DE GRUYTER
OPEN

JAISCR, 2016, Vol. 6, No. 4, pp.255 – 265
10.1515/jaiscr-2016-0019

AUTONOMOUS VIEWPOINT SELECTION OF ROBOT BASED ON AESTHETIC EVALUATION OF A SCENE

Kai Lan, Kosuke Sekiyama

Department of Micro-Nano System Engineering, Nagoya University,
Furo-cho, Chikusa-ku, 464-8601 Nagoya, Japan

Abstract

In this paper, we propose an optimal viewpoint selection system for monitoring robots to search for the optimal viewpoint of a scene with the highest aesthetic property. Using the information of the targets, we propose a novel method for predicting human aesthetic sense for a scene. We construct evaluation functions based on certain known composition rules using three factors, namely, target size, visual balance, and composition fitting value. Then a score, which is a reflection of human evaluation, will be obtained using these functions. The optimal viewpoint will be selected from a number of candidates around the target group, by evaluating the aesthetic properties of scenes for each candidate viewpoint. Finally, once the optimal viewpoint is confirmed, path planning and path following controls are implemented for the robots during the moving process.

Keywords: optimal viewpoint selection system, aesthetic evaluation, viewpoint searching

1 Introduction

In recent years, it has been seen that a robot photographer can be used in various situations such as live sports broadcast, film shooting, and aerial photographing. Through its swift mobility, a robot photographer can capture the targets from a viewpoint that cannot be easily reached by humans [1]. In particular, due to its portability and low cost, a robot photographer can lead to immense growth in the field of photography facilitating both professionals and amateurs worldwide.

The composition of photos is an arrangement of visual components within a picture frame [2] and it is a basic technique to capture better pictures. In the field of photography, there are several principles that can be used to attain natural and aesthetic compositions. *Rule of thirds*, *Diagonal composition* and *Triangle composition* are the most commonly used compositions [3].

However, conventional robot photographers have not dealt with the issue of computational aesthetic evaluation for compositions and also that of the autonomous searching for an appropriate viewpoint. The selection of these compositions and viewpoints has been done by a remote operator. In this study, we will develop a viewpoint selection system that can realize autonomous viewpoint selection for robot photographers based on the aesthetic evaluation of a scene.

Previous studies have performed aesthetic evaluation of compositions by quantifying the composition components as evaluation factors. This is called *Computational Aesthetics* that is related to the computational evaluation of aesthetics in the field of human creative expression [4]. Mathematical formulas that represent aesthetic principles have been proven to correlate with human evaluation [5], [6]. For example, our previous work [7] succeeded in sorting aesthetically good and bad scenes for a teleoperation task by a support vector machine with the target balance and size. Although this system is feasible for evaluating the composition, it cannot be used for aesthetic evaluation. Liu et al. [8] and Zhang et al. [9] proposed a crop-and-retarget operator by repositioning foreground targets in a scene using an after-processing method. The aesthetic features are transformed into mathematical formulas in conjunction with specialized algorithms. They were realized to digitalize the subjective evaluation of humans in the research. Furthermore, Byers et al. [10], [11] tried to combine composition evaluation with viewpoint selection. In case of events or a ceremony where people are moving and talking to each other, they proposed a complete, end-to-end system for capturing an aesthetic photograph based on the *Rule of thirds*. However, due to the lack of a well-planed viewpoint search process, the success ratio of their system in obtaining satisfactory compositions is only 29%.

In this paper, we propose a novel system for optimal viewpoint selection based on aesthetic composition evaluation. By exploiting the mobile capability of monitoring robots, good compositions can be determined based on composition principles such as the *Rule of thirds*, *Diagonal composition* and *Triangle composition*. A monitoring robot is supposed to move around during the events or ceremony to record memorable moments of activities. The optimal viewpoint is selected from a number of candidate viewpoints around the target group by estimating the composition of a scene.

This paper is composed of five sections. In Section 2, we will introduce the optimal viewpoint selection system (OVSS) for a robot photographer that can be used to search for the optimal viewpoint of a scene. In Section 3, we explain composition evaluation functions with visual balance, target size, and composition fitting value. In Section 4, we introduce a novel optimal viewpoint selection method. The optimal viewpoint is selected from a number of candidates around the target group by estimating the composition of targets. In Section 5, the results of a hardware experiment are given for the proposed viewpoint selection system for an omni-directional robot, where three composition evaluation functions introduced in Section 3 are employed.

2 Optimal Viewpoint Selection System

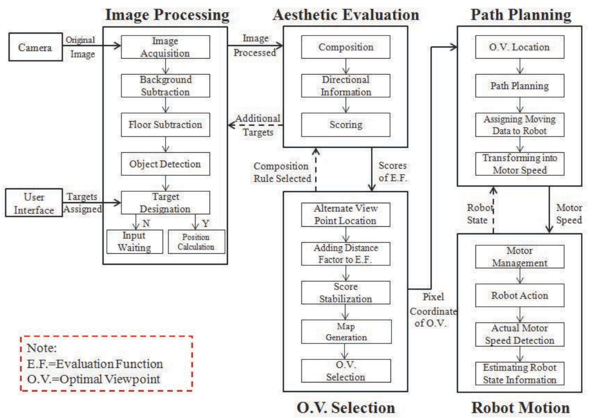


Figure 1. Optimal viewpoint selection system. Solid line shows the tasks that have been realized; dotted line shows future work.

In this paper, we propose an optimal viewpoint selection system for a monitoring robot to find the optimal viewpoint of a scene with the highest aesthetic property. This system consists of five modules: *Image Processing*, *Aesthetic Evaluation*, *Optimal Viewpoint Selection*, *Path Planning* and *Robot Motion*. Figure 1 shows the schematics of the proposed system.

In the *Image Processing* module, original images are inputed into the system along with a cloud image generated by an RGB-D camera. All targets are detected and their position and size in terms of the pixel coordinates of the images can be estimated. However, the detection process for objects and humans is different. For objects, after obtaining the original image and the corresponding cloud image, the background and floor of the scene will be subtracted by using a segmentation method called ransac [12]. The objects will be retained after the subtraction process. On the other hand, human detection depends on the face detection and recognition. We use the Shenzhen University frontal face detector, which is a free library provided by the Face Detection Data Set and Benchmark. We develop an interface for users so that they can select any object or human as a target from the operational screen. The objects or humans that are not selected as targets will be ignored in the composition.

In the *Aesthetic Evaluation* module, we propose a novel method to estimate the aesthetic properties of a scene. We develop the evaluation functions based on the composition rules with three elements: target size, visual balance, and composition fitting value. Then an evaluation score, reflecting the result of human subjective evaluation, will be generated by these functions. The highest score corresponds to the optimal composition for the recognized targets.

In the *Optimal Viewpoint Selection* module, we set some viewpoint candidates around the target group, presuming and evaluating the composition of targets observed from each position using the proposed composition evaluation methods. Then, the optimal viewpoint of all viewpoint candidates will be selected by using the composition evaluation result for each candidate and a distance factor. The distance factor corresponds to the energy and time efficiency as the robot is required to determine the optimal viewpoint within a limited time and power. As a result, we set the distance factor such that the monitoring robot can select the viewpoints near its present position.

In the *Path Planning* module, we can estimate the actual position of the optimal viewpoint in the robot coordinate system. In order to realize smooth and continuous motion of the robot, we design a gentle curve as the moving path of the robot. Although the path will update every three seconds, we apply a proportional feedback control to the robot before the path changes. The moving speed and direction of the robot are decided based on its distance from the goal position (optimal viewpoint) and the result of the feedback control, respectively. For an omni-directional robot, the rotation speed and orientation are decided by a visual feedback control.

In the *Robot Motion* module, the microcomputer of the robot controls the rotation speed and orientation of each wheel after receiving the movement information from the *Path Planning* module.

3 Aesthetic Evaluation by Element Judgment

3.1 Composition elements

Visual balance, region size, and composition fitting value are three important elements when considering the aesthetic properties of photographs. In this Section, we will define three subfunctions to evaluate these elements and create a combined function using them to determine the aesthetic values of scenes. We temporarily neglect the influence of the background because it does not change drastically when we try to observe a scene in a limited space. The subfunctions of visual balance and region size are the same for different compositions, while the fitting value subfunction varies with the composition rules because it is a special factor that reflects the features of different compositions. The mathematical formation of each element depends on the position and region size of targets; hence, we have to deal with the pixel information of targets when dealing with composition evaluation functions and subfunctions. Sample compositions used in our research are shown in Figure 2.

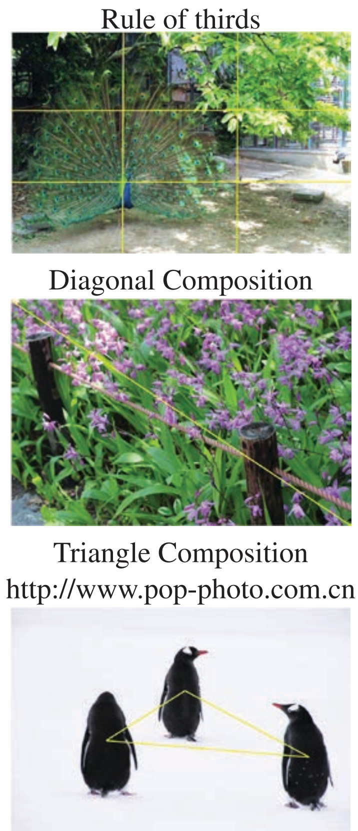


Figure 2. Typical compositions of photographs

Before discussing composition evaluation functions, we want to define 'distance' for computational aesthetics. In a digital photograph, a pixel is the smallest addressable and controllable element. Although the positions of pixels can be localized in a Cartesian coordinate system, it is not appropriate to measure the distance between two pixels using the traditional Euclidean distance as a pixel cannot be divided. As a result, we bring in the concept of Manhattan distance d to describe the relative distances between two points $p_1(x_1,y_1)$, $p_2(x_2,y_2)$. If w and h denote the width and height of a photo, respectively, a normalized Manhattan distance can be defined as follows:

$$d(p_1,p_2) = \frac{|x_1 - x_2|}{w} + \frac{|y_1 - y_2|}{h}. \tag{1}$$

3.1.1 Visual balance

We utilize the normalized Manhattan distance d_{vb} between the center of the target group (G) and the center of the image frame (C) as a measurement for visual balance,

$$d_{vb} = d(G,C). \tag{2}$$

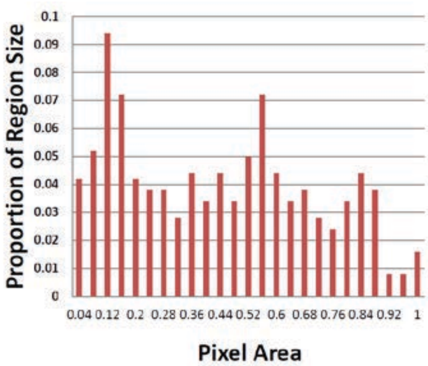
The definition of the visual balance subfunction E_{vb} is in the form of Equation (3). E_{vb} varies from 0 to 1 and a higher score implies better balance of vision. σ_1 is a constant of value 0.05, which is obtained from experiments.

$$E_{vb} = e^{-\frac{d_{vb}^2}{\sigma_1}}. \tag{3}$$

3.1.2 Region size

The region size is determined by the ratio of the target size to the photo size. Liu et al. [8] conducted a survey by considering over 200 professional photographs to confirm the appropriate size of a target which makes it look the most appealing. They plotted a histogram of the result, which is shown in Figure 3(a). Three dominant peaks were found that represent human preference when they appreciate artistic works in aesthetic vision.

(a) Survey result of Liu et al.



(b) Approximate result by Gaussian mixture.

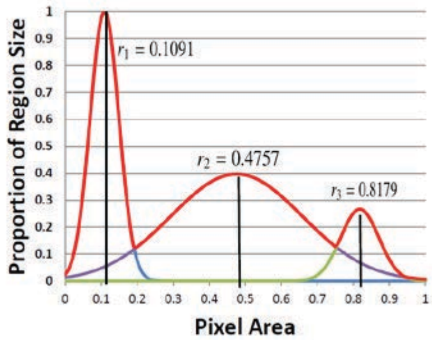


Figure 3. Inclination of human on region size

If r is the ratio of the target size to the photo size, the targets in aesthetic photos are mostly distributed around the values: $r_1 = 0.12$, $r_2 = 0.56$, and $r_3 = 0.82$. We use this result for size evaluation. As the region size subfunction should be applied for targets of all sizes, we transform these discrete values into three continuous Gaussian curves corresponding to the three dominant peaks. For each value on the horizontal axis, we choose the highest value from the three Gaussians and link all of them. Finally, we normalize the largest vertical value to 1 because we need the subfunction E_{sz} to vary from 0 to 1. $M(S_i)$ is the product of region area $A(S_i)$ and a parameter that corresponds to its importance value $I(S_i)$. $M(S_i) = A(S_i)I(S_i)$. n denotes the number of targets. The subfunction E_{sz} can be defined as follows:

$$E_{sz} = \max_{j=1,2,3} e^{-\frac{\left(\sum_{i=1}^{n} M(S_i) - r_j\right)^2}{\omega_j}}, \tag{4}$$

where $r_1 = 0.109, r_2 = 0.476, r_3 = 0.818$ from the curve of Figure 3(b). w_1, w_2, w_3 are constants with values 0.003, 0.067, and 0.006 determined by experimental results.

3.1.3 Composition fitting value

Every composition has particular characteristics that make it distinct from others. In this paper, we employ three kinds of compositions as an example and construct a fitting value subfunction to measure the similarity for each composition rule.

a) Rule of thirds

The guideline states that an image should be imagined to be divided into nine equal parts by two equally spaced horizontal lines and two equally spaced vertical lines, and the important compositional elements should be placed on their intersections P_j. The intersections play a very important role in evaluating the positions of targets in this kind of composition. The *Rule of thirds* is an application of the golden ratio [13]. High aesthetic evaluation will be obtained near an intersection. If we use $d(S_i) = \min_{j=1,2,3,4} d(C(S_i), P_j)$ to represent the minimum Manhattan distance between the center of targets $C(S_i)$ and intersection P_j, the fitting value subfunction for *Rule of thirds* E_{rt} can be formulated as follows:

$$E_{rt} = \frac{\sum_{i=1}^{n} M(S_i) e^{-\frac{d_{S_i}^2}{\sigma_2}}}{\sum_{i=1}^{n} M(S_i)}. \tag{5}$$

σ_2 is a constant with value 0.04. The results of this subfunction vary from 0 to 1.

b) Diagonal composition

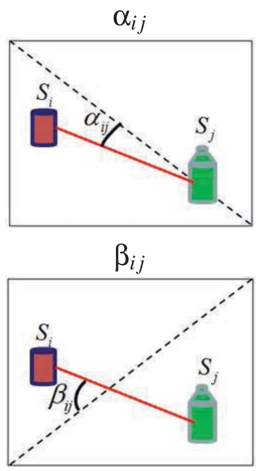


Figure 4. Angles of diagonals and the crossing line

In diagonal composition, the centers of all targets are positioned in a line coinciding with the diagonals of frame, as shown Figure 3(b). By calculating the absolute values of the angles with a diagonal, which are smaller than $\pi/2$, we design a function E_{dr} for this kind of composition. Because there are two diagonals in one frame, two sets of angles α_{ij} and β_{ij} will be formed as shown in Figure 4. The angles in the same set are formed with the same diagonal. We compute E_{dr} with two sets of angles as follows:

$$E_{dr1} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} e^{-\frac{4\alpha_{ij}^2}{\pi^2}}, \tag{6a}$$

$$E_{dr2} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} e^{-\frac{4\beta_{ij}^2}{\pi^2}}. \tag{6b}$$

Then, we choose the larger value as the evaluating result,

$$E_{dr} = max\{E_{dr1}, E_{dr2}\}. \tag{7}$$

c) Triangle composition

Triangle composition is also a common composition which has a wide usage in photography. Some kind of particular triangles, for instance, isosceles triangles, regular triangles and right angled triangles, are acknowledged to have a mysterious beauty introduced in many works [14] with a mathematical explanation. However, triangle composition doesn't demand targets, which are usually treated as vertexes placed in a position where they can shape a particular triangle. Triangles are in a more flexible form to professional photographers. We can merely find targets in a shape of particular triangles from any artistic works.

As the golden ratio is now well-accepted as a famous aesthetic and natural law, we use the golden ratio to construct this subfunction. We focus on the pixel area of the hollow triangle formed by the targets as a measurement for the triangle composition. We first prescribe a model triangle where the ratio of the width to the longer side of the picture is equal to the golden ratio while the ratio of the height to the shorter side is also equal to the golden ratio. We compute the pixel areas of the model triangle and formulate the fitting value subfunction of the triangle composition E_{tr} as Equation (8).

$$E_{tr} = \exp\left\{-\left(\frac{S_0}{S\sigma_3} - \frac{\Phi^2}{2\sigma_3}\right)\right\}, \tag{8}$$

where S_0 is the pixel area of the hollow triangle formed by objects; S shows the pixel area of the entire frame; Φ represents the golden ratio; σ_3 is a constant equal to 0.04 determined through experiments.

3.2 Weight determination

In order to determine the importance of each element for composition evaluation, we conducted a questionnaire survey for 17 ($m = 17$) people. We prepared 50 ($M = 50$) photos for the *Rule of thirds*, *Diagonal composition*, and *Triangle composition* and allowed the participants to evaluate their aesthetic properties with a score of 1 to 5. As humans have different personal preferences, we received different scores for the same sample photo. Assuming that the score of Photo no. p from the qth tester is recorded as h_p^q, we can calculate the average score of each sample photo to represent the human aesthetic evaluations for it. The average score for Photo no. p ($\bar{h}_p$) is as follows:

$$\bar{h}_p = \frac{1}{m}\sum_{q=1}^{m} h_p^q. \tag{9}$$

As the results of the evaluation function should have a positive correlation with the human evaluation, we examined the consistency between the subfunction values and the evaluation scores. Taking the *Rule of thirds* as an example, we plotted the distribution of the subfunction results and human evaluation scores, as shown in Figure 5. Applying the least squares method, we confirmed that a straight line shows the variation tendency of the results and then made use of the root mean square error (RMSE) to evaluate the dispersion of the three distributions based on the variation tendency. The expected values can be estimated from this straight line. If realistic values of E_{rt}, E_{vb}, and E_{sz} for Photo no. p are recorded as E_{rt}^p, E_{vb}^p, and E_{sz}^p, the expected values are $\bar{E}_{rt}^p$, $\bar{E}_{vb}^p$, and $\bar{E}_{sz}^p$, then, RMSE can be calculated as follows:

$$R_{rc} = \sqrt{\frac{1}{M}\sum_{p=1}^{M}(E_{rt}^p - \bar{E}_{rt}^p)^2}, \tag{10a}$$

$$R_{vb} = \sqrt{\frac{1}{M}\sum_{p=1}^{M}(E_{vb}^p - \bar{E}_{vb}^p)^2}, \tag{10b}$$

$$R_{sz} = \sqrt{\frac{1}{M}\sum_{p=1}^{M}(E_{sz}^p - \bar{E}_{sz}^p)^2}. \tag{10c}$$

Our evaluation function should have a positive correlation with human subjective evaluation. Therefore, we assume that each subfunction has a positive correlation with human evaluation. We analyzed the consistency between each subfunction and human evaluation and removed those subfunctions with low or negative correlation. The correlations between the subfunctions and humans are shown in Table 1.

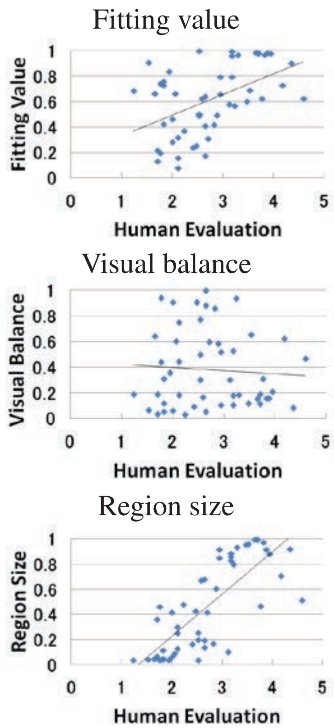


Figure 5. Distribution of subfunction and human evaluation

Table 1. Correlation Analysis

Correlation	E_{tr}	E_{vb}	E_{sz}
Human	0.49	-0.07	0.77

(a) Rule of thirds

Correlation	E_{tr}	E_{vb}	E_{sz}
Human	0.76	-0.10	0.75

(b) Diagonal composition

Correlation	E_{tr}	E_{vb}	E_{sz}
Human	0.72	0.70	0.13

(c)Triangle composition

As mentioned above, we select the subfunctions with high consistency with human evaluation to construct the composition evaluation functions. We normalized the reciprocals of the RMSE of the remaining subfunctions as their weights (w) in order to decrease the dispersion of functions distributions. The value of w for the *Rule of thirds* is determined as follows:

$$w_{rc} = \frac{\frac{1}{R_{rc}}}{\frac{1}{R_{rc}} + \frac{1}{R_{sz}}} = \frac{R_{sz}}{R_{rc}+R_{sz}}, \tag{11a}$$

$$w_{sz} = \frac{\frac{1}{R_{sz}}}{\frac{1}{R_{rc}} + \frac{1}{R_{sz}}} = \frac{R_{rc}}{R_{rc}+R_{sz}}. \tag{11b}$$

Composition evaluation functions E_{rc}, E_{dc}, and E_{tc} are as follows:

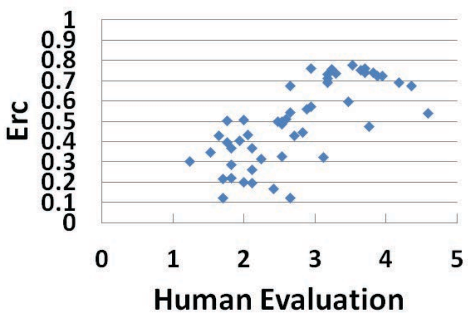
$$E_{rc} = 0.49E_{rt} + 0.51E_{sz}, \tag{12a}$$

$$E_{dc} = 0.50E_{dr} + 0.50E_{sz}, \tag{12b}$$

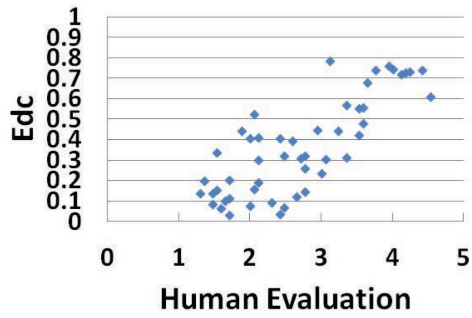
$$E_{tc} = 0.63E_{tr} + 0.37E_{vb}. \tag{12c}$$

3.3 Consistency with human evaluation

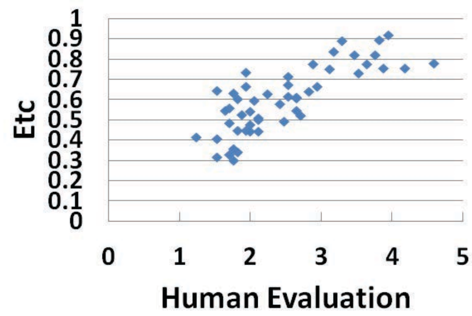
Figure 6 shows the distribution of the function score and the human evaluation result for sample photos of the questionnaire. The abscissas are the evaluation scores of humans and the ordinates are scores from our functions. In order to examine the consistency between the system and human evaluation, we employ the correlation coefficient as a measurement. If value of the correlation coefficient is over 0.7, it indicates a strong correlation. After calculating the correlation coefficients for Figure 6 (a), (b), and (c), we obtained the results of 0.76, 0.82, and 0.84, which shows high consistency between our composition evaluation functions and human subjective evaluation.



(a)Rule of thirds



(b)Diagonal composition



(c)Triangle composition

Figure 6. Distribution of the function score and human evaluation

4 Viewpoint Selection

We developed a novel algorithm for robots to search for the best observation position [15]. By estimating the relationships among targets in different observation positions based on the present coordinates and angles in the real world, we can obtain the results of each parameter before a robot reaches the position where optimal composition can be obtained. Figure 7 shows the parameters used to obtain pixel information in different observation positions. Pixel coordinates (x_i', y_i') and area $M'(S_i)$ of target i are estimated as follows:

$$x_i' = \frac{2}{W}\frac{\tan\theta_i}{\tan\varphi}, y_i' = \frac{2}{V}\frac{\tan\gamma_i}{\tan\varphi_2}, \tag{13-1}$$

$$M'(S_i) = M(S_i)\left(\frac{d_i}{d}\right)^2, \tag{13-2}$$

where W and V denote the pixel width and height of a picture's frame, d is the distance between the target i and the camera, while d_i is distance between the expected observation position and the camera.

Using x_i', y_i', and $M'(S_i)$, we can estimate the values of the factors of different composition rules.

Then, the total evaluation functions can be calculated by using equation (10) for all the alternate positions. Hence, when a monitoring robot is at one observation position, we can evaluate the aesthetic values of all other alternate positions. The best observation position is determined with the highest score according to the evaluation functions. Figure 8 shows a viewpoint map from a vertical view. Here, different function scores are represented in different colors. Higher scores are shown in blue color while lower scores are shown in red. S represents the current position of the monitoring robot, while G is the best observation position estimated.

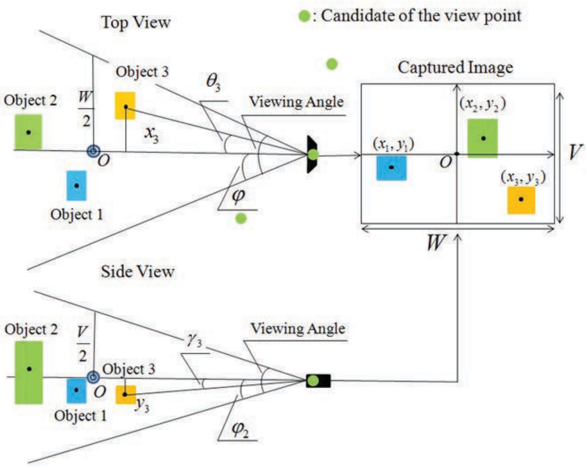


Figure 7. Top view and side view of argets

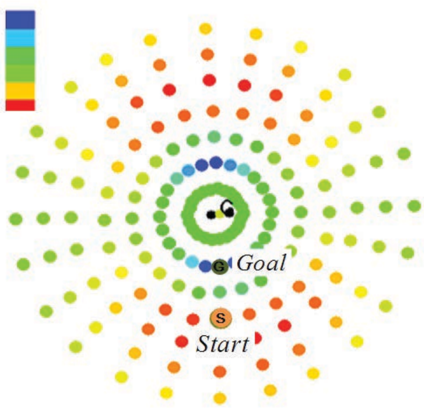


Figure 8. Composition map

5 Experiment

5.1 Experiment condition

Figure 9 shows the experimental condition. The monitoring robot is an omni-directional mobile robot, which can move in all directions with a PC and an RGB-D camera (Xtion Pro Live). The motors of the robot are controlled by the Arduino microcomputer. The targets of the photographs can be both objects and humans.

We performed two groups of experiments. In the first group, the robot was placed at a random initial position. Based on each composition principle, we repeated viewpoint selection experiments 10 times proving that our viewpoint selection system has high robustness and success ratio. In the second group, we changed the importance of each target in the scene showing the capability of our evaluation function to compose various aesthetic compositions based on human preference.

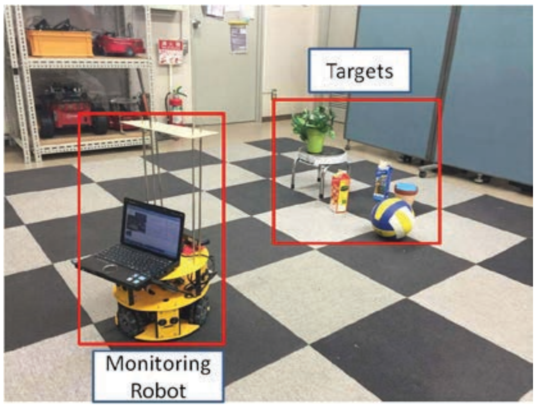
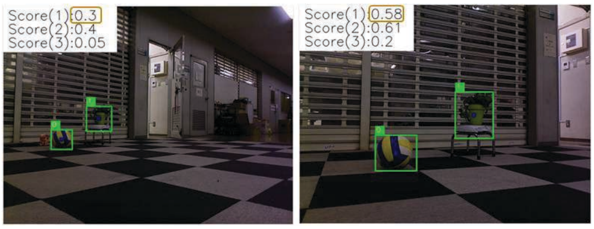


Figure 9. Experimental condition

5.2 Experiment results

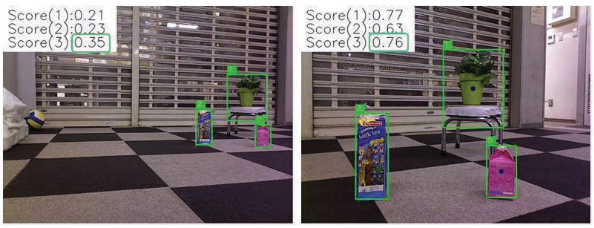
Figure 10 and Figure 11 show the initial and optimal scenes captured using an RGB-D camera for the first experiment. The pictures on the left show the composition of the initial viewpoints and those on the right are from the optimal viewpoint selected by our system. Figure 10 show the compositions of humans considering the face orientation. Figure 11(a) is judged by the *Rule of thirds*, Figure 11(b) is based on *Diagonal composition*, while Figure 11(c) is based on *Triangle composition*. Figure 12 shows the evaluation score for each frame in the viewpoint search process.

Figure 10. Initial and optimal compositions for humans



(a) Rule of thirds

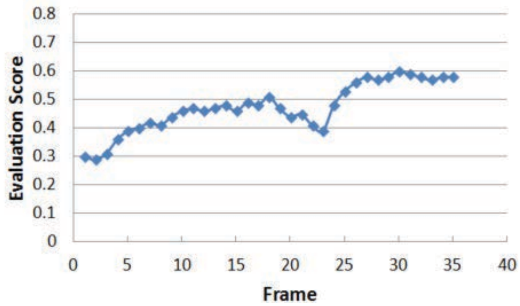
(b) Diagonal composition



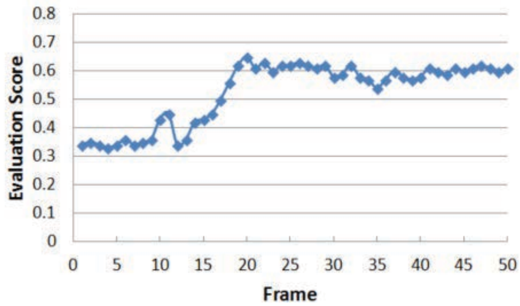
(c) Triangle composition

Figure 11. Initial and optimal compositions for objects

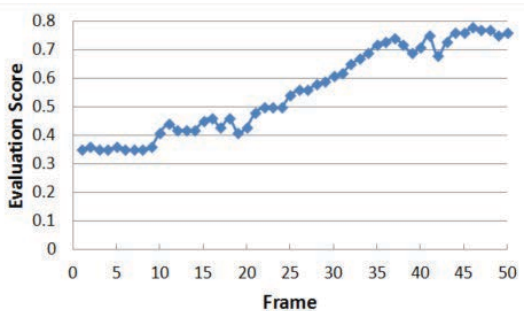
We repeated the experiment 10 times and recorded the evaluation score for the initial and optimal viewpoint, as shown in Figure 13. Although the values of the evaluation score for the initial scenes vary drastically due to different initial viewpoints, the score for the optimal viewpoints has low variance. Compared to the initial compositions, the aesthetic properties of scenes in the optimal viewpoint selected by our system have been improved to approximately 196%, 65%, and 94%.



(a) Rule of thirds



(b) Diagonal composition



(c) Triangle composition

Figure 12. Evaluation score for each frame in the viewpoint search process

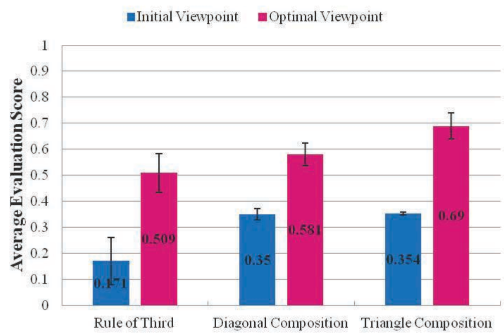
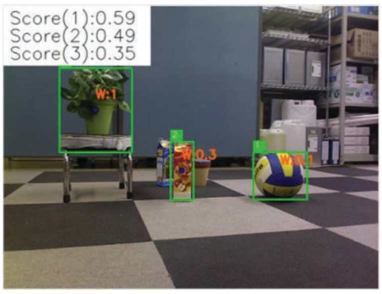
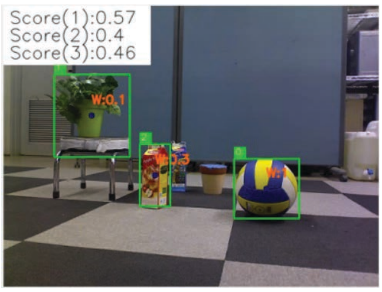


Figure 13. Evaluation scores of initial and optimal compositions

Considering the situations in which people take photos, the targets in the same scene usually play different roles for composition; hence, each target in a scene will have its own importance for the composition. The optimal composition changes when the main target, on which the focus is expected to be, changes. In the second experiment, each target had different importance although the actual arrangement of targets is retained during the experiments. We selected three targets from all the candidates and assigned different importance values to them (0.1, 0.3, 1.0) in each experiment. The robot starts from the same initial position and the same initial orientation. By changing the arrangement of the importance values and the target order, our optimal viewpoint selection system can obtain different optimal compositions based on the emphasis of different targets. Figure 14 shows the optimal composition based on the evaluation function of the *Rule of thirds*. First, we should concentrate on the plant and pay least attention to the volleyball; hence, when composing the image, the plant will be first placed at the position of a crossing point. Then, we focus on the volleyball and reduce the influence of the plant; hence, the ball is expected to be deposited at the position of a crossing point. Our optimal viewpoint selection system can result in an adaptable composition based on human preference.



(a) Plant emphasized



(b) Ball emphasized

Figure 14. Optimal composition considering the importance of targets

6 Conclusions

In this paper, we proposed a viewpoint selection system for a monitoring robot to search for the optimal viewpoint of a scene with the highest aesthetic property. We demonstrated that different composition rules in the art field can be computationally evaluated. Then, a novel method for predicting human aesthetic sense for scenes was proposed. We constructed three evaluation functions based on the composition rules with three elements. Our evaluation functions were proven to have high consistency with human evaluation. The result showed that the aesthetic evaluation of the optimal viewpoints selected by our system could be greatly improved compared with initial compositions.

However, we did not consider the background of photographs. Since the backgrounds influence the aesthetic values, a background evaluation element needs to be introduced in our future work.

References

[1] Faria J, Bagley S, Ruger S, et al. Challenges of finding aesthetically pleasing images. Proc. of the 14th IEEE International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), 2013, pp. 1-4.

[2] Dunstan B, Composing your paintings, Watson-Guptill Publications, 1971.

[3] Krages B, Photography: the art of composition, Skyhorse Publishing, Inc., 2012.

[4] Datta R, Joshi D, Li J, Studying Aesthetics in Photographic Images Using a Computational Approach, Proc. of the European Conference on Computer Vision, 2006, pp. 288–301.

[5] Wu Y, Bauckhage C, Thurau C, The good, the bad, and the ugly: Predicting aesthetic image labels, Proc. of the 20th IEEE International Conference on Pattern Recognition (ICPR), 2010, pp. 1586-1589.

[6] Wong L K, Low K, Saliency-enhanced image aesthetics class prediction, Proc. of the 16th IEEE International Conference on Image Processing (ICIP), 2009, pp. 997–1000.

[7] Ito M, Sekiyama K, Optimal viewpoint selection for cooperative visual assistance in multi-robot systems, Proc. of the 2015 IEEE/SICE International Symposium on System Integration (SII), 2015, pp. 605–610.

[8] Liu L, Chen R, Wolf L, et al. Optimizing photo composition. Computer Graphics Forum, Blackwell Publishing Ltd, Vol. 29, No. 2, 2010, pp. 469–478.

[9] Zhang F L, Wang M, Hu S M, Aesthetic Image Enhancement by Dependence-aware Object Recomposition, IEEE Transactions on Multimedia, vol. 15, no. 7, 2013, pp. 1480–1490.

[10] Byers Z, Dixon M, Goodier K, Grimm C, Smart W, An Autonomous Robot Photographer, Proc. of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 3, 2003, pp. 2636–2641.

[11] Byers Z, Dixon M, Smart W D, Grimm C, Say cheese! Experiences with a robot photographer. AI magazine, Vol. 25, No. 3, 2004, pp. 37–46.

[12] Toyama K, Krumm J, Brumitt B, Wallflower: Principles and practice of background maintenance, Proc. of the 7th IEEE International Conference on Computer Vision, 1999, pp. 255–261.

[13] Martinez B, Visual forces, an introduction to design, Pearson College Division, 1995.

[14] Maor E, Trigonometric delights, Princeton university press, 2013.

[15] Lan K, Sekiyama K, Autonomous Viewpoint Selection of Robots Based on Aesthetic Composition Evaluation of a Photo, Proc. of the 2015 IEEE Symposium Series on Computational Intelligence, 2015, pp. 295–300.

K. Lan received his ME from Nagoya University, Aichi, Japan, in 2016; he is currently working toward a PhD with the Department of Micro-Nano Systems Engineering at Nagoya University. His research interests include aesthetic evaluation of scenes for robots.

K. Sekiyama received the BE degree in 1992, the ME degree in 1994 and Dr. Eng. in 1997 from Nagoya University, Japan. From 1998 to 2001, he was a Lecturer at the Science University of Tokyo and, he was an associate professor at University of Fukui from 2001 to 2006. Since 2006, he has been an associate professor at Department of Micro-Nano Systems Engineering, Nagoya University. He has involved in distributed autonomous systems, cognitive robot systems, artificial intelligence, multi-locomotion robot systems and DNA computing. He is a member of Robotic Society of Japan (RSJ), Japan Society of Mechanical Engineering (JSME) and Society of Instrument and Control Engineers (SICE).