Journal of Robotics, Networking and Artificial Life
Vol. **8**(**2**); September (2021), *pp.* 134–138
DOI: https://doi.org/10.2991/jrnal.k.210713.013; ISSN 2405-9021; eISSN 2352-6386
https://www.atlantis-press.com/journals/jrnal

Research Article

Detection of a Fallen Person and its Head and Lower Body from Aerial Images

Joo Kooi Tan*, Haruka Egawa

Department of Mechanical and Control Engineering, Kyushu Institute of Technology, Sensuicho 1-1, Tobata, Kitakyushu 804-8550, Japan

ARTICLE INFO

Article History

Received 23 November 2020
Accepted 23 May 2021

Keywords

Fallen person detection
head and lower body detection
aerial images
rotation-invariant HOG
rotation-invariant LBP
Random Forest
peak of gradient histograms

ABSTRACT

This paper proposes a method of detecting a person fallen on the ground and its head and lower body from aerial images. The study intends to automate discovering victims of disasters such as earthquakes from areal images taken by an unmanned aerial vehicle (UAV). Rotation-invariant histogram of oriented gradients and rotation-invariant local binary pattern are used as features describing a fallen person so as to detect it regardless of its body orientation. The proposed method also detects the head and the lower body of a fallen person using the peak of the gradient histogram. Experimental results show satisfactory performance of the proposed method.

1. INTRODUCTION

Japan is one of the world's leading countries of earthquakes [1] and has been hit by disasters caused by earthquakes. In the event of such a disaster, it is often very difficult to search for the victims by human search activities. Therefore, recently, attention has been paid on the development of an automatic search method for those victims from aerial images taken by UAVs.

There are various automatic human detection methods for searching people [2–4], but most of them are based on the premise that the target fallen person's body orientation is unified and the subject is upright [3,4]. However, there is obviously no uniformity in the orientation of a fallen person's body taken with an UAV. Therefore, in this paper, we propose an automatic detection method of a fallen person that does not depend on its body orientation.

The proposed method employs Rotation-invariant Histogram of Oriented Gradients (Ri-HOG) features [5–7] and Rotation-invariant Local Binary Pattern (Ri-LBP) [8–10] as the features describing texture objects, or a person fallen on the ground, since they are the features robust to object rotation. In addition, Random Forest [11,12] is used for designing a classifier.

Having detected a fallen person [9], the proposed method also detects the head and lower body areas using the peak value of a gradient histogram. This enables quick support of the fallen person for communication, medical care, food, etc.

The proposed method is examined its performance by the experiment using areal images.

2. MATERIALS AND METHODS

2.1. Ri-HOG Features

The Ri-HOG feature [5,6] is the HOG feature [13] with rotational invariance. Unlike the original HOG, which uses rectangular cells, the cell arrangement derived from dividing concentric circles is used in Ri-HOG feature. This section describes the Ri-HOG feature.

2.1.1. Calculation of luminance gradient information

The input color image is gray-scaled and the luminance gradient at each pixel is calculated. From the obtained luminance gradient, the intensity and the direction of the gradient are calculated in the Cartesian coordinate system.

Assuming that the origin O is at the center of an input image, the relative luminance gradient direction θ' in the polar coordinate system is defined by the difference between the luminance gradient direction $\theta(x, y)$ of a pixel at (x, y) and the declination $\varphi(x, y)$ of the pixel provided by the polar coordinate system, as given by

$$\theta' = \theta(x, y) - \varphi(x, y) \tag{1}$$

*Corresponding author. Email: etheltan@cntl.kyutech.ac.jp

The value θ' is quantized to eight directions having an interval of 45°.

2.1.2. Creating a 2-D histogram

In the method, a concentric circle consisting of three (large, medium and small) circles is divided into 36 areas in the angular direction, and each small area is defined as a cell (The total number of cells is 108.). The cells in the smallest concentric circle are given numbers as 1, 4, 7, …, 106 clockwise starting from the area in the 0° direction. Similarly, the cells in the middle circle are numbered 2, 5, 8, …, 107 and the cells in the largest circle 3, 6, 9, …, 108.

Given a pixel $\boldsymbol{p}(x,y)$ in a cell, six offset regions $W_1(\boldsymbol{p})$, $W_2(\boldsymbol{p})$, …, $W_6(\boldsymbol{p})$ are further defined by dividing a semicircular region existing in the radial direction of $\boldsymbol{p}$. Let $\boldsymbol{q}$ be a pixel in one of the offset regions of $\boldsymbol{p}$. A 2-D histogram is calculated with pixels $\boldsymbol{p}$ and $\boldsymbol{q}$ to use the co-occurrence information with respect to their gradient intensities and luminance gradient directions. Six 2-D histograms are created with each cell as it has six offsets. The 2-D histogram is calculated using the following formula:

$$H_{ij}(o_1, o_2) = \sum_{\boldsymbol{p}\in S_i} \sum_{\boldsymbol{q}\in W_j(\boldsymbol{p})} (m(\boldsymbol{p}) + m(\boldsymbol{q}))K(\theta'(\boldsymbol{p}) \in o_1)K(\theta'(\boldsymbol{q}) \in o_2) \quad (2)$$

Here, i (i = 1, 2, …, 108) is the cell number, j (j = 1, 2, …, 6) is the offset number, o_1 is the luminance gradient direction of pixel $\boldsymbol{p}$ after quantization, and o_2 is that of the offset pixel $\boldsymbol{q}$ after quantization. S_i is the cell of present concern, $W_j(\boldsymbol{p})$ is the offset region of pixel $\boldsymbol{p}$ and $K(*)$ is the function whose value is 1, if the argument is true, and 0, otherwise.

To reduce the influence of changes in local brightness and contrast, each 2-D histogram obtained by Eq. (2) is normalized.

2.1.3. Gradient histogram and Ri-HOG feature vector

Since the value of luminance gradient direction θ of the entire circular region in the Cartesian coordinate system ranges from 0° to 360°, it is quantized in 36 directions every 10° intervals. A gradient histogram of the entire circular region is then made and its peak angle is found as the reference direction. The cells are rearranged in the clockwise order from the reference direction.

The cell numbers of the smallest concentric circle are expressed as follows:

$$i_u = 1, 4, 7, \cdots, N \ (u = 1, 2, 3, \cdots, 36) \quad (3)$$

Using Notation (3), the cells are arranged in the order of i_1 $i_1 + 1$, $i_1 + 2$, i_2, $i_2 + 1$, $i_2 + 2$, i_3 …, $i_{36} + 2$, if the peak of the gradient histogram is 0°, and they are connected in the order of i_{36}, $i_{36} + 1$, $i_{36} + 2$, i_1, …, $i_{34} + 2$, i_{35}, $i_{35} + 1$, $i_{35} + 2$, if the peak of the gradient histogram is 350°. Since each of the 108 cells has 6 2-D histograms and the histogram has 64 components, concatenation of all of these components defines an Ri-HOG feature vector whose dimension is 41,472.

2.2. Ri-LBP Features

Rotation-invariant local binary pattern feature is a feature obtained by adding rotational invariance to the LBP [14]. This section describes Ri-LBP features.

2.2.1. Local binarization

An input image is gray-scaled as shown in Figure 1a, and the brightness values of the pixel of interest and its eight adjacent pixels are compared. If the gray value of an adjacent pixel is larger than or equal to that of the center pixel, the adjacent pixel is given 1, otherwise given 0. By aligning the obtained 0 and 1 from the top-left pixel in the direction of the arrow shown in Figure 1b, an 8-digit binary number is derived, which represents the center pixel.

In Ri-LBP, the starting point is changed in turn to acquire 8-digit binary numbers, and the minimum value out of the eight numbers represents the center pixel. For example, in Figure 1, the LBP value is $01000011_b = 67_d$, but the Ri-LBP value is $00001101_b = 13_d$.

2.2.2. Histogram creation and Ri-LBP feature vector

The cell of the Ri-LBP feature is the same as the cell of Ri-HOG feature.

Since LBP is given as an 8-digit binary number, there are 256 values from 0 to 255, but in Ri-LBP, there are 8-digit binary numbers that have a same value by changing the start point. For example, $10000000_b = 128_d$ in LBP is expressed as $00000001_b = 1_d$ when the start point is changed. Hence 10000000_b and 00000001_b are equivalent in Ri-LBP. After all, the total number of Ri-LBP values is 36.

Since there are cases where the number of pixels in one cell is less than 36, the proposed method divides 36 patterns into nine patterns and creates a histogram with each cell. This histogram is normalized. The cells are rearranged based on the reference angle that gives the peak in the brightness gradient direction obtained in Section 2.1.3, and the histograms of all the cells are concatenated. This provides the Ri-LBP feature vector having the dimension of 972, since each of the 108 cells has a histogram of nine components.

After all, an image for training and test in the proposed method is described by the 42,444-D feature vector containing rotation-invariant HOG and LBP features.

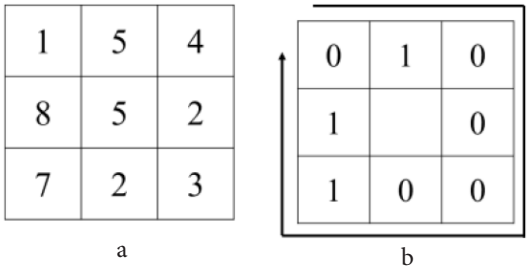

Figure 1 | Example of a local binary pattern: (a) an input gray-scale image, (b) binarization by comparing the brightness of the center pixel with its eight neighbors.

2.3. Detection of a Fallen Person

The proposed method uses Random Forest as a classifier. The advantages of Random Forest include that it is possible to learn efficiently even with high-dimensional features by random learning, and that the influence of the noise contained in learning data can be suppressed efficiently by their random selection.

3. HEAD AND LOWER BODY DETECTION

The head and the lower body of a fallen person are detected using the angle at which the gradient histogram in the circular region obtained in Section 2.1.3 has a peak. When a fallen person exists in a detection window in the orientation shown in Figure 2, the horizontal gradient becomes large and the peak angle of the gradient histogram is 0° or 180°. Therefore, the areas shown by the red frame in Figure 2 are the estimated head position. Let us assume that the posture of a fallen person is straight. Then, if the peak angle in the gradient histogram is denoted by ω_{peak}, the angle ω indicating the orientation of the head is provided by

$$\omega = \omega_{peak} \pm 90° \tag{4}$$

On the other hand, the lower body position is estimated by (i) searching for the area opposite to the detected head on the fallen body, or (ii) searching for the end area of the fallen body if the head is not found.

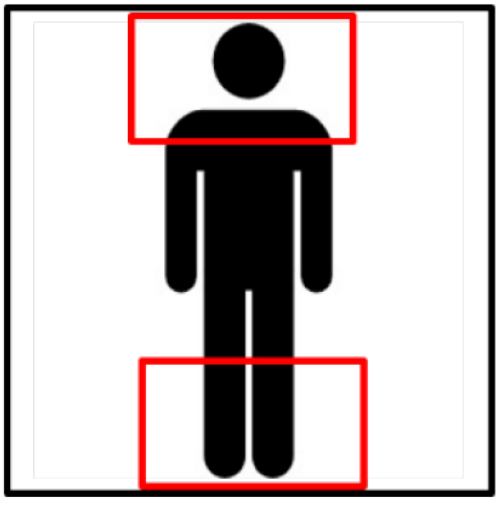

Figure 2 | Example of a presumed head position in a fallen person detection window.

4. EXPERIMENTAL RESULTS

Two experiments are performed. In Experiment 1, a fallen person is detected from a video, whereas the head and the lower body of the person are detected in Experiment 2.

4.1. Experiment 1

In this experiment, 669 fallen person images are collected and used for a positive data class, whereas 1935 images which do not contain persons are used for a negative data class. It is noted that the negative data class contains 1218 negative images chosen from INRIA Person Dataset [15] to get sufficient negative data. Each training image is resized to a 61 × 61 pixels image. A Random Forest classifier is constructed using these training data.

The classifier is applied to eight bird-eye view videos to detect a fallen person. The ground is flat in this particular experiment and the fallen person is assumed to be lying mostly straight on the ground. Each of these videos contains a single fallen person. The processed video frame is resized to 171 (height) × 300 (width) pixels. The size of the search window is 61 × 61 pixels. The judgment on the detection is done with each detection window at each image frame in the video using Intersection over Union (IoU). The threshold of the IoU is experimentally set to 0.6.

As the result, the average recall, the average precision and the average F-value with respect to the eight videos are 0.881, 0.831 and 0.843, respectively. The average computation time is 6.60s/frame. Some results are shown in Figure 3. The green square is the manually set Ground Truth area and the red square shows a detected fallen person.

4.2. Experiment 2

In the second experiment, the head and the lower body are searched and detected on the image of a detected fallen person. A head detector and a lower body detector are designed using Random Forest. For this learning, 879 head images (positive) and 2126 other images (negative) are used with the head, whereas 416 lower body images (positive) and 1858 other images (negative) are used with the lower body.

The designed detectors are applied to nine bird-eye view videos, in three of which a fallen person's head part is occluded and in other three of which the person's lower body is occluded. For the judgment on the detection, IoU is again used. The threshold is set to 0.4 in this particular experiment.

Figure 3 | Results of Experiment 1. The green square is the Ground Truth area, whereas the red square shows the result of the detection.

The experimental results are as follows: The average *F*-value of head detection is 0.539, whereas that of lower body detection is 0.790. The computation time is 7.42 s/frame in average.

Some results of the detection are shown in Figure 4. In Figure 4a, there is no occlusion with the fallen person. On the contrary, the head part of the fallen person is occluded in Figure 4b and its lower body is occluded in Figure 4c. The red square indicates a detected fallen person: The blue square indicates a detected head and the yellow square indicates a detected lower body.

5. DISCUSSION

The proposed method detects a person fallen on the ground from areal images using rotation-invariant image features. Experimental results show effectiveness of the method. The training data still needs improvement, however, in the sense of its amount and content, since a spot of disaster is normally a cluttered environment and a fallen person may have various postures. The evaluation on the performance of the proposed method is expected to be enhanced by learning more amount and varieties of data.

The proposed method also finds the head and the lower body of a detected fallen person. The information on a person's head location is particularly important for further assistance such as communication or providing food.

Another intention of this body part detection is to realize direct detection of a fallen person with certain occlusion. In the performed experiment, a fallen person is assumed to be detected first and then the body parts are found. The method needs to be improved to detect the body parts of a fallen person directly from an aerial image, even if its certain part is occluded.

The total dimension of the used feature vector is large. This may have caused the long computation time of approximately 6–7 s/frame in both experiments. An idea to decrease the dimension is to reduce the number of cells by considering the peak angle of the gradient histogram described in Section 2.1.3. Since the peak angle indicates a person's fallen orientation, some cells perpendicular to the orientation may be discarded.

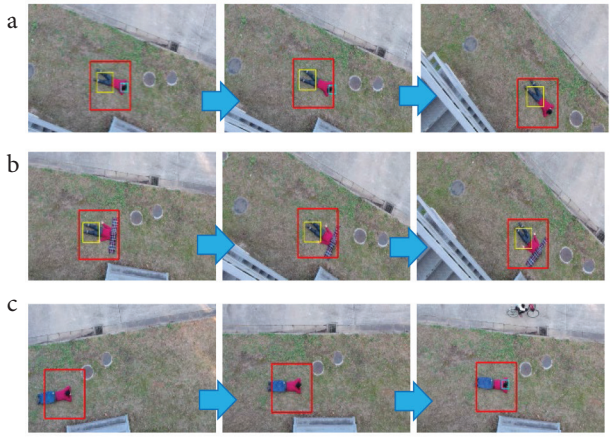

Figure 4 | Results of the head and lower body detection. (a) No occlusion, (b) head part is occluded, (c) lower body is occluded.

6. CONCLUSION

In this paper, we proposed a method of detecting a fallen person and its head and lower body from areal images. The method used rotation-invariant HOG and LBP features to describe a fallen person and Random Forest classifiers were designed using the features. Experimental results showed effectiveness of the method.

Future work includes collecting larger amount of training data to increase the detection rate and improving the method so that it may detect a fallen person with various postures and occlusion. Decreasing the dimension of the employed feature vector is also necessary to reduce the computation time.

CONFLICTS OF INTEREST

The authors declare they have no conflicts of interest.

REFERENCES

[1] White Paper Disaster Management in Japan, saigaipamphlet, Cabinet Office, 2020, p. 3.
[2] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, Kauai, HI, USA, 2001, pp. 511–518.
[3] R. Sato, S. Koshimura, Development of a method for searching survivors in disaster-affected areas with UAV and image analysis, J. Japan Soc. Civil Eng. Ser. B2 (Coastal Engineering) 69 (2013), I_1461–I_1465.
[4] A. Gaszczak, T.P. Breckon, J. Han, Real-time people and vehicle detection from UAV imagery, Proceedings of the SPIE Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques, vol. 7878, IS&T/SPIE Electronic Imaging, San Francisco Airport, CA, USA, 2011, pp. 78780B–78780B–13.
[5] T. Fujimoto, Y. Kameda, I. Matsuda, S. Itoh, Human detection from top view images using rotation invariant HOG features, Proceedings of the Institute of Image Information and Television Engineers, The Institute of Image Information and Television Engineers (ITE), 2014, pp. 17_10_1–17_10_2 (in Japanese).
[6] K. Liu, H. Skibbe, T. Schmidt, T. Blein, K. Palme, T. Brox, et al., Rotation-invariant HOG descriptors using Fourier analysis in polar and spherical coordinates, Int. J. Comput. Vis. 106 (2014), 342–364.
[7] Z. Luo, J. Chen, T. Takiguchi, Y. Ariki, Rotation-invariant histograms of oriented gradients for local patch robust representation, Proceedings of the 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), IEEE, Hong Kong, China, 2015, pp. 196–199.
[8] M. Pietikäinen, T. Ojala, Z. Xu, Rotation-invariant texture classification using feature distributions, Pattern Recognit. 33 (2000), 43–52.
[9] H. Egawa, S. Ishikawa, J.K. Tan, Detection of a fallen person from UAV images using rotation invariant features, Proceedings of the Annual Conference of Biomedical Fuzzy Systems Association, Biomedical Fuzzy Systems Association, 2020, pp. 91–94 (in Japanese).

[10] T. Ojala, M. Pietikäinen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002), 971–987.
[11] L. Breiman, Random forests, Mach. Learn. 45 (2001), 5–32.
[12] G. Yu, J. Yuan, Z. Liu, Unsupervised random forest indexing for fast action search, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Colorado Springs, CO, USA, 2011, pp. 865–872.
[13] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, San Diego, CA, USA, 2005, pp. 886–893.
[14] T. Ojala, M. Pietikäinen, D. Harwood, A comparative study of texture measures with classification based on featured distributions, Pattern Recognit. 29 (1996), 51–59.
[15] INRIA Person Dataset. Accessed January 10, 2019.

AUTHORS INTRODUCTION

Prof. Dr. Joo Kooi Tan

She is currently with Department of Mechanical and Control Engineering, Kyushu Institute of Technology, as Professor. Her current research interests include ego-motion, three-dimensional shape/motion recovery, human detection and its motion analysis from videos. She was awarded SICE Kyushu Branch Young Author's Award in 1999, the AROB Young Author's Award in 2004, Young Author's Award from IPSJ of Kyushu Branch in 2004 and BMFSA Best Paper Award in 2008, 2010, 2013 and 2015. She is a member of IEEE, The Information Processing Society, The Institute of Electronics, Information and Communication Engineers, and The Biomedical Fuzzy Systems Association of Japan.

Ms. Haruka Egawa

She received B.E. and M.E. in Control Engineering from the Graduate School of Engineering, Kyushu Institute of Technology, Japan. Her research includes aerial images processing, human/fallen human detection and motion recognition.