Human Tracking by a Multi-rotor Drone Using HOG Features and Linear SVM on Images Captured by a Monocular Camera

Yusuke Imamura, Shingo Okamoto, and Jae Hoon Lee, *Member, IAENG*

***Abstract*—In recent years, many researches of the drone (unmanned vehicles) have been carried out. Above all, the drone such as a multi-rotor craft might monitor and track suspicious person and find sufferers from disasters because it can move freely in the air. In the present research, a method was proposed that a multi-rotor drone can track a human by processing the two-dimensional images captured by a monocular camera installed on the multi-rotor drone. Furthermore, it can detect human without the differences of colors and movements of a target by using the Histograms of Oriented gradients (HOG) features and the linear Support Vector Machine (SVM). Then, it was shown that the multi-rotor drone could track a human by the proposed method.**

***Index Terms*—Multi-rotor drone, Human tracking, Image processing, HOG, Linear SVM.**

I. Introduction

IN recent years, the progress and evolution of drones (unmanned vehicles) are remarkable, and many researches of the drones have been carried out [1]-[3]. Above all, the drone such as a multi-rotor craft might monitor and track suspicious person and find sufferers from disasters because it can move freely in every direction in the air. However, the general robots as well as the multi-rotor drones may lose sight of targets like humans or objects in tracking. Therefore, the some researches that a multi-rotor drone tracks a target are performed. Juhng-Perng Su *et al.* and Tayyad Naseer *et al.* report that a multi-rotor drone tracks an object or human by detecting a targeted object with a stereo camera or a depth camera [4]-[5]. However, the flight-time of a multi-rotor drones become short in using a stereo camera or a depth camera because they consume electric power of a battery much. Furthermore, it takes much time to perform an image processing in using three-dimensional data captured by a stereo camera or a depth camera. For those reasons, Thomas Müller and Markus Müller took notice to a monocular camera because it is more lightweight and reasonable than a stereo camera or a depth camera as well as the processing time becomes short since two-dimensional data are used in an image processing [6]. Then, they carried out the research where a multi-rotor drone tracks a human who has a different color against background colors by using two-dimensional images captured by a monocular camera. Furthermore, Gonzalo R. Rodríguez-Canosa *et al.* performed the research where a multi-rotor drone track moving human and object in which they are expressed by vectors in two-dimensional images captured by a monocular camera [7]. Furthermore, Ashraf Qadir *et al.* reported that an unmanned miniature plane tracks an object by detecting the image similar to an image called “template” in two-dimensional images captured by a monocular camera [8]. In general, the detection method using the color has difficulty in detecting a target when the color of a target is similar to those of the background. Then, another detection method using the movements of a human and an object is difficult to detect a stationary target. The detection method using the image of “template” also has difficulty in detecting a pedestrian whose shape changes in the images because it is weak to a change of shape of a target.

In the present research, the method was proposed that a multi-rotor drone can track a human. The proposed method tracks a human while detecting a human without the differences of colors and the movements of a target. The Histograms of Oriented Gradients (HOG) features and the linear Support Vector Machine (SVM) were used in two-dimensional images captured by a monocular camera installed on the multi-rotor drone. Then, the validity and effectiveness of proposed method were examined in the experiment which a multi-rotor drone used.

II. Method to Detect a Human by a Monocular Camera

A. Detection of a Targeted Human

It is necessary to perform an image processing on images captured by a monocular camera in the case of detecting a human using the monocular camera. In the present research, the Histograms of Oriented Gradients (HOG) features and the linear Support Vector Machine (SVM) were used in detecting a human using a monocular camera [9]. The HOG features can express a rough shape of a human by making a histogram of the luminance-gradient vectors in a local area of an image. Then, the linear SVM is one of pattern recognition models.

Fig. 1 shows whole photo image, detection window, block, cell and image pixel. The whole photo image is an image captured by a monocular camera. The *o-xy* coordinate frame

Manuscript received December 8, 2015; revised January 20, 2016.
Every author is with Mechanical Engineering Course, Graduate School of Science and Engineering, Ehime University, 3 Bunkyo-chyo,Matsuyama 790-8577, Japan
E-mail: Yusuke Imamura < b480005m@mails.cc.ehime-u.ac.jp >,
Shingo Okamoto < okamoto.shingo.mh@ehime-u.ac.jp >,
Jae Hoon Lee < jhlee@ehime-u.ac.jp >

(Global coordinate frame) is fixed to the whole photo image. The o_w-$x_i y_{js}$ coordinate frame, namely Window coordinate frame is fixed to the detection window. The o_b-$x_k y_l$ coordinate frame, namely Block coordinate frame is fixed to the block. The o_s-$x_m y_n$ coordinate frame, namely Cell coordinate frame is fixed to the cell. The o_p-$x_p y_p$ coordinate frame, namely Pixel coordinate frame is fixed to the image pixel.

Fig. 2 shows luminance-gradient vectors in the case that the luminance value on coordinate(x_m, y_n) is denoted by L(x_m , y_n). Then, the luminance-gradient vector made in Fig. 2 denotes a feature in each image pixel. The magnitude, G_{mn} and argument, θ_{mn} of luminance-gradient vector on coordinate(x_m, y_n) are calculated by (1), (2), (3) and (4).

$$d_x(x_m , y_n) = L(x_{m+1} , y_n) - L(x_{m-1} , y_n) \quad (1)$$

$$d_y(x_m , y_n) = L(x_m , y_{n+1}) - L(x_m , y_{n-1}) \quad (2)$$

$$G_{mn} = \sqrt{d_x(x_m , y_n)^2 + d_y(x_m , y_n)^2} \quad (3)$$

$$\theta_{mn} = arctan \frac{d_y(x_m , y_n)}{d_x(x_m , y_n)} \quad (4)$$

The range of calculated θ_{mn} is generally from 0 to 2π [rad]. However, in the present research, the range was treated as from 0 to π [rad] excluding the direction of luminance-gradient vectors.

Fig. 3 shows histogram of luminance-gradient vectors. The histogram is made from the data of luminance-gradient vectors in each cell in a detection window. The luminance-gradient vector made in Fig. 2 denotes a feature in each image pixel. Then, the luminance-gradient vector is weak to a change of shape of a target. However, the histogram is strong to the change of shape because it is made in each cell. The histogram also denotes the sum of magnitude of luminance-gradient vectors in each π/9[rad] on the arguments of luminance-gradient vectors.

Fig. 4 shows normalized histograms on luminance-gradient vectors in a block. The histograms are normalized to adapt well the change of brightness. There are four histograms in each block. The histograms is normalized by (5) and (6).

$$G_{kl} = \sum_{m=1}^{8} \sum_{n=1}^{8} G_{mn} \quad (5)$$

$$G_{kl}{}^* = \frac{G_{kl}}{\left(\sqrt{\sum_{k=1}^{2} \sum_{l=1}^{2} G_{kl}{}^2}\right) + \varepsilon} \quad (\varepsilon = 1) \quad (6)$$

The G_{kl} denotes the sum of magnitude of luminance-gradient vector on coordinate(x_k, y_l). The G_{kl}* is the normalized G_{kl}. The ε is used to avoid that the denominator in (6) becomes zero.

Fig. 5 shows the way to raster-scan in a detection window. The raster-scan is performed by moving the block in a detection window. Then, the histograms are normalized in every moving the block in the detection window. At the start of the raster-scan, the upper-left corner of a block accords with the upper-left corner of a detection window. The direction of raster-scan is the left one. The distance, D_w by sliding of raster-scan is 1 [cell]. The block is moved to a new line by the line-feed width of 1 [cell] when the right edge of a block exceeds the right edge of a detection window. The raster-scan in a detection window finishes when the lower edge of a block exceeds the lower edge of a detection window.

Fig. 6 shows the flowchart for calculation of HOG features in a detection window. Then, the histograms normalized by the raster-scan denotes the HOG features in a detection window. In the present research, a Region Of Interest (ROI) that is a range where image processing is performed is set on a whole photo image. Because the processing time in using ROIs is shorter than that in using a whole photo image when the image processing to detect a human is performed.

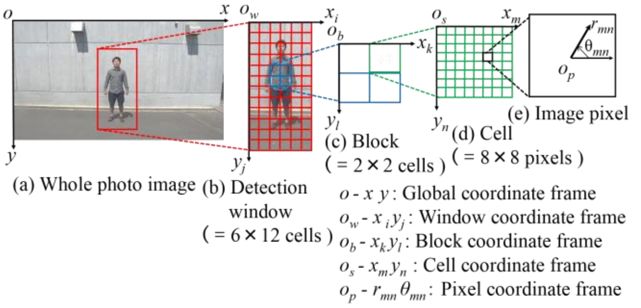

Fig.1. Whole photo image, detection window, block, cell and image pixel

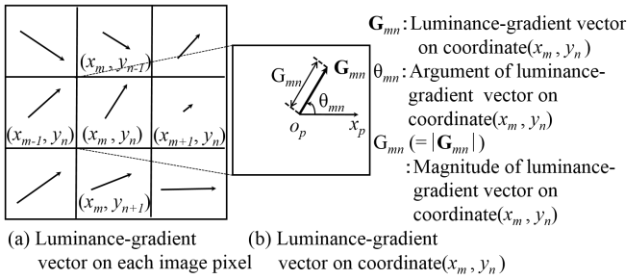

Fig. 2. Luminance-gradient vectors

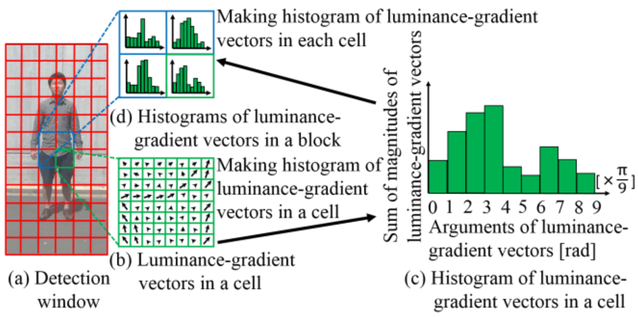

Fig. 3. Histogram of luminance-gradient vectors

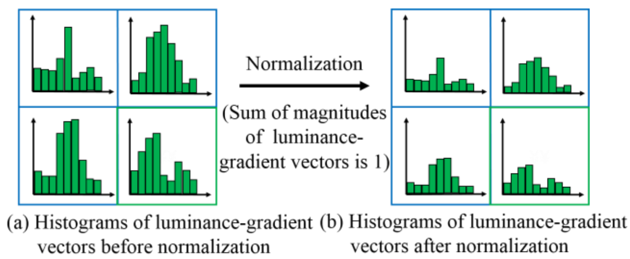

Fig. 4. Normalized histograms on luminance-gradient vectors in a block

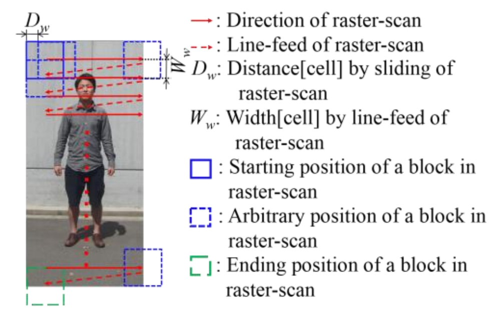

Fig. 5. Raster-scan in a detection window

Fig. 7 shows the way to raster-scan in a ROI. The raster-scan is performed by moving the detection window in a ROI. Then, the raster-scan in a detection window is performed in every moving the detection window in the ROI. At the start of the raster-scan, the upper-left corner of a detection window accords with that of a ROI. The direction of raster-scan and the distance, D_r by sliding of raster-scan as well as the width, W_r by line-feed of raster-scan never change in the raster-scan in a detection window. The detection window is moved to a new line when the right edge of a detection window exceeds that of a ROI. The ROI is reduced by the reduction ratio, α when the lower edge of a detection window exceeds that of a ROI. Then, the execution of raster-scan in a ROI are kept after making the upper-left corner of a detection window accord with that of a reduced ROI. The compression is performed until a reduced ROI becomes smaller than a detection window. The reduction of a ROI is performed until the size of a reduced ROI becomes smaller than that of a detection window. It is impossible to calculate the HOG features in the case that a human is inside a detection window when the size of a human is bigger than that of a detection window in a ROI. The detection windows where there are humans are detected by using the linear SVM and the HOG features of detection windows in an initial ROI and reduced ROIs. Then, a position of the upper-left corner of a detection window in a ROI reduced by α^N ,where α is the redaction ratio of a ROI and N is the number of times that an ROI is reduced, is converted into that in an initial ROI. The coordinates of upper-left corner of a detection window in a reduced ROI is expressed by $(x/\alpha^N, \ y/\alpha^N)$. The linear SVM is the learning model to perform a pattern recognition using the discriminant function. The linear SVM can learn the parameters of discriminant function. The discriminant function used in the present research is expressed by

$$f(\boldsymbol{x}) = sign(\boldsymbol{w}^T\boldsymbol{x}+b) = \begin{cases} 1 \text{ if } (\boldsymbol{w}^T\boldsymbol{x}+b) \geq 0 \\ -1 \text{ if } (\boldsymbol{w}^T\boldsymbol{x}+b) < 0 \end{cases} \tag{7}$$

where, $\boldsymbol{x}$ denotes the vector calculated using the HOG features in detection window, then $\boldsymbol{w}$ and b denote the vector and the scalar, respectively, calculated using the parameters learned by the linear SVM. For example, the $f(\boldsymbol{x})$ should result in 1 when $\boldsymbol{x}$ is the vector calculated using the HOG features in the detection window where there is a human. Then, the $f(\boldsymbol{x})$ should result in -1 when $\boldsymbol{x}$ is that in the detection window where there is no a human. In the present research, the values opened to the public in the Open Source Computer Vision Library (Open CV) were used as the $\boldsymbol{w}$ and b in the discriminant function [10].

Fig. 8 shows the renewal of an initial ROI. The height, h_r and breath, b_r of a renewed initial ROI are β_1 times h_w and β_2 times b_w, respectively, because the coordinates in the center of a renewed initial ROI coincides with those of a detected detection window.

Fig. 9 shows the flowchart for human detection. The image processings shown in Figs.1-8 are performed for the sequential whole-photo-images made from a moving image until the signal of termination command is read. It becomes possible to detect a human by the above method.

B. Position of a Human in a coordinate frame fixed to a monocular camera

Fig. 10 shows the position of a human in the camera coordinate frame fixed to a monocular camera. It is necessary to find a position of a human in the coordinate frame fixed to a monocular camera in the case of a human tracking using the whole photo image captured by the monocular camera. The o_c-$x_c y_c z_c$ coordinate frame denotes the camera coordinate frame fixed to the monocular camera. The P_c (x_c , y_c , z_c) [pixel] denotes the position of the human of center in the camera coordinate frame, o_c -$x_c\, y_c\, z_c$ fixed to the monocular camera, where B is breadth [pixel] and H is height [pixel] of a whole photo image.

Fig. 11 shows the relationship between the camera coordinate frame, o_c-$x_c\, y_c\, z_c$ fixed to a monocular camera and a global coordinate frame. It is possible to calculate the coordinates(y_c , z_c) [pixel], because the whole photo image is two-dimensional image. However, it is difficult to calculate the distance in the x_c –direction. Then, the x_c is calculated by (8) and (9) using the triangulation method with the angle, θ_0 [°] of view-field of a monocular camera and the coordinate, y_c [pixel].

$$\theta_1 = \theta_0 \times \frac{2|y_c| + b_w}{2B} \tag{8}$$

$$x_c = \frac{2|y_c| + b_w}{2tan\ |\theta_1|} \tag{9}$$

Where θ_0 is the angle [°] of view-field of monocular camera, θ_1 is the angle [°] between z_c-axis and the straight line that the slope is expressed by (| y_c |+ b /2) / x_c through the origin o_c in Fig. 11. It becomes possible to calculate the position of human of center in the camera coordinate frame, o_c -$x_c\, y_c\, z_c$ fixed to a monocular camera by the above method.

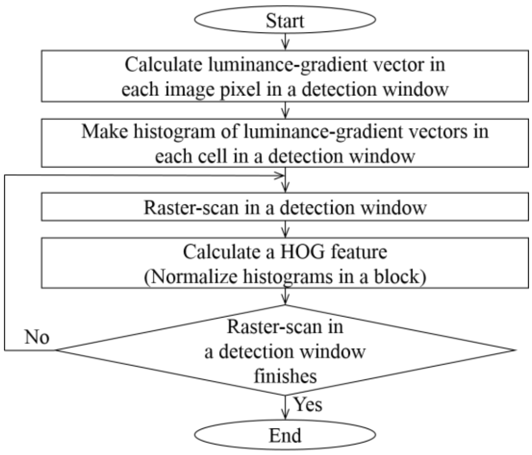

Fig. 6. Flowchart for calculation of HOG features in a detection window

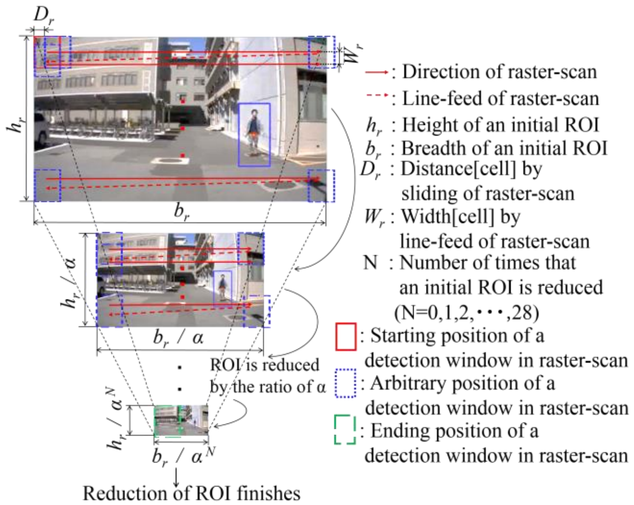

Fig. 7. Raster-scan in a ROI

ISBN: 978-988-19253-8-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

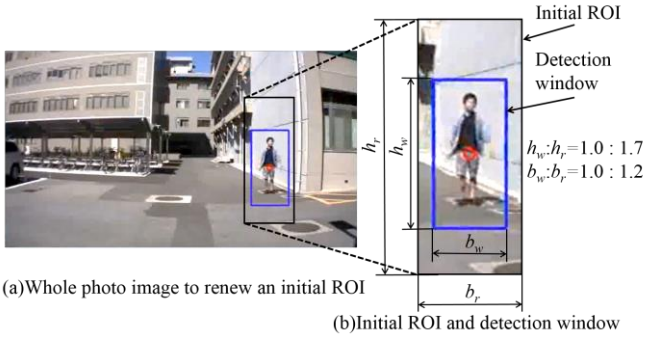

Fig. 8. Renewal of an initial ROI

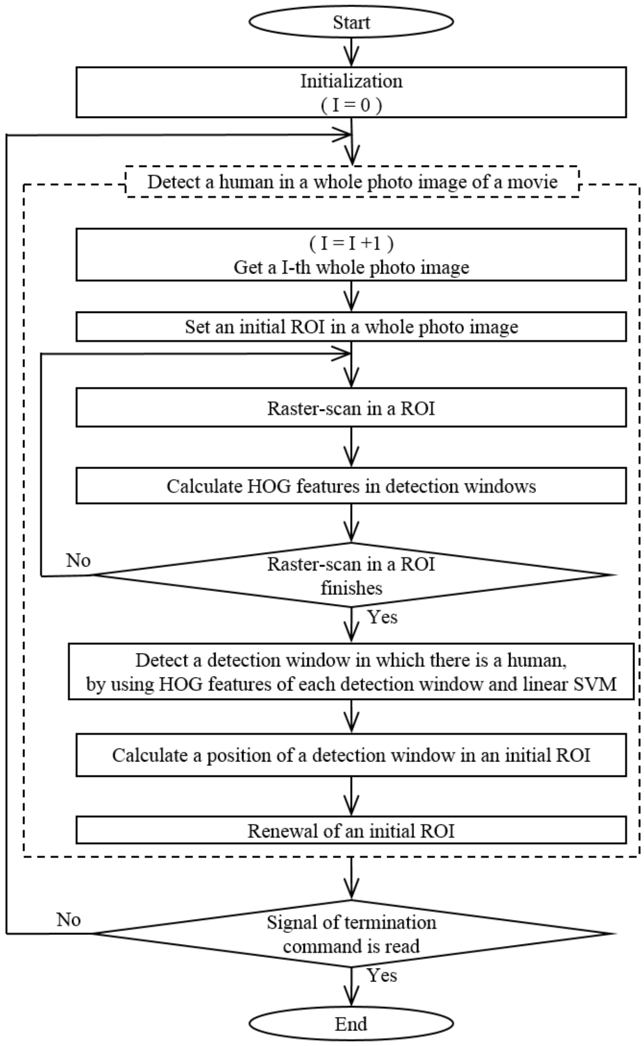

Fig. 9. Flowchart for human detection (Refer to Fig.6)

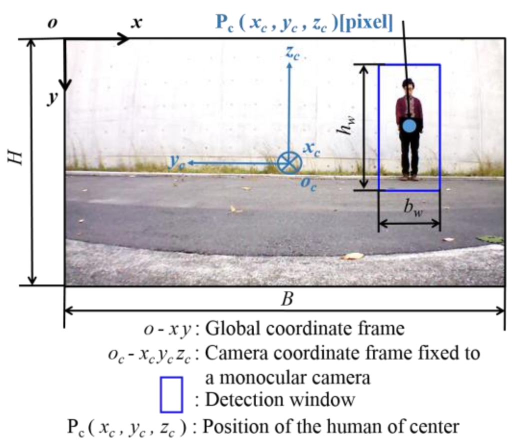

Fig. 10. Position of a human in the camera coordinate frame fixed to a monocular camera

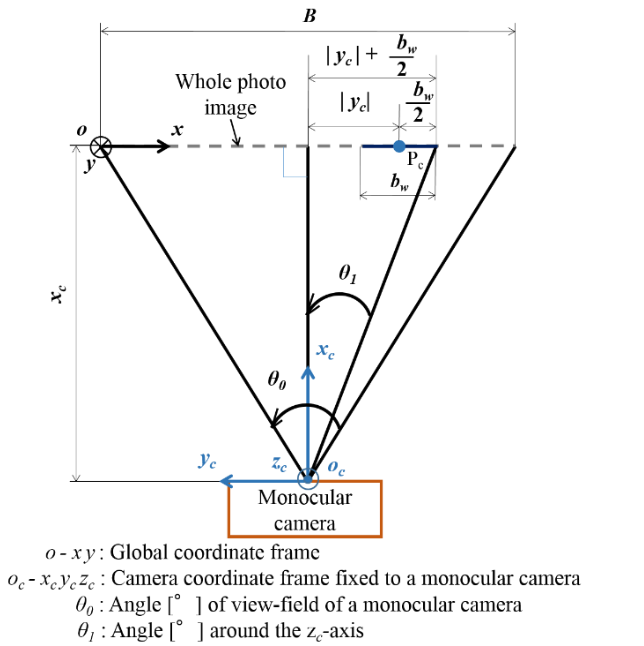

Fig. 11. Relationship between the o_c-$x_c y_c z_c$ camera coordinate frame fixed to a monocular camera and the global coordinate frame.

III. METHOD TO TRACK A HUMAN USING A MULTI-ROTOR DRONE

Fig. 12 shows the multi-rotor drone used as a platform in the present research. This drone is AR.Drone2.0 installing a monocular camera developed by Parrot Corporation [11].

Fig. 13 shows the configuration of control system for the multi-rotor drone. The drone can transmit the images captured by the monocular camera to a PC using the Wi-Fi communication. The PC firstly calculates the velocity such that the position P_c (x_c , y_c , z_c) of a human can come to the center ($y_c = z_c = 0$) of image sent from the drone to the PC. Then, the PC transmits the calculated velocity to the drone using the Wi-Fi communication.

Fig. 14 shows the coordinate frames used for human tracking using the multi-rotor drone. The o_d -$x_d y_d z_d$ in Fig. 4 denotes the coordinate frame fixed to the gravity center of the drone. The v_x [m/s] and v_z [m/s] are the translational velocities of the drone in x_d - and z_d - directions, respectively, in the o_d -$x_d y_d z_d$. The ω_z [rad/s] is the angular velocity of the drone around the z_d -axis in the o_d -$x_d y_d z_d$. The desired human position is $\overline{P_c}$($\overline{x_c}$ (= const.), $\overline{y_c}$ (= 0), $\overline{z_c}$ (= 0)) in Fig. 14.

Fig. 15 shows the block diagram of proportional control on the velocity of the multi-rotor drone for human tracking. It is necessary to give velocities such that P_c (x_c , y_c , z_c) comes to the center ($y_c = z_c = 0$) of images captured by the monocular camera to the drone while keeping the distance (x_c = const.) between the drone and a human when the drone tracks a human. Then, the translational velocities, v_x [m/s], v_z [m/s] and the angular velocity, ω_z [rad/s] that the drone can track a human are given by (10), (11) and (12).

$$v_x = K_x \, (\overline{x_c} - x_c) \tag{10}$$

$$v_z = K_z \, (\overline{z_c} - z_c) \tag{11}$$

$$\omega_z = K_{\theta z} \, (\overline{y_c} - y_c) \tag{12}$$

Where, the K_x [m/(pixel· s], K_z [m/(pixel· s] and $K_{\theta z}$ [rad/pixel· s] are the gains of proportional control.

Fig. 12. Multi-rotor drone used as a platform in the present research (AR.Drone2.0. Parrot Corporation)

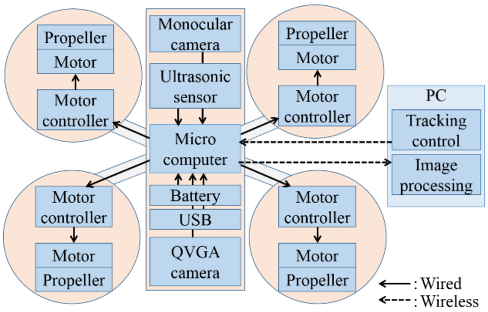

Fig. 13. Configuration of control system for the multi-rotor drone

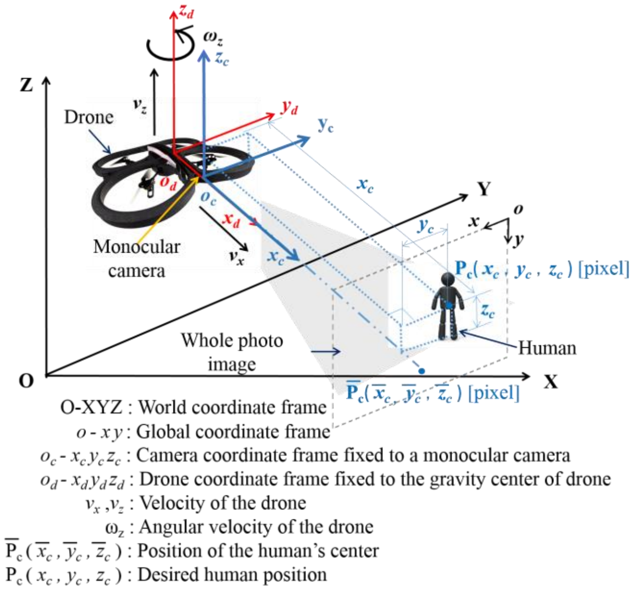

Fig. 14. Coordinate frames used for human tracking using the multi-rotor drone

Fig. 16 shows the flowchart for human tracking. The values of v_x [m/s], v_z [m/s] and ω_z[rad/s] are set to zero when a human is not detected in the ROIs. The human tracking by drone is performed until the signal of terminal command is read. The above method can make the drone track a human.

A. *Experimental Method*

Fig. 17 shows the experimental environment and the experimental result. In Fig. 17, the solid and broken lines show the trajectories of the drone and the human, respectively. The numbers, (a)-(h) on the trajectory of the drone, in Fig. 17 denote the positions at the times of the figures, (a)-(h) in Fig. 18. Then, the distance between the drone and the human was kept by 4.0 [m].

B. *Experimental Result*

Fig. 18 shows the experimental result that the multi-rotor drone could track the human under the experimental plane explained in the previous section. The Figs. (a)-(h) in Fig. 18 show the photo images captured in every 2.0 seconds when the drone had tracked the human.

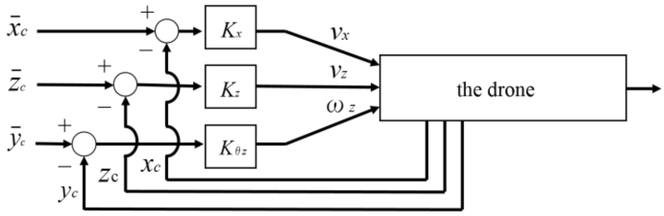

Fig. 15. Block diagram of proportional control on the velocity of the multi-rotor drone for human tracking

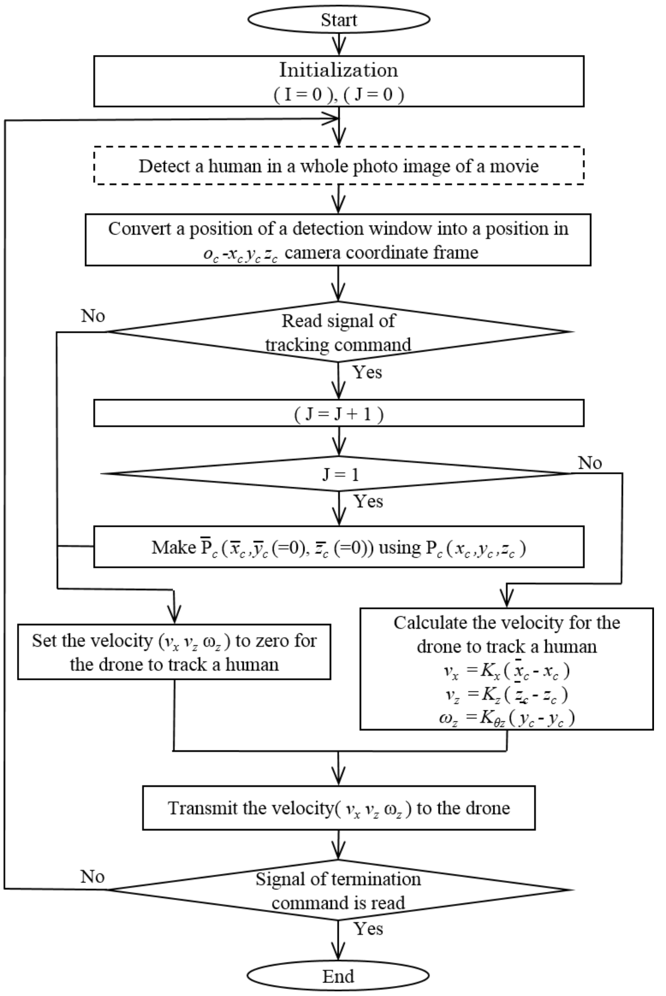

Fig. 16. Flowchart for human tracking (Refer to Fig.11)

IV. Experiment to Human Tracking

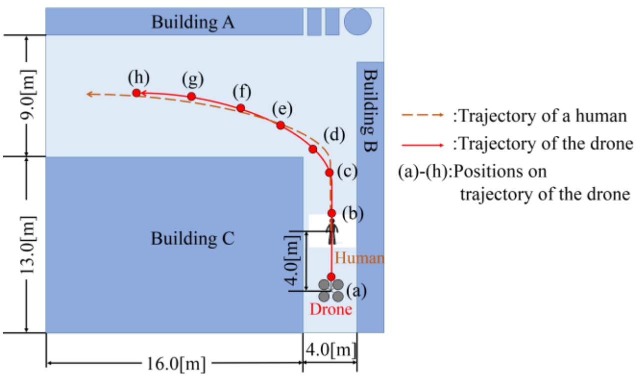

Fig. 17. Experimental environment and the experimental result (Refer to Fig.18)

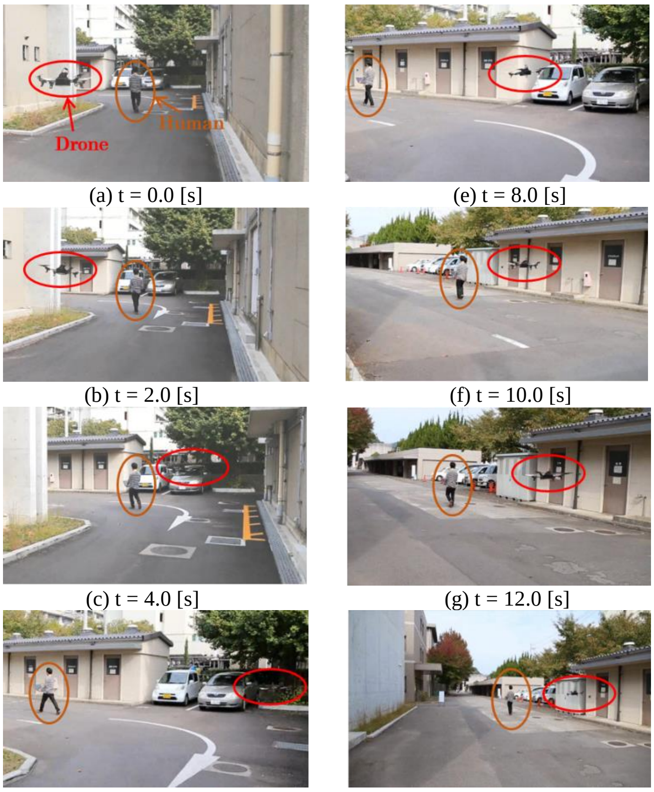

Fig.18. Experimental result that the multi-rotor drone could track the human (Refer to Fig.17)

V. CONCLUSION

In the present research, the method was proposed that a multi-rotor drone can track a human while detecting a human by using the HOG features and the linear SVM on images captured by a monocular camera installed on the drone. Furthermore, it was shown that the multi-rotor drone could track a human in the experiment where the proposed method was used.

APPENDIX

Fig. a-1 shows an example of the histogram of luminance-gradient vectors in a cell. In Fig. a-1(c), the 0~8 [×π/9] and monochrome gradation denote arguments divided in every π/9 [rad] from 0 to 2π [rad] and magnitudes, respectively, of luminance-gradient vectors. The monochrome gradation becomes white as the magnitudes becomes larger.

Fig. a-2 show an example of visualized histograms on luminance-gradient vectors. Fig. a-2 made by the method explained in Fig. a-1(c) shows an example of the visualized histograms of luminance-gradient vectors in a photo image of a detection window.

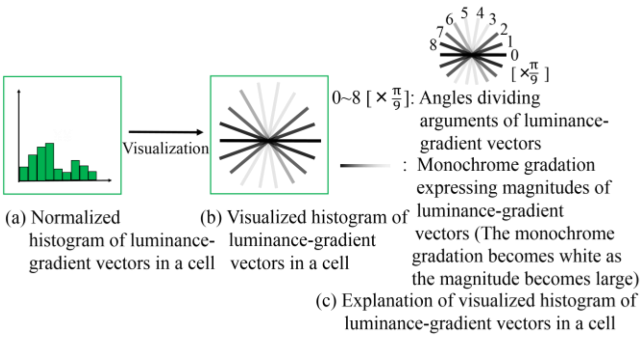

Fig. a-1. Example of the histogram of luminance-gradient vectors in a cell

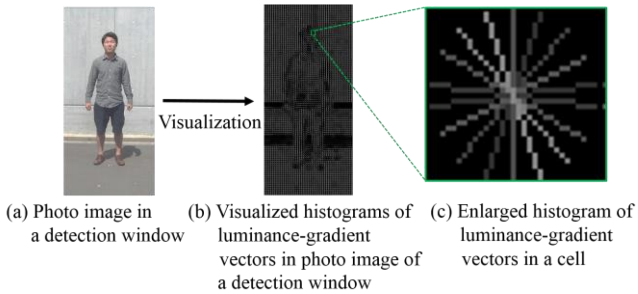

Fig. a-2. Example of visualized histograms on luminance-gradient vectors

REFERENCES

[1] Akhmad Taufik, Shingo Okamoto, and Jae Hoon Lee, "Multi-Rotor Craft with Single-Lens Camera that Can Autonomously Fly along a River", *Asia-Pacific Conference on Engineering and Applied Science (APCEAS)*, August, 25-27, 2015

[2] Akhmad Taufik, Shingo Okamoto, and Jae Hoon Lee, "Multi-Rotor Drone to Fly Autonomously along a River Using a Single-Lens Camera and Image Processing", *International Journal of Mechanical Engineering (IJME)*, vol. 4, issue 6, Oct-Nov 2015, pp.39-50

[3] Stefan Hrabar, "3D Path Planning and Stereo-based Obstacle Avoidance for Rotorcraft Unmanned Aerial Vehicles", *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Acropolis Convention Center Nice, September, 22-26, 2008

[4] Juhng-Perng Su and Kuo-Hsien Hsia, "Height Estimation and Image Tracking Control of an Indoor Quad-Rotor Craft via Multi-Vision Systems", *International Journal of Computer, Consumer and Control (IJ3C)*, vol. 2, no. 4, 2013

[5] Tayyad Naseer, Jurgen Sturm and Daniel Cremers, "FollowMe: Person Following and Gesture Recognition with a Quadrocopter", *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS)*, November, 3-7, 2013

[6] Thomas Müller and Markus Müller "Vision-based drone flight control and crowd or riot analysis with efficient color histogram based tracking", *SPIE 8020, Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications VIII*, 80200R, May, 25, 2011

[7] Gonzalo R. Rodríguez-Canosa, Stephen Thomas, Jaime del Cerro, Antonio Barrientos and Bruce MacDonald, "A Real-Time Method to Detect and Track Moving Objects (DATMO) from Unmanned Aerial Vehicles (UAVs) Using a Single Camera", *Remote Sensing 2012*, vol. 4, issue 4, 2012, pp.1090-1111

[8] Ashraf Qadir, Jeremiah Neubert, and William Semke, "On-Board Visual Tracking with Unmanned Aircraft System (UAS)", *Infotech@ Aerospace 2011*, March, 29-31, 2011

[9] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." *Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2005, pp. 886-893,

[10] Open CV, http://opencv.jp/

[11] AR.Drone2.0, http://ardrone2.parrot.com/