Paper:

The Proposal of Region Proposal Method for Outdoor- Camera's Image by Fuzzy Inference System

Junsuke Yokoseki **Yukinobu Hoshino**

Kochi University of Technology
Kami, Kochi 782-8502, Japan
E-mail: {245065k@gs., hoshino.yukinobu@}kochi-tech.ac.jp

Abstract. In recent years, wild monkeys have been damaging crops in field area in Japan. Since monkeys live in groups, it is necessary to capture all the monkeys in a group to prevent damage. Therefore, we need a system that can detect when a group of monkeys enter a trap by using image recognition. The goal of this research is to detect a group of monkeys by image recognition. We expect to use images from a fixed-point camera installed in an outdoor cage which capture monkeys. The problem with these images are that the color distribution changes dynamically over time by sunlight. In this paper, we propose a method for extracting the regions in an image that contain monkeys. This method converts the image to L*a*b* color space and extracts the regions by fuzzy inference. The result shows that the proposed method is an effective region proposal method. In general, fuzzy reasoning requires less computation and reduces the entropy. Since the capture system will be installed in field areas where power supply is difficult. We consider that the proposed method can assist artificial intelligence and edge computing.

Keywords: Wildlife Capture System, Fuzzy Reasoning, Image Recognition, Edge Computing,

1. INTRODUCTION

In recent years, damage caused by wildlife has become a problem in Japan. In many of the countryside in Japan, agriculture is the main industry. Damage caused by wildlife not only reduces agricultural production, but also makes it impossible to grow crops that are susceptible to attack by wildlife. In other words, damage caused by wildlife is a very serious problem for farmers. In addition, the damage caused by wildlife is not only to crops, but also to human about directly attack. Because the monkey is the most intelligent in wildlife species, it is difficult to take measures against damage. As a result of population decline and aging in countryside, monkeys are not afraid of humans and invade human's houses. As a basic policy of countermeasures against damage by animals, methods of extermination and control are considered. In the case of control, the whole community drive away the monkeys or electric fences is set on field to protect crop by them, etc. However, the affected area may be spread, we consider that capture and extermination is more desirable. There are two methods of extermination: direct extermination and trapping. Direct extermination is difficult because a lot of hunters are currently aging. Therefore, we consider that trapping is the most appropriate method. Many efforts are being made to reduce the burden of trapping work by technologies like IoT (Internet of Things) and image recognition. One of the systems that can reduce the burden of monitoring the trapping status is the remote monitoring system. But in the field area where trap is set, the necessary communication environment is often not available for remote control. For deer fences, Kato et al[1]. have developed a vibration detection system for monitoring deer fences. The system acquires sensor information from the field by long-distance wireless communication. In other studies, wildlife capture systems have become more high spec with automatic detection, notification, and automatic capture functions which based on camera and sensor information. The more advanced the functions, the larger the capture system. Oishi et al[2]. focused on the problem of the weight of the capture system when transporting the equipment to the field area in the mountains. They developed and evaluated a small system using IoT technology.

The characteristics of monkeys and the points to note when capturing them are as follows. Wild monkeys (Japanese monkeys) usually live in groups of 40 to 50. They are diurnal animals and sleep at night. In addition, they are highly intelligent. So, if we capture a monkey by itself, it is likely that other monkeys in the group will learn the trap and will not come near the trap. Therefore, it is necessary to capture a whole group of monkeys by using a large cage. In addition, it is necessary to capture the parent monkeys in order to prevent damage by the group. However, the parent monkeys are very careful. When the monkeys recognize a new feeding site, the less careful monkeys will visit the site. But the parent monkeys will not visit. Thus, the number of monkeys in the cage is an index to determine the timing of capture. Inada et al[3]. have been studying a method to measure the number of wildlife. In the report, they use an infrared camera installed in the cage to measure the number of deer. They succeeded that to detect the number of deer with high accuracy by taking the background difference of depth information in images. However, the experiment was conducted at night. It can only use with wild deer and wild boar, which are active even at night. A different method is required for monkeys, which are mainly active during the day.

Studies on image recognition of wild animals have been reported[4][5].The purpose of these studies is to investigate the ecology of wild animals, and to estimate the species of animals using CNN (Convolutional Neural Network). Methods using moving images have been proposed[6][7]. These methods combine CNN and moving object detection methods. The purpose of these researchers are to investigate the ecology of wild animals. Therefore, computer specs is not taken into account.

In the field, it is difficult to supply a large amount of electric power to the capture system. It is necessary to control the capture system and recognize the image with a microcomputer. Because the microcomputer use low electric power. Kamesaka et al[8]. Created and verified a CNN that can embedded on microcomputer in a capture system. Image processing needs to be accelerated in order to know the appropriate capture timing. Research aimed at high-speed processing and low power consumption was conducted by Kamesaka et al[9]. This is achieved by mounting the CNN circuit on the FPGA board. And CNN architecture which can embedded on the board was shown in below. 1st input layer require a image which is 28 pixels × 28 pixels. 2nd hidden layer has 60 neuros, 3rd hidden layer has 20 neuros and output layer has 2 neuros. The result shows that small size CNN need small size input image which is pre-proposed small area from the whole image. There are many kind of CNNs that automatically extract and discriminate regions. But these architecture have many parameter. and use many memory of microcomputer. If the CNN model was embedded on server that set in remote location, CNN is not possible to acquire images in real time by poor communication environment. In this paper, region proposal method is that combines the background subtraction method and the distance of the color distribution. The reasons why we use these method are poor communication environment, poor memory storage on microcomputer and no images of monkey in the trap.

After the background subtraction as pre-processing, the region focused by pre-processing is calculated the difference of color distribution from the template image. And the region is narrowed the focus area. The color distribution of template images of monkey changes dynamically by the influence of sunlight. The first reason is that the brightness changes over time. The second reason is that it is affected by shadows. The fuzzy inference system was used as a way to construct a measure as the membership grade of being a monkey depend on time and position in the image. In this paper, we propose a method to detect the region which include a monkey from whole image.

1.1. Data Source

In this experiment, we use the moving image at the Awaji Island Monkey Center[11] which is published on website. The images are pictures from the fixed camera. We captured the image every 30 seconds. A sample of image is provided in Fig. 1.

Fig. 1. example image : Awaji Island Monkey Centor

Fig. 1 is a good sample because this image shows some difficult problem to solve to detect monkeys. First, the color of the monkey is different for each region in the image. (e.g., the monkey capturing in the front of picture is bright. But the monkey in the behind of picture is dark). Second, since the shadow of the tree is moving slowly in the image, the background subtraction method only can't remove the shadow. In addition, if the monkey is captured in the same region on the image, the color of the monkey is different depending on the time. We recognise these problems had to solve when apply the system on the real field.

We approach these problem with the proposal method. The difference of the color is converted to similarity and measured by antecedent part of fuzzy reasoning. The difference of the region is converted to distance and measured by consequent part of fuzzy reasoning. The effect to the color by the time is controlled by proposing with template image which captured near the time on previous day.

2. METHODOLOGY

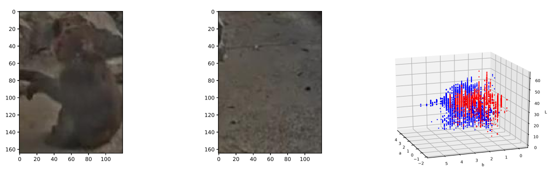
2.1. Background Subtraction

The background subtraction method is used to detect moving objects. In this study, we use this method because we want detect the object as an animal which suddenly appear in front of image. A sequential image is expressed the function of V(x,y,t) which is depends on time. where t is the time dimension, x and y are variables of the pixel location. In this case, simple Median-based Background Subtraction was used to be a method. The method compute the median for each pixel over time. And the value of median on pixel become the value of pixel of Background Subtraction.

2.2. L*a*b* Space

The L*a*b* Space is designed to approximate human vision. The three coordinates of L*a*b* represent the lightness of the color, its position between red and green, and its position between yellow and blue. The differences between any two colors in L*a*b* can be calculated as the Euclidean distance between them. In the field of image processing, a method of mapping to the L*a*b* color space and analyzing in the space for removing shadows has been proposed [10].

e.g. monkey and Ground are mapped in L*a*b* space which are provided in Fig. 2c. Where the red plots are



(a) monkey (b) Ground (c) 3d plot

Fig. 2. color distribution in L*a*b* space

color distribution of ground image and blue plots are monkey. The result shows that the difference of color distribution between object and background may be used to recognize the monkey.

2.3. Kullback–Leibler Divergence

The Kullback–Leibler divergence, (D_{KL}) is a kind of measure of difference between the two probability distributions. If the two probability distributions match, the entropy is 0. The more different the mean or variance values, the greater the entropy. When probability distributions $p(x)$ and $q(x)$ are defined as follows:

$$p(x) = N(m_p, \sigma_p^2) = \frac{1}{\sqrt{2\pi\sigma_p^2}} \exp\left(-\frac{(x-m_p)^2}{2\sigma_p^2}\right) \quad (1)$$

$$q(x) = N(m_q, \sigma_q^2) = \frac{1}{\sqrt{2\pi\sigma_q^2}} \exp\left(-\frac{(x-m_q)^2}{2\sigma_q^2}\right) \quad (2)$$

D_{KL} is defined as follows:

$$D_{\mathrm{KL}}(P\|Q) = \log\frac{\sigma_q}{\sigma_p} + \frac{\sigma_p^2 + (m_p - m_q)^2}{2\sigma_q^2} - \frac{1}{2} \quad . \quad (3)$$

2.4. Fuzzy Leasoning System

Fuzzy reasoning is described as follows:

$$\begin{array}{llll}
R_A: \ If\ X_L^{PA}\ is\ X_L^Q, & X_a^{PA}\ is\ X_a^Q, & X_b^{PA}\ is\ X_b^Q, & thenY_A\ is\ S_A \\
R_B: \ If\ X_L^{PB}\ is\ X_L^Q, & X_a^{PB}\ is\ X_a^Q, & X_b^{PB}\ is\ X_b^Q, & thenY_B\ is\ S_B \\
R_C: \ If\ X_L^{PC}\ is\ X_L^Q, & X_a^{PC}\ is\ X_a^Q, & X_b^{PC}\ is\ X_b^Q, & thenY_C\ is\ S_C \\
 & \vdots & & \\
Fact: \ X_L^{Q*}, & X_a^{Q*}, & X_b^{Q*} & \\ \hline
Result: & & & Y^*
\end{array}$$

where $X_L^{Pi},\ X_a^{Pi},\ X_b^{Pi} (i \in \{A,B,C,...\})$ are the labels of fuzzy sets of antecedent part. For example, the membership function of X_L^{Pi} is given by $p_L^A = N(m_{pL}, \sigma_{pL}^2)$. when m_{pL} is mean and σ_{pL}^2 is variance. Those use L channel for each pixel of template image. $X_L^Q,\ X_a^Q,\ X_b^Q$ are the labels of fuzzy sets of input. The membership function of X_L^Q is given by $N(m_{qL}, \sigma_{qL}^2)$ in the same way. S_i is the label of fuzzy set of consequent part . The membership function is singleton g. Let g be the following function:

$$g(S) = \begin{cases} 1, & S = S_i \\ 0, & S \neq S_i \end{cases} \quad . . . . \quad (4)$$

In this study, we use the simplified fuzzy reasoning system. The system is provided in Fig. 3.

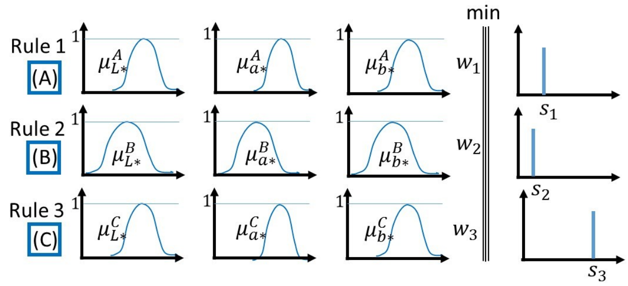


Fig. 3. simplified fuzzy reasoning

2.4.1. antecedent part

The membership function μ_j^i which calculate similarity of color distribution of j channel between P and Q is described as follows:

$$\mu_j^i = h_1(D_{KL}^j(P_i\|Q)) \quad . . . . \quad (5)$$

where $i \in \{A,B,C,...,\}$ and $j \in \{L,a,b\}$. h_1 is normalize function and described as follows:

$$h_1(x) = -\frac{2}{1+e^{-x}} + 2 \quad . . . . \quad (6)$$

The graph of h_1 function is provided in Fig. 4. The function maps the value of D_{KL} to $[0,1]$ space. If the value of D_{KL} is large due to low similarity, then the h_1 function outputs small real number which is nearly zero.

The degree of match of the antecedent part of fuzzy reasoning is w_i and is defined as follows:

$$w_i = \min_{j\in\{L,a,b\}} \mu_j^i \quad . . . . \quad (7)$$

w_i is crisp output and is mapped on $[0,1]$ space.

2.4.2. consequent part

The membership function $g(S)$ is singleton. The variable s express the similarity of coordinate between the region P and Q on the same image. In this time, we selected Manhattan distance (D_M). D_M is defined as follows:

$$D_M^i = |x_{pi} - x_q| + |y_{pi} - y_q| \quad . . . . \quad (8)$$

Where the coordinate of pixel of upper left on the template region is shown as $\boldsymbol{p}_i = [x_{pi}, y_{pi}]^T \ (i \in A,B,C,...)$. And the coordinate of pixel of upper left on the input region is shown as $\boldsymbol{q} = [x_q, y_q]^T$.

The similarity S_i of coordinate between the region P and Q on the same image can be expressed as

$$S_i = h_2(D_M^i) \quad \ldots\ldots\ldots\ldots\ldots \quad (9)$$

where h_2 is normalize function and described as follows:

$$h_2(x) = -\frac{1}{1+e^{-0.001\times x}} + 1.5 \quad \ldots\ldots \quad (10)$$

The graph of h_2 function is provided in Fig. 5. The function maps the value of D_M to $[0.5, 1]$ space. If the value of D_M is large due to long distance, then the h_2 function outputs real number which is nearly 0.5. Where 1.5 and -0.001 are constant value which were gotten by experimental result.

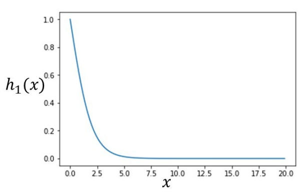


Fig. 4. The h1 function

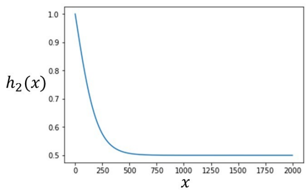


Fig. 5. The h2 function

2.4.3. Defuzzification

The overall output Z is taken as the weighted average of each rule's output w_i. The method can be expressed as

$$Z = \frac{\sum_{i=1}^{n} w_i S_i}{\sum_{i=1}^{n} w_i} \quad \ldots\ldots\ldots\ldots\ldots \quad (11)$$

2.5. Creating the Heat Map

The process of proposed method to the target image $V(x,y,t)$ is followed.

1. Let $V(x,y,t_i)$ is a target image.
2. Choose an image $V'(x,y,t)$ among images on the previous day near time t_i.
3. Read patch images $(A', B', C', \ldots)$ which have a object region from $V'(x,y,t)$
4. Read the upper left coordinate as $(\boldsymbol{p}_A, \boldsymbol{p}_B, \boldsymbol{p}_C, \ldots)$ and center region of patch images as template images $(A,B,C,\ldots)$ from $(A', B', C', \ldots)$.
5. Calculate the difference between $V(x,y,t_i)$ and background image.
6. Bound the minimum area which include the difference region binarized.
7. Calculate the color similarity between template image P and input area Q which included in bounded area.
8. Calculate the region similarity between template image P and input area Q which included in bounded area.
9. Add Z which is degree of similarity between P and Q to the same area as Q in heat map.
10. Take raster scan in bounded area and repeat processes from the 7th to 9th.

The processes from 1st to 6th are detection process of template region and are provided in Fig. 6. Where the box(1) on previous day is shown as $V'(x,y,t)$. And the box(1) on the day is shown as $V(x,y,t_i)$. The boxes(2) on previous day are patch images $(A', B', C', \ldots)$. The boxes(3) which is inside boxes(2) are template images $(A,B,C,\ldots)$. The box(4) which is inside box(1) is the area of 6th process and expressed by dashed line.

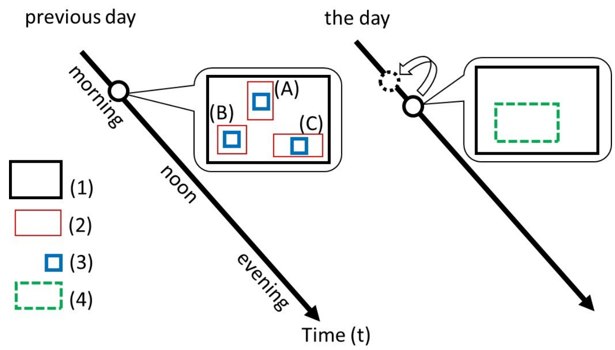


Fig. 6. Detection process of template region

The processes from 7th to 8th are calculation process of distance between input area Q and template area P. The processes are provided in Fig. 7. Where the area filled with diagonal line is input area Q and slid by raster scan pixel by pixel. The template area P is shown at upper right of input area Q. The template area P is the same as boxes(3) in Fig. 6. The result of 7th process is used to calculate antecedent part of fuzzy reasoning. The result of 8th process is used to calculate consequent part of fuzzy reasoning. The process of 8th is provided in Fig. 8. Where the coordinate of pixel of upper left on the template area $i (i \in A,B,C,\ldots)$is shown as $\boldsymbol{p}_i = [x_{pi}, y_{pi}]^T$. And the coordinate of pixel of upper left on the input area is shown as $\boldsymbol{q} = [x_q, y_q]^T$. These variables are used in Eq. 8. The process 9th is calculation process of similarity by fuzzy reasoning using result of 7th and 8th. The value of Z is result of Eq. 11

2.6. Region Proposal

Before propose the region, The heat-map and absolute difference image are binarized. Finally, calculate the min operation. These binarized-map is provided in Fig. 9. The bottom image is raw data. The second is heat-map made by fuzzy reasoning. The third is absolute difference image made by the process of background subtraction. The

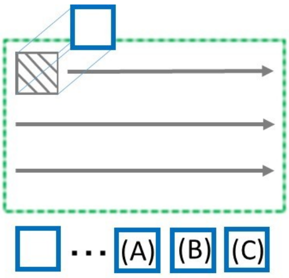


Fig. 7. raster scan

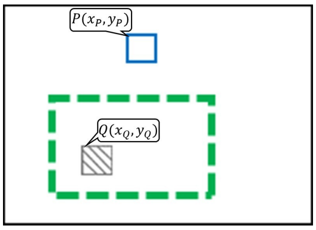


Fig. 8. the distance between P and Q

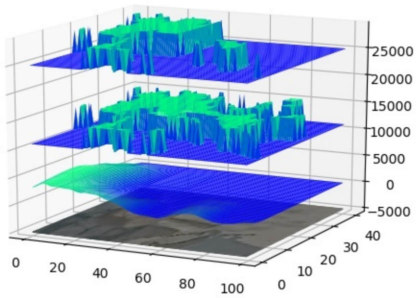


Fig. 9. 3d heat-map

top image is the result of min operation between absolute difference image and heat-map from fuzzy.

3. RESULTS

The images the part of the result which applied our method are provided in Fig. 10, 11 and 12. These base images which are proposed by background subtraction are not good region. This is because the places other than the monkey's body is detected by shadows. The region where propose by our method is shown in min heat-map of Fig. (d). The result shows that our proposal method couldn't clearly detect the region of target object.

4. DISCUSSION

in this experiment, D_{KL} was used to measure the similarity between input image and template images. In the region where the color distribution is uniform, such as sandboxes, the variance of the color distribution is 0. It was impossible to measure the D_{KL} of some region in image by this problem. It shows that it is necessary to change from DKL to the method of treating the mean and variance of the color distribution as vector components and measuring the distance between the vectors.

5. CONCLUSION

In this paper, we investigate a method for detecting individual monkeys in a cage of a group of monkeys. First, we performed individual detection using the background subtraction method and FIS, and confirmed that the detection area specified by the background subtraction method is affected by shadows. Next, we paid attention to the influence of shadows, and specified the detection area using the background subtraction method in combination with the FIS for shadow removal. As a result, our proposal method couldn't clearly detect the region of target object and there were some false responses such as deletion of a part of the monkey's body. In the future, we plan to use the shadow removal FIS for monkey images of various time periods, sizes, and image positions to identify the monkey images that are prone to false reactions. In addition, when monkeys overlap each other on the screen, it is difficult to detect multiple monkeys by the background subtraction method, so we plan to use other information such as color distribution to detect individual monkeys or groups of monkeys.

References:

[1] Takuya Kato, Hiroyuki Yokota, Yuichiro Hei, Izuru Nogaito, and Eiji Utsunomiya, "A Prototype and Evaluation of Vibration Detection System for Surveillance of Deer Fence", IEICE Technical Report. Vol. 120, No. 73, pp. 1-6, 2020.

[2] Kai Oishi and Shinichiro Mori, "Beast damage countermeasures using IoT", The Special Interest Group Technical Reports of IPSJ, Vol. 2021-CDS-31, No. 1, pp 1-8, 2021.

[3] Tatsuki Inada, Akie Yagura, Yousuke Yamamoto, Saori Hamaguchi, Kazufumi Nakai, Nobuo Ezaki, Naoto Yamabata and Osamu Takahashi, "Measurement of Number of Vermin around a Cage", ITE technical report, Vol. 41, No. 5, pp. 203-208, 2017.

[4] Masaki Ando, Shunsuke Nakatsuka, Hiroaki Aizawa, Satsuki Nakamori, Takashi Ikeda, Junji Moribe, Kazunori Terada, and Kunihito Kato, "Recognition of wildlife using deep learning in images taken by camera traps", The Japanese Journal of the Mammalian science, Vol. 59, No. 1, pp. 49-61, 2019.

[5] Yusuke Uehara, Naoya Otani, Keisuke Nishimoto, Kunimasa Kawasaki, Yasuhiro Matsui, Shigemi Nagata, Shumpei Kitamura, and Yutaka Yada, "A wildlife recognition system based on deep learning", The Journal of the Chubu forestry research, Vol. 68, No. 1, pp.25-30, 2020.

[6] Ryosuke Kabata, Hidetaka Hayashi, Kazunori Terada, Masataka Ueno, and Kazunori Yamada, "Visual Identification and Tracking of Individual Japanese Macaque", Proceedings of Fuzzy system Symposium. Vol. 33, No. 1, pp. 343-348, 2017.

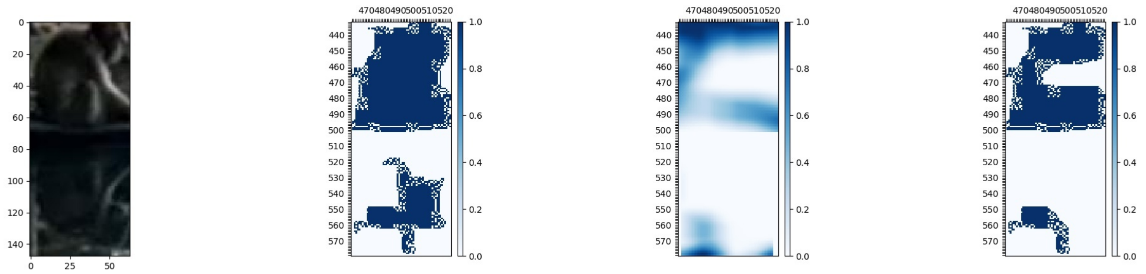
[7] Ayana Nakajimaa, Hiroyuki Okua, Kazuhiro Motegia and Yoichi Shiraishia, "Wild Animal Recognition Method for Videos Using a Combination of Deep Learning and Motion Detection", The Japanese Journal of the Institute of Industrial Applications Engineers, Vol. 9, No. 1, pp. 38-45, 2021.

[8] Ryoki Kamesaka, and Yukinobu Hoshino. "Development of a Prevention System for Beast Damage of Agricultural Products Using Deep Learning", Proceedings of Joint 10th International Conference on Soft Computing and Intelligent Systems and 19th International Symposium on Advanced Intelligent Systems. pp. 747-752, 2018.

[9] Ryoki Kamesaka, Yukinobu Hoshino, "Design of the Convolution Layer Using HDL and Evaluation of Latency Using a Camera Signal", Journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol. 33, No. 1, pp. 515-519, 2021.

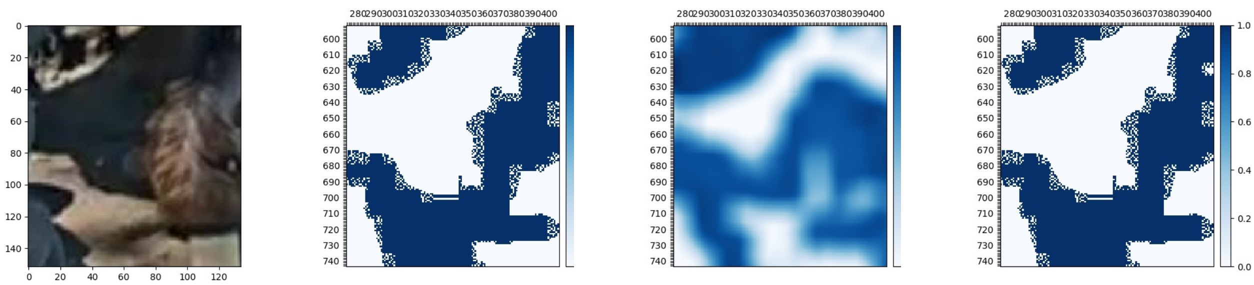
[10] Saritha Murali, V. K. Govindan, "Shadow Detection and Removal from a Single Image Using LAB Color Space", Journal of Cybernetics and Information Technologies, Vol. 13, No. 1, pp. 95 103, 2013.

[11] Awaji Island Monkey Centor, HP: https://monkey-center.jp/english.php



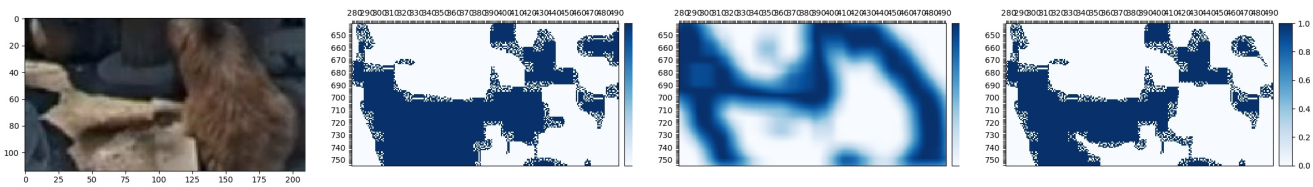
(a) base img (b) background subtraction (c) fuzzy heatmap (d) min heatmap

Fig. 10. 119th image



(a) base img (b) background subtraction (c) fuzzy heatmap (d) min heatmap

Fig. 11. 1300th image



(a) base img (b) background subtraction (c) fuzzy heatmap (d) min heatmap

Fig. 12. 1306th image