \{\boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_L\}$, $\boldsymbol{\mu} = \{\boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_L\}$.

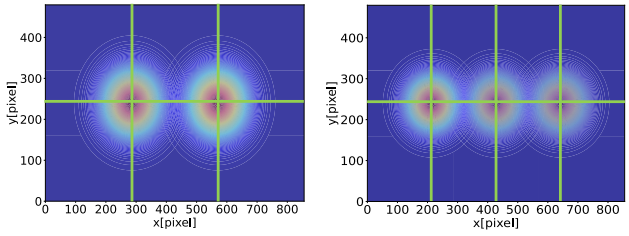
$$ \begin{aligned} P(\boldsymbol{a}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \sum_{l=1}^{L}\pi_l \mathcal{N}(\boldsymbol{a}|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l) \\ &= \sum_{l=1}^{L}\frac{\pi_l}{2\pi|\boldsymbol{\Sigma}_l|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{a}-\boldsymbol{\mu}_l)^T\boldsymbol{\Sigma}_l^{-1}(\boldsymbol{a}-\boldsymbol{\mu}_l)\right\} \end{aligned} \quad (3) $$

The composition evaluation value is given by the KL divergence for the distribution (3) and the user-defined reference composition distribution, $Q(\boldsymbol{a}; \boldsymbol{\mu}', \boldsymbol{\Sigma}')$, on the screen, which is defined as follows:

$$ D^c = -\iint_H P(\boldsymbol{a}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\log\frac{Q(\boldsymbol{a}|\boldsymbol{\mu}', \boldsymbol{\Sigma}')}{P(\boldsymbol{a}|\boldsymbol{\mu}, \boldsymbol{\Sigma})}dH. \quad (4) $$

where H is the range of the pixel screen in the photograph. The photographic composition improves as $D^c \ge 0$ decreases. The calculation of the KL divergence in a multivariate mixed Gaussian distribution requires an approximation of the distribution, and it is given in [16], [17]. Based on this approximation, the KL divergence is calculated in accordance with a previous work [4].

Next, we perform an experiment on the composition evaluation. Fig.4(a) shows the reference composition in the case of 2 subjects. Figs.5(a) and 5(b) are the images that evaluate the composition based on the reference composition of Fig.4(a). The results calculated using the (4) are shown in the upper left of each image, where $D^c = 6.53$ for Fig.5(a) and $D^c = 1.07$ for Fig.5(b). As shown in these figures, the case where D^c is smaller shows a better composition in terms of the reference composition of Fig.4(a).



(a) $L = 2$ (b) $L = 3$

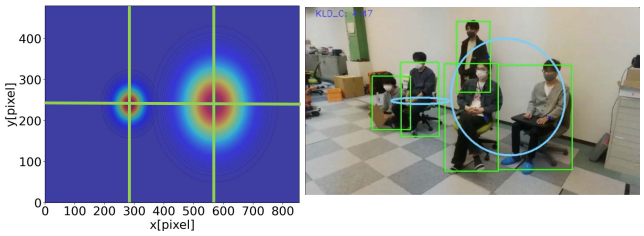
FIGURE 4. Reference composition.

B. CLUSTERING SUBJECTS BASED ON VARIATIONAL BAYES

The number of classes for the reference 2D mixed Gaussian distribution is set as $K = 3$ for simplicity. If there are numerous subjects, the group of (3) is clustered and represented as a single mixture element.

(a) $D^c = 6.53$ (b) $D^c = 1.07$

FIGURE 5. Composition evaluation value D^c.



(a) Composition with a sense of perspective (b) Photo with a reference of Fig.6(a)

FIGURE 6. Changes in camera viewpoint due to changes in reference composition.

Therefore, in this work, we use variational Bayes to classify all the subjects into $1 \sim 4$ groups according to their arrangement on the drone's camera screen and their position and posture. It is possible to obtain a flexible distribution. Therefore, in the proposed system, all subjects are classified into one of several classes according to the pixel position. Variational Bayes is an approximate solution for Bayesian estimation of probabilistic models. Variational Bayes is applied to determine the latent posterior distribution in the closed form when the probability problem is difficult. In this experiment, a subject is represented by a 2D mixed Gaussian distribution for evaluating the photographic composition. Therefore, variational Bayes can be used for clustering the subjects. The specific implementation method is as follows. The mixed Gaussian distribution with $K = 4$ contains a $D = 2$ dimensional Gaussian distribution as a mixed element. Additionally, the observation data, $\boldsymbol{a} = (x, y)$, are mixed with the probability of belonging to class k. In variational Bayes clustering, the initial value is $K = 4$. However, in most cases, the class distribution is smaller than K. The 2D mixed Gaussian distribution with multiple subjects optimized using variational Bayes is evaluated on the basis of the KL divergence with the reference composition distribution, in the same manner as that used in (4). The specific calculation procedure is performed with reference to [19].

C. VERIFICATION OF VIEWPOINT BY CHANGING REFERENCE COMPOSITION

We examine whether a change in the reference composition affects the optimal viewpoint position. Fig.6(a) shows the change in the reference composition of Fig.4(a), and Fig.6(b) shows the resultant change in the optimal viewpoint. It is confirmed that the optimal viewpoint position corresponding to Fig.6(a) is obtained from the photograph shown in Fig.6(b). These results demonstrate that the proposed method can be applied to different viewpoints by setting various reference compositions in advance.

V. OPTIMAL VIEWPOINT SELECTION

A. EVALUATION VALUE PREDICTION BY GAUSSIAN PROCESS

The flow of the optimal viewpoint search is explained. The viewpoint selection method is presented in Algorithm 1.

Algorithm 1 Optimal Viewpoint Selection Method

Require: Observation data $\boldsymbol{X}, \boldsymbol{y}$
Ensure: Optimal photo viewpoint position(x'_n, y'_n, z'_n)

```
while n ≦ 6 do
    Randomly observed and create a data set (6)
end while
while After 120s or until the observed values converge do
    Create a predicted distribution (15), (16) from a deta set (6)
    Find the maximum point of (17)
    PSO search by (19) using the maximum value of (17)
    if Difference between current position and next observation point d < 0.1 then
        Add observation data (6)
    end if
end while
```

As only one drone is used in the experiment, PSO cannot perform a wide-area search and tends to lead to a local solution. Therefore, a Gaussian process [19] is used to compensate for the small number of observations. The Gaussian process is used to interpolate the unsearched area. The mean and variance for each observation point are the predicted outputs. Therefore, the ambiguity of prediction can be expressed. As an example, the composition evaluation value at the third observation position is predicted when the number of observations is $N = 2$. Therefore, the input is the drone position, $\boldsymbol{x} = (x, y, z) \in \mathbb{R}^3$, and the output is the composition evaluation value, $y = u \in \mathbb{R}$. The orientation of the drone with respect to the subjects and the angle of the camera should be considered as inputs for the Gaussian process. However, in this experiment, the direction of the drone's viewpoint is controlled by the visual feedback for the subjects. Hence, the drone is always oriented toward the subjects during flight. Observations are obtained when the center of a subject is within a certain area on the camera screen. In addition, the camera angle of the drone is fixed for capturing a photo. Therefore, the prediction of the camera direction is not considered, and only the shooting position of the drone is considered. The linear model is expressed by the (5) with the

regression coefficient $\boldsymbol{b} \in \mathbb{R}^3$.

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b}$$
$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} \tag{5}$$

The regression coefficient is assumed to follows a Gaussian distribution, $p(\boldsymbol{b}) = \mathcal{N}(\boldsymbol{b} \mid \boldsymbol{0}, \boldsymbol{\sigma}_b^2)$. Then, the linear model (5) is extended to a nonlinear model using the kernel method. $\phi(\boldsymbol{X})$ is a design matrix that represents a function in a nonlinear space. A nonlinear mapping, $\boldsymbol{X} \rightarrow \phi(\boldsymbol{X})$, of the set, $\boldsymbol{X} = \{\boldsymbol{x}_n | 1 \leqq n \leqq N\}$, of the drone position, $\boldsymbol{x}_n = (x_n, y_n, z_n)^T \in \mathbb{R}^3$, after n times provides the following nonlinear regression model:

$$\boldsymbol{y} = \phi(\boldsymbol{X})\boldsymbol{b} \tag{6}$$

Let σ_{ij}^2 be the kernel function K, as given by (7).

$$\begin{aligned} \sigma_{ij}^2 &= \sigma_b^2 \phi(\boldsymbol{x}_i)\phi(\boldsymbol{x}_j)^T \\ &= K(\boldsymbol{x}_i, \boldsymbol{x}_j) \end{aligned} \tag{7}$$

The RBF kernel is used as a kernel function, as given by (8).

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left\{-\frac{\theta}{2}||\boldsymbol{x}_i - \boldsymbol{x}_j||^2\right\} \tag{8}$$

Furthermore, in practice, measurement error of $\boldsymbol{e}$ is considered in (6). Assuming that the measurement error is independent of the observation position and follows $p(e) = \mathcal{N}(e|0, \sigma_e^2)$, the nonlinear regression model (6) is given by (9).

$$\boldsymbol{t} = \boldsymbol{y} + \boldsymbol{e} \tag{9}$$

The mean vector, $\boldsymbol{\mu}_t$, of $\boldsymbol{t}$ and the variance–covariance matrix, $\boldsymbol{\Sigma}_t$, are as expressed follows:

$$\boldsymbol{\mu}_t = \boldsymbol{0} \tag{10}$$
$$\boldsymbol{\Sigma}_t = \begin{bmatrix} \sigma_b^2 \boldsymbol{x}_1 \boldsymbol{x}_1^T + \sigma_e^2 & \sigma_b^2 \boldsymbol{x}_1 \boldsymbol{x}_2 \\ \sigma_b^2 \boldsymbol{x}_2 \boldsymbol{x}_1^T & \sigma_b^2 \boldsymbol{x}_2 \boldsymbol{x}_2 + \sigma_e^2 \end{bmatrix} \tag{11}$$

The covariance is 0 because e is independent of the observation position.

Then, the composition evaluation value, u_3, at the third observation position is predicted. In other words, we determine the conditional distribution, $p(u_3|\boldsymbol{y})$, of u_3 given $\boldsymbol{y}$. Therefore, $p(u_3|\boldsymbol{y})$ can be transformed into (12).

$$p(u_3|\boldsymbol{y}) = \frac{p(u_3, \boldsymbol{y})}{p(\boldsymbol{y})} \tag{12}$$

The mean, $\boldsymbol{m}'$, and the variance–covariance matrix, $\boldsymbol{\Sigma}'$, of the joint distribution, $p(\boldsymbol{y}, u_3) = p(\boldsymbol{y}')$, are obtained using the matrix of kernel function (13).

$$\boldsymbol{k} = \begin{bmatrix} K(\boldsymbol{x}_1, \boldsymbol{x}_3) & K(\boldsymbol{x}_2, \boldsymbol{x}_3) \end{bmatrix}^T \tag{13}$$
$$\boldsymbol{m}' = \boldsymbol{0}, \quad \boldsymbol{\Sigma}' = \begin{bmatrix} \boldsymbol{\Sigma}_t & \boldsymbol{k} \\ \boldsymbol{k}^T & K(\boldsymbol{x}_3, \boldsymbol{x}_3) + \sigma_e^2 \end{bmatrix} \tag{14}$$

Therefore, as the joint distribution (14) and the prior distributions(10) and (11) are obtained, the conditional distribution can be derived based on [19]. The mean and variance of the conditional distribution are

$$m(\boldsymbol{X}_3) = \boldsymbol{k}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{t} \tag{15}$$
$$\sigma^2(\boldsymbol{X}_3) = K(\boldsymbol{x}_3, \boldsymbol{x}_3) + \boldsymbol{k}^T \boldsymbol{\Sigma}_t^{-1} \boldsymbol{k} \tag{16}$$

When the observation data are obtained by performing such a calculation, the photographic composition evaluation value of un unobserved point is expressed by the mean and the variance of the Gaussian distribution.

B. CREATION OF COMPOSITION EVALUATION MAP 3D SPACE

The composition evaluation function, $M(\boldsymbol{X}', m, \sigma^2)$ is created using the expected value, $m(\boldsymbol{X}')$, and the variance, $\sigma^2(\boldsymbol{X}')$, of the predicted distribution of the KL divergence at a certain unobserved point $\boldsymbol{X}'$ using the Gaussian process given in (15) and (16).

$$M(\boldsymbol{X}', m, \sigma^2) = \frac{1}{1 + e^{\alpha_M(m/\sigma^2 - \beta_M)}} \tag{17}$$

α_M and β_M are the coefficients of the sigmoid function; and $\alpha_M = 0.1$ and $\beta_M = 10$ in this study. In addition, E_{KL} is obtained by dividing the expected value, m, of the predicted distribution of the KL divergence by variance σ^2. This is used to reduce the uncertainty of evaluating observation points even if the predicted photographic composition evaluation value is small. As the composition evaluation function, $M(\boldsymbol{X}', m, \sigma^2)$, increases, the KL divergence and variance decrease. Therefore, a good composition can be expected.

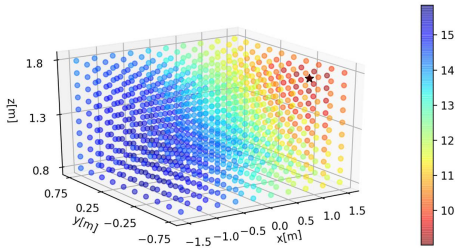


FIGURE 7. 3D photographic composition evaluation map.

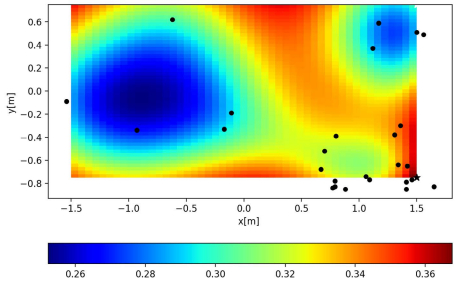


FIGURE 8. Composition map projected on XY plane.

Fig.7 shows the 3D composition map created by the Gaussian process in the actual experiment. The drone search

range, (x:-1.5 to 1.5m, y:-0.75 to 0.75m, z: 0.8 to 1.8m), is divided into $10\times10\times10$ points, and each point is used as the drone observation point. The composition evaluation value at an observation point is expressed using color, which changes from blue to red. $M(\boldsymbol{X}', m, \sigma^2)$ increases as the red color becomes stronger. It is difficult to evaluate the composition of the entire search space in the 3D composition map. Therefore, $M(\boldsymbol{X}', m, \sigma^2)$ is marginalized in the Z direction to create a 2D composition map projected onto the XY plane. The 2D composition map is shown in Fig.8. The colors denote the composition map obtained $M(\boldsymbol{X}', m, \sigma^2)$. Hence, the darker the red color, the more reliable the point where a photograph with a good composition can be obtained.

As the evaluation values of the Z axis are averaged, they are slightly different between the 2D and 3D composition maps. However, as the composition map is used only to simplify the composition evaluation, we use the 2D composition map to evaluate the composition evaluation.

C. SEARCH UPDATE BY PSO

The point in the drone search area is the input, $\boldsymbol{x} = (x, y, z) \in \mathbb{R}^3$, and the objective function is the composition evaluation value. As the basic calculation of PSO, the $n+1$ search point update formula (position: $\boldsymbol{x}(n+1)$, speed: $\boldsymbol{v}(n+1)$) is given by (18a) and (18b). The parameters are w, ρ_1 and ρ_2. $\boldsymbol{p}$ is the personal best, and $\boldsymbol{g}$ is the global best.

$$\boldsymbol{x}(n+1) = \boldsymbol{x}(n) + \boldsymbol{v}(n) \tag{18a}$$

$$\boldsymbol{v}(n+1) = w\boldsymbol{v}(n) + \rho_1(\boldsymbol{p} - \boldsymbol{x}(n)) + \rho_2(\boldsymbol{g} - \boldsymbol{x}(n)) \tag{18b}$$

Only one drone is used in this experiment. Based on this, the $n+1$ search point update formula with PSO is defined by

$$\boldsymbol{x}(n+1) = \boldsymbol{g} + \frac{C \times \rho}{1 + e^{\alpha \times \{m(\boldsymbol{g}) - \beta\}}}. \tag{19}$$

ρ is a random number and coefficient $C = 0.5$. The parameters of the sigmoid function are $\alpha = 10$ and $\beta = 0.5$. In addition, $\boldsymbol{g} \in \mathbb{R}^3$ is the observation position of the drone (global best) [18], which is predicted to obtain the photograph with the best composition. Subsequently, $m(\boldsymbol{g})$ is the composition evaluation value predicted by $\boldsymbol{g}$. We use a sigmoid function that inputs $m(\boldsymbol{g})$ to the search update. This makes it easier for the search to converge as it approaches the optimal viewpoint. In this study, it is possible to efficiently search for a viewpoint even with a single drone by interpolating and predicting the unobserved points using the Gaussian process for the wide-area search of PSO.

VI. EXPERIMENT ON OPTIMAL VIEWPOINT SELECTION METHOD

A. EXPERIMENTAL SETTINGS

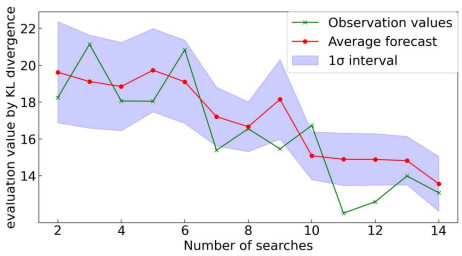
This section validates the Gaussian process and the 3D viewpoint selection method using PSO through actual machine experiments. The Parrot BEBOP2 drone is used in the experiments. It is assumed that images are obtained in an indoor environment, and there is no disturbance due to wind. The reason for conducting the experiment indoors is to minimize the error in photographic composition evaluation due to disturbance caused by wind. Therefore, the shooting environment is not limited. In addition, there are no obstacles in the search range of the drone. The subjects are 5 people, and variational Bayes is employed. The subjects are static, and there are no restrictions on the posture. The reference composition used for the optimal viewpoint is shown in Fig.4(a). The search range is set from the origin of the SLAM coordinate system (x: -1.5 to 1.5m, y: -0.75 to 0.75m, z: 0.8 to 1.8m). This range is selected from a sufficient range that can be ensured in our laboratory; the autonomous imaging system used in this study is not limited to this range. The condition for completing the process up to shooting is to obtain the log-likelihood from the actual composition evaluation value and the mean and variance of the Gaussian process. The value exceeds a certain standard, and the observed value is predicted. It is assumed that the data fit within the 1σ interval for 5 consecutive times or 120s have passed since the start of the search.

TABLE 1. Subject distribution information for each class($n = 15$).

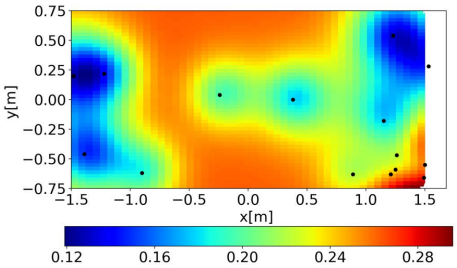
Class number	$k=1$	$k=2$
Center of distribution(pixel)	$(273, 317)$	$(461, 223)$
Size of distribution(pixel)	$(330, 351)$	$(259, 211)$
Responsibility of mixed elements	42%	53%

B. EXPERIMENTAL RESULTS

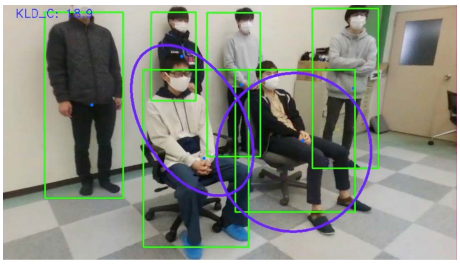
Fig.9(a) shows the observed KL divergence, number of searches in the KL divergence predicted by the Gaussian Process, and the change in the KL divergence. The 1σ interval becomes smaller because the prediction accuracy of the Gaussian process increases with the number of searches. The results for $n = 15$ at the optimal viewpoint show that the proposed method makes it possible to search for a viewpoint with a better composition compared to the start of the search. Furthermore, the observation results indicated by the green line converge to the viewpoint with a smaller KL divergence as the number of searches increases. It is considered that the KL divergence during the search is due to the observation error in the position control of the drone or the deviation due to the random search of PSO. It is also possible that the probabilistic prediction is wrong. Figs.9(c) and 9(d) show the photographic composition for $n = 4, 15$ in the Fig.9(a). The composition of the photograph obtained at $n = 15$ at the end of the search is similar to the reference composition, and the entire subject is well balanced. The distribution information clustered by variational Bayes for $n = 15$ is summarized in Table.1. The responsibility of mixed elements negligibly affects the composition evaluation and is omitted. The mean and variance vectors of the distribution clustered by variational Bayes are represented by a pixel coordinate system on the drone camera screen. The variance vector is expressed by the major and minor axis of the ellipse. These results confirm that the variational Bayes clustering is accurate. The black



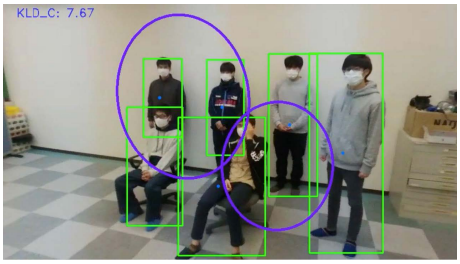
(a) KL divergence with the viewpoint in search



(b) 2D composition map(experiment)



(c) Photographic composition with 4 searches



(d) Photographic composition with 15 searches

FIGURE 9. Optimal viewpoint selection experiment using variational Bayes.

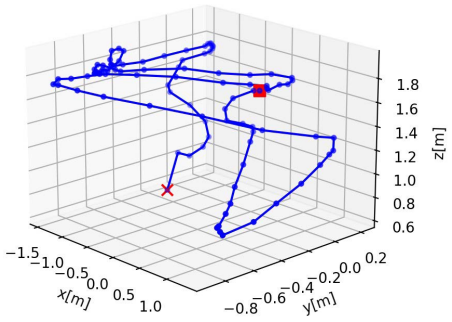


FIGURE 10. PSO search trajectory of the drone.

dots drawn on the 2D composition map in Fig.9(b) represent the observation points in the drone search. The white star mark represents the optimal viewpoint position for $n = 15$. The dispersion of these observation points shows that the drone can search over a wide area. The drone search route is shown in Fig.10. The red cross and red square indicates the starting point and optimal viewpoint position, respectively. These observation points show that the search is does not fall into a local solution and that the composition evaluation value is predicted over the entire search range by combining PSO and the Gaussian process.

VII. EXPERIMENTAL ANALYSIS

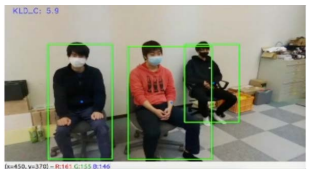
A. ANALYSIS OF OPTIMAL VIEWPOINT POSITION

The experimental analysis of the proposed Gaussian process and 3D viewpoint selection method with PSO is performed. The experiment is performed multiple times for the same subject positions and the initial position ($(x, y, z) = (1.5\text{m}, -0.75\text{m}, 0\text{m})$) of the drone, and the variation in the optimal viewpoint is verified. The experiment is performed 10 times with 3 static subjects. Fig.11(c) shows a diagram of the optimal viewpoint positions. The optimal viewpoint position is considerably scattered in three places. However, the most common optimal viewpoint is on the left side with respect to the origin of the global coordinate system. As the search is performed under a time constraint of 120s, the optimal viewpoint will vary more as the arrangement of subjects become more complex and the number of subjects increases. The optimal viewpoint can be selected more efficiently and accurately by changing the search time according to the arrangement and number of subjects.

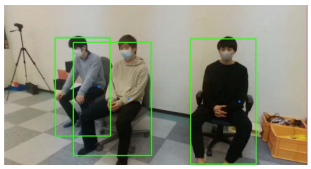
It is verified whether the proposed method is a better than only PSO. Fig.11(d) shows the optimal viewpoints obtained using only PSO under the same experimental settings as Fig.11(c). The initial position of the drone is set as the position of the red square in Fig.11(d). In the case of a search using only PSO, since only one drone is used, the search often failed to converge within 120s. Convergence to the optimum viewpoint position depends on the initial position in many cases. It is confirmed that the proposed search method with Gaussian process can find the optimum viewpoint without depending on the initial position. Next, Table.3 shows the average and variance of the composition evaluation value, D^c. D^c is lower for the proposed method, and a better viewpoint is obtained. As the value of D^c at the optimal viewpoint tested 10 times is also small, there is a slight bias in the search solution even after multiple experiments. The proposed method supplements PSO and provides better viewpoint selection.

TABLE 2. Comparison of D^c and convergence time.

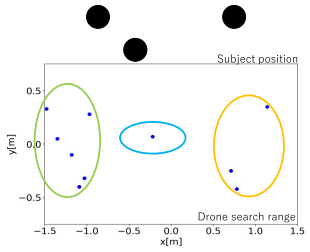
	PSO and Gaussian Process	Full range search experiment
Time	95.7s(Min: 64s)	120s
Optimal viewpoint (x, y, z)	(-1.17m, -0.42m, 1.36m)	(-1.5m, -0.25m, 1.69m)
D^c at optimal viewpoint	5.63	6.53
D^c at the origin	8.23	9.31
D^c at $(x : 1.0\text{m}, y : 0.0\text{m}, z : 1.0\text{m})$	12.02	8.56



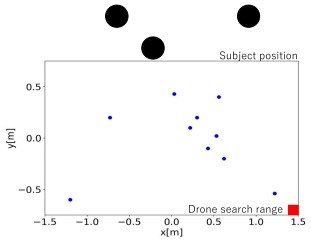
(a) Photograph from the left viewpoint



(b) Photograph from the right viewpoint



(c) Summary of optimal viewpoint position

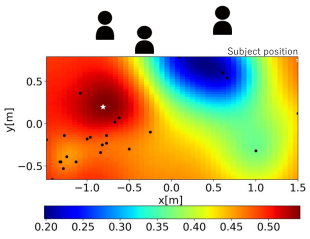


(d) Search experiment using only PSO

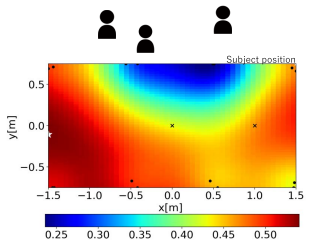
FIGURE 11. Optimal viewpoint position and photographs in each experiment.

TABLE 3. Mean and variance of D^c in each experiment.

	PSO and Gaussian process	PSO
Average of D^c	9.95	12.6
Variance of D^c	7.61	13.22



(a) PSO and Gaussian process



(b) Full range search experiment

FIGURE 12. Comparison of composition maps.

B. COMPOSITION MAP ANALYSIS

We verify whether the proposed optimal viewpoint selection method can derive the optimum viewpoint in a time-efficient manner. The proposed method is compared with the composition evaluation map and search time of the experiment in which the entire range is observed over a wide area (the search range is divided into $4 \times 2 \times 2$). There are 3 static subjects in the experiment. The composition maps obtained using the proposed method and full-range search experiment are shown in Figs.12(a) and 12(b), respectively.

Both composition maps are similar. This shows that a wide-area search using PSO and interpolation using the Gaussian process are possible. A summary of the search time and the optimal viewpoint position is presented in Table.2. The search time in the full-range search experiment is 120s. In the case of proposed method, the average and shortest search time are 95.7s and 64s, respectively. This demonstrates that the proposed method is a time-efficient approach. Table.2 shows that the entire range is sufficiently predicted. This confirms that the proposed method can perform a search without hindering the prediction of the search range, even though it is more time efficient than the full-range search experiment.

VIII. CONCLUSIONS

An optimal viewpoint selection method is proposed for an indoor drone photographer. The method is applied to a commercially available drone without modification. The drone can capture a photograph with a well-balanced composition according to a user's preference using ORB-SLAM for visual feedback control. A group of subjects is visually classified into a set of the clusters represented by a mixed Gaussian distribution. The captured scene changes according to the movement of the drone, but the clustering is dynamically adjusted using variational Bayes. The KL divergence between the subjects' distribution and the user-defined reference composition, which is represented as a Gaussian mixture model, is defined as the composition evaluation value. PSO and a Gaussian process are integrated in the drone's search. This enables the drone to predict the composition evaluation value in an unobserved area. Hence, a time-efficient search is possible within a 120s in our experimental setting. Moreover, the results are similar to those obtained in the entire search of the space. However, the search should be energy efficient in addition to being time efficient. Additionally, the proposed system is limited to the static subjects, and collision avoidance must be considered for practical use. As part of future work, we are investigating a viewpoint selection method that can track mobile subjects.

REFERENCES

[1] M. Germen, "Alternative cityscape visualisation: Drone shooting as a new dimension in urban photography," in *Proc. Electron. Workshops Comput.*, Jul. 2016, pp. 1–8.
[2] D. Gozen and S. Ozer, "Visual object tracking in drone images with deep reinforcement learning," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 10082–10089.
[3] L. Liu, R. Chen, L. Wolf, and D. Cohen-Or, "Optimizing photo composition," *Comput. Graph. Forum*, vol. 29, no. 2, pp. 469–478, 2010.
[4] K. Lan and K. Sekiyama, "Autonomous robot photographer with KL divergence optimization of image composition and human facial direction," *Robot. Auto. Syst.*, vol. 111, pp. 295–300, Jan. 2019.
[5] M. Zabarauskas and S. Cameron, "Luke: An autonomous robot photographer," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 1809–1815.
[6] R. C. Luo, W. U. Chan, and P.-J. Lai, "Intelligent robot photographer: Help people taking pictures using their own camera," in *Proc. IEEE/SICE Int. Symp. Syst. Integr.*, Dec. 2014, pp. 322–327.
[7] R. Newbury, A. Cosgun, M. Koseoglu, and T. Drummond, "Learning to take good pictures of people with a robot photographer," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 11268–11275.
[8] Z. Byers, M. Dixon, K. Goodier, C. M. Grimm, and W. D. Smart, "An autonomous robot photographer," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2003, pp. 2636–2641.
[9] J. Kenndy and R. Eberfart, "Particle swarm optimization," in *Proc. ICNN Int. Conf. Neural Netw.*, Nov. 2015, pp. 1942–1948.
[10] D. Wang, D. Tan, and L. Liu, "Particle swarm optimization algorithm: An overview," *Soft Comput.*, vol. 22, no. 2, pp. 387–408, 2017.
[11] K. Kameyama, "Particle swarm optimization," *IEICE Trans. Inf. Syst.*, vol. E92-D, no. 7, pp. 1354–1361, 2009.
[12] W. Kongkaew and J. Pichitlamken, "A Gaussian process regression model for the traveling salesman problem," *J. Comput. Sci.*, vol. 8, no. 10, pp. 1749–1758, Oct. 2012.
[13] J. Tian, Y. Tan, J. Zeng, C. Sun, and Y. Jin, "Multiobjective infill criterion driven Gaussian process-assisted particle swarm optimization of high-dimensional expensive problems," *IEEE Trans. Evol. Comput.*, vol. 23, no. 3, pp. 459–472, Jun. 2019.
[14] K. Nonami, "Current status and challenges of drone technology and the forefront of business," *Translated From Jpn. Inf. Manage.*, vol. 59, no. 11, pp. 755–763, 2017.
[15] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: A versatile and accurate monocular SLAM system," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015.
[16] J. R. Hershey and P. A. Olsen, "Approximating the Kullback Leibler divergence between Gaussian mixture models," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2007, pp. IV-317–IV-320.
[17] J.-L. Durrieu, J.-P. Thiran, and F. Kelly, "Lower and upper bounds for approximation of the Kullback–Leibler divergence between Gaussian mixture models," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 4833–4836.
[18] T. Saito, "Particle swarm optimization and nonlinear system," *IEICE Fundam. Rev.*, vol. 5, no. 2, pp. 155–161, 2011.
[19] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.