sensors

MDPI

Article

Robust Behavior Recognition in Intelligent Surveillance Environments

Ganbayar Batchuluun, Yeong Gon Kim, Jong Hyun Kim, Hyung Gil Hong and Kang Ryoung Park *

Division of Electronics and Electrical Engineering, Dongguk University, 30 Pildong-ro 1-gil, Jung-gu, Seoul 100-715, Korea; ganabata87@dongguk.edu (G.B.); csokyg@dongguk.edu (Y.G.K.); zzingae@naver.com (J.H.K.); hell@dongguk.edu (H.G.H.)

* Correspondence: parkgr@dongguk.edu; Tel.: +82-10-3111-7022; Fax: +82-2-2277-8735

Academic Editor: Vittorio M. N. Passaro
Received: 16 May 2016; Accepted: 25 June 2016; Published: 30 June 2016

Abstract: Intelligent surveillance systems have been studied by many researchers. These systems should be operated in both daytime and nighttime, but objects are invisible in images captured by visible light camera during the night. Therefore, near infrared (NIR) cameras, thermal cameras (based on medium-wavelength infrared (MWIR), and long-wavelength infrared (LWIR) light) have been considered for usage during the nighttime as an alternative. Due to the usage during both daytime and nighttime, and the limitation of requiring an additional NIR illuminator (which should illuminate a wide area over a great distance) for NIR cameras during the nighttime, a dual system of visible light and thermal cameras is used in our research, and we propose a new behavior recognition in intelligent surveillance environments. Twelve datasets were compiled by collecting data in various environments, and they were used to obtain experimental results. The recognition accuracy of our method was found to be 97.6%, thereby confirming the ability of our method to outperform previous methods.

Keywords: intelligent surveillance system; visible light camera; thermal camera; behavior recognition

1. Introduction

Accurate recognition of behavior is very important and a challenging research topic in intelligent surveillance systems. Although this system should be also operated during the nighttime, objects cannot be visualized in images acquired by a conventional visible light camera. Therefore, a near infrared (NIR) camera or thermal camera (based on medium-wavelength infrared (MWIR) or long-wavelength infrared (LWIR) light) has been considered for usage during the nighttime as an alternative.

The NIR camera usually requires an additional NIR illuminator, which should illuminate a wide area over a great distance during nighttime. However, a thermal camera usually does not need an additional illuminator. In addition, an intelligent surveillance system should be operated during both the daytime and nighttime. Considering all these factors, a dual system of visible light and thermal cameras is used in our research, and we propose a new behavior recognition in intelligent surveillance environments.

A thermal camera image clearly reveals the human body at night and in winter by measuring the human body temperature, which is detectable in the range of MWIR of 3–8 µm and LWIR of 8–15 µm [1–12]. Based on these characteristics, Ghiass et al. [12] introduced various methods of infrared face recognition, also. However, the distinction between the human and the background in the thermal image diminishes when the background temperature is similar to that of the human in some situations during the daytime, which can reduce the consequent recognition accuracy of behavior.

Therefore, the use of both thermal and visible light images enables us to enhance the recognition accuracy of behavior in intelligent surveillance system.

There have been previous studies where the fusion of visible light and thermal cameras is used for various applications [13–16]. Davis et al. [13] proposed the fusion method for human detection from visible light and thermal cameras based on background modeling by statistical methods. Arandjelović et al. [14] proposed the face recognition method by combining the similarity scores from visible light and thermal camera-based recognition. Leykin et al. [15] proposed the method for tracking pedestrians based on the combined information from visible light and thermal cameras. In addition, Kong et al. [16] proposed the face recognition method robust to illumination variation based on the multiscale fusion of visible light and thermal camera images.

Although there have been previous studies combining visible light and thermal cameras in various applications, we focus on the behavior recognition in this research. Previous research on behavior recognition can be categorized into two groups: single camera-based and multiple camera-based approaches. Rahman et al. [17–20] proposed the single camera-based approaches which used negative space analysis based on a background region inside the detected box of the human body. However, these approaches deliver limited performance in cases in which there is a large difference between the detected box and the box that was defined based on the actual human area. Fating and Ghotkar [21] used the histogram of the chain code descriptor to analyze the boundary of the detected area. Although the histogram represents the efficient features, it is computationally expensive to obtain the chain codes from all the shapes of the detected area. A Fourier descriptor was used to extract the scale and rotation invariant features for the detected foreground area [22–25]. This method has the advantage of correctly representing the foreground shape, but it has difficulty in discriminating behavior with a similar foreground shape, such as walking and kicking. Sun et al. [26] proposed the method using a local descriptor based on scale invariant feature transform (SIFT) and holistic features by Zernike moments for human action recognition. However, their method has the disadvantage of requiring much processing time to extract the features by using both the SIFT and Zernike moments. Schüldt and Laptev et al. [27–29] performed behavior recognition based on the local spatiotemporal features, although a clear background is required to guarantee highly accurate results. In addition, corner detection-based feature extraction can cause the false detections of corners, which can degrade the accuracy of behavior recognition. Wang et al. and Arandjelović [30,31] proposed the concept of motionlets, a motion saliency method, which achieves high performance in terms of human motion recognition provided the foreground is clearly segmented from the background. Optical flow-based methods were used to represent motion clearly using a feature histogram. However, the process of obtaining optical flow is computationally expensive, and it is also sensitive to noise and illumination changes [27,28,32]. Various groups [32–37] have performed behavior recognition based on a gait flow image (GFI), gait energy image (GEI), gait history image (GHI), accumulated motion image (AMI), and motion history image (MHI). These studies focused on obtaining a rough foreground region in which motion occurs, but the attempt to determine the exact path of motion of hands and legs was less satisfactory. Wong et al. [38] proposed a method of faint detection based on conducting a statistical search of the human's height and width using the images obtained by a thermal camera. Youssef [39] proposed a method based on the convexity defect feature point for human action recognition. This method finds the tip points of hand, leg, and head. However, incorrect points can be detected as a result of an increase in the number of points for large-sized humans surrounded by noise. In addition, this method cannot discriminate the tip point of a hand from that of a leg in cases in which the leg is positioned at a height similar to that of the hand when kicking motion happens. Besides, it is computationally expensive because of the need to calculate the contour, polygon, convex hull, and convexity defect in every frame.

All this research based on images acquired with a single visible camera has the limitation of performance enhancement when a suitable foreground is difficult to be obtained from the image in situations containing extensive shadows, illumination variations, and darkness. In addition, previous research based on thermal cameras was performed in constrained environments such as indoors or without considering situations in which the background temperature is similar to the foreground in hot weather.

Therefore, multiple camera-based approaches were subsequently considered as an alternative. Zhang et al. [40] proposed a method of ethnicity classification based on gait using the fusion of information of GEI with multi-linear principal component analysis (MPCA) obtained from multiple visible light cameras. Rusu et al. [41] proposed a method for human action recognition by using five visible light cameras in combination with a thermal camera, whereas Kim et al. [42] used a visible light camera with a thermal camera. However, all these studies were conducted in indoor environments without considering temperature and light variations associated with outdoor environments.

We aim to overcome these problems experienced by previous researchers, by proposing a new robust system for recognizing behavior in outdoor environments including various temperature and illumination variations. Compared to previous studies, our research is novel in the following three ways:

- Contrary to previous research where the camera is installed at the height of a human, we research behavior recognition in environments in which surveillance cameras are typically used, i.e., where the camera is installed at a height much higher than that of a human. In this case, behavior recognition usually becomes more difficult because the camera looks down on humans.
- We overcome the disadvantages of previous solutions of behavior recognition based on GEI and the region-based method by proposing the projection-based distance (PbD) method based on a binarized human blob.
- Based on the PbD method, we detect the tip positions of the hand and leg of the human blob. Then, various types of behavior are recognized based on the tracking information of these tip positions and our decision rules.

The remainder of our paper is structured as follows. In Section 2, we describe the proposed system and the method used for behavior recognition. The experimental results with analyses are presented in Section 3. Finally, the paper is concluded in Section 4.

2. Proposed System and Method for Behavior Recognition

2.1. System Setup and Human Detection

In this section, we present a brief introduction of our human detection system because the behavior recognition is carried out based on the detection results.

Figure 1 shows the examples of our system setup and human detection method. Detailed explanations of the system and our human detection method can be found in our previous paper [43]. We have implemented dual camera systems (including an FLIR thermal camera capturing the images of 640 × 480 pixels [44] and a visible light camera capturing the images of 800 × 600 pixels), where the two axes of the visible light and thermal cameras are oriented parallel to the horizontal direction as shown in the upper left image of Figure 1. Therefore, the visible light and thermal images are simultaneously captured with minimum disparity. We prevented rain from entering the thermal and visible light cameras by attaching a glass cover (germanium glass, which is transparent to MWIR and LWIR light [45] and conventional transparent glass for the thermal and visible light cameras, respectively) to the front of each camera, as shown in the upper left image of Figure 1. Our dual camera system is set at a height of 5–10 m above the ground, as shown in Figure 1. In addition, we tested it at five different heights (see details in Section 3.1) to measure the performance of our system in various environments.

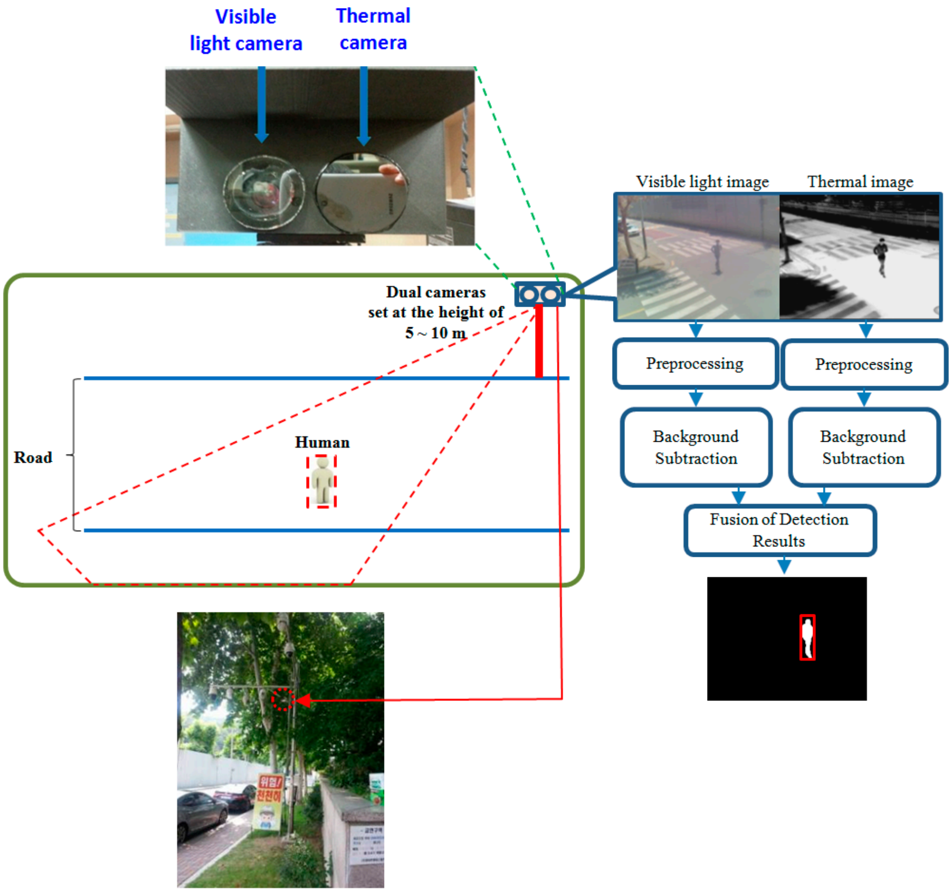


Figure 1. Examples of system setup with brief explanations of our human detection method.

The flow diagram on the right-hand side of Figure 1 summarizes our human detection algorithm. Through background subtraction after preprocessing to remove noise, two human regions are detected from the visible light and thermal images. Then, the two regions are combined based on a geometric transform matrix, which is obtained by the calibration procedure of both cameras in advance. Finally, the correct human area is detected as shown on the right of Figure 1 after post-processing for noise reduction, morphological operation, and size filtering [43,46]. Our detection method can be applied to images including multiple humans. The use of dual camera systems enables our method to robustly detect human areas in the images of various environments such as those with severe shadow, illumination variations, darkness, and cases in which the background temperature is similar to human area in hot weather.

Most previous research on behavior recognition was performed with images captured by a camera at low height. That is, the camera was set at a height of about 2 m above the ground and can capture frontal-viewing human body parts as shown in Figure 2a. This arrangement allowed the motions of human hands and feet to be observed more distinctively. However, our camera is set at a height of 5–10 m, which is the normal height of the camera of a surveillance system. This causes the human area in the image to shrink in the vertical direction as shown in Figure 2b. In addition, our camera arrangement could also reduce the motions of human hands and feet in the vertical direction, change the ratio of height to width of the human box, and reduce the vertical image resolution of the human box in the image, all of which can complicate the correct recognition of behavior. However, our research needs to consider that our system is intended for use in a conventional surveillance system; therefore, the camera setup shown in Figure 2b is used in our research, and we extract the tip points of the hand and leg to track them for accurate behavior recognition rather than using the entire region covering the detected human body.

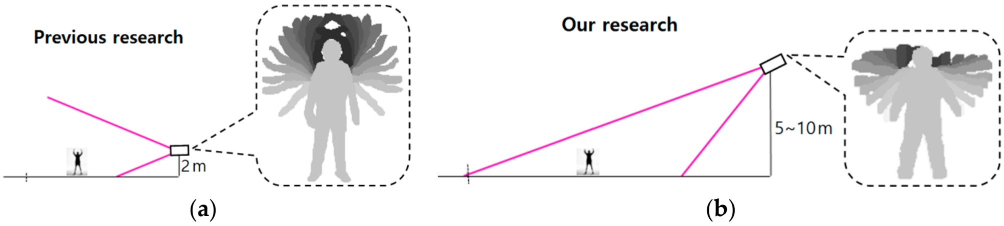


Figure 2. Comparison of different camera setup used in (**a**) previous research; (**b**) our research.

2.2. Proposed Method of Behavior Recognition

2.2.1. Overall Procedure of Behavior Recognition

The method proposed for behavior recognition is presented in Figure 3. Detailed explanations of Rules 1–4 are provided in Section 2.2.2. We adopted an IF-THEN rule-based method to recognize behavior. As shown in Figure 3, behaviors are roughly divided into two types according to the horizontal activeness measured by the change in the width of detected human box. Our research aims to recognize the following 11 types of behavior:

- Waving with two hands
- Waving with one hand
- Punching
- Kicking
- Lying down
- Walking (horizontally, vertically, or diagonally)
- Running (horizontally, vertically, or diagonally)
- Standing
- Sitting
- Leaving
- Approaching

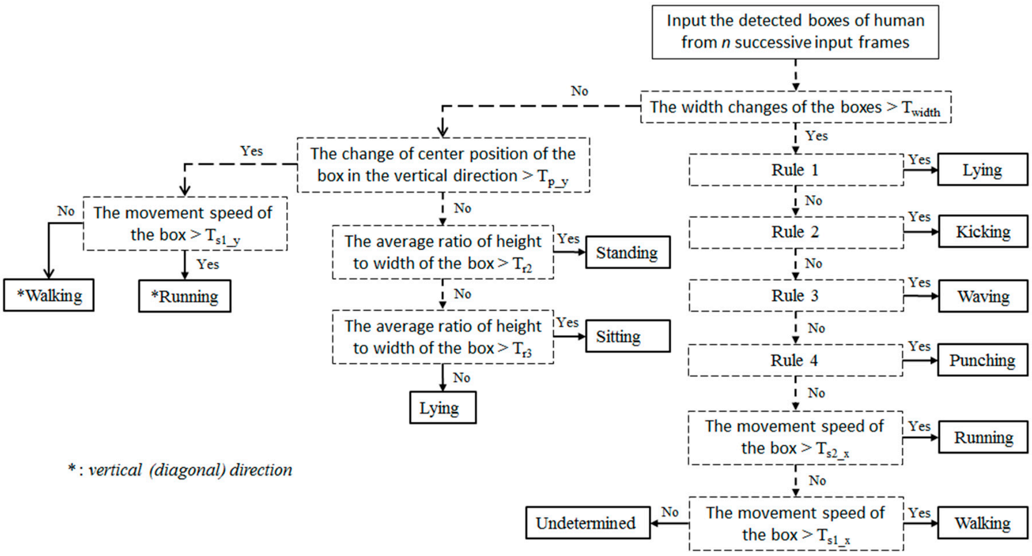

Input the detected boxes of human from *n* successive input frames
No
The width changes of the boxes > T_{width}
Yes
The change of center position of the box in the vertical direction > T_{p_y}
Yes
The movement speed of the box > T_{s1_y}
No
Yes
*Walking
*Running
No
The average ratio of height to width of the box > T_{r2}
Yes
Standing
No
The average ratio of height to width of the box > T_{r3}
Yes
Sitting
No
Lying
*: vertical (diagonal) direction
Rule 1
Yes
Lying
No
Rule 2
Yes
Kicking
No
Rule 3
Yes
Waving
No
Rule 4
Yes
Punching
No
The movement speed of the box > T_{s2_x}
Yes
Running
No
The movement speed of the box > T_{s1_x}
Yes
Walking
No
Undetermined

Figure 3. Flowchart of the proposed method of behavior recognition.

Among the 11 types of behavior, leaving and approaching are recognized simply based on the change in the distance between the two center positions of the detected boxes of two people. If the distance is reduced, our system determines it as approaching, whereas if it increases, our method determines it as leaving.

Behavior such as hand waving, punching, and kicking has very complicated motion patterns. For example, waving can either be full waving or half waving. We assume the directions of 3, 6, 9, and 12 o'clock to be 0°, 270°, 180°, and 90°, respectively.

- Full waving, where the arm starts from about 270° and moves up to 90° in a circle as shown in Figure 4a,c.
- Half waving, where the arm starts from about 0° and moves up to 90° in a circle as shown in Figure 4b,d.
- Low punching, in which the hand moves directly toward the lower positions of the target's chest shown in Figure 4e.
- Middle punching, where the hand moves directly toward the target's chest as shown in Figure 4f.
- High punching, where the hand moves directly toward the upper positions of the target's chest as shown in Figure 4g.
- Similarly, in low, middle, and high kicking, the leg moves to different heights as shown in Figure 4h–j, respectively.

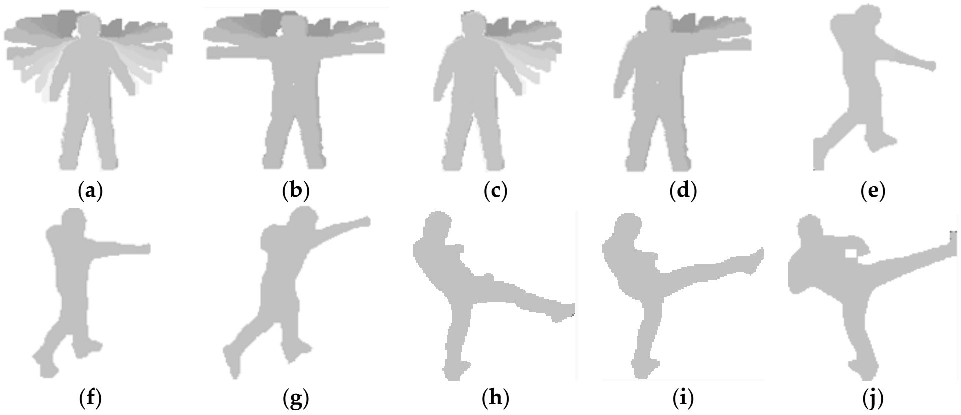


Figure 4. Comparison of motion patterns of each type of behavior. Waving: (**a**) full, by using two hands; (**b**) half, by using two hands; (**c**) full, by using one hand; (**d**) half, by using one hand. Punching: (**e**) low; (**f**) middle; (**g**) high. Kicking: (**h**) low; (**i**) middle; (**j**) high.

In addition, according to martial arts, kicking and punching can be shown in many different ways.

In our research, the 11 types of behavior are categorized into three classes, with each class representing behavior with different intention and meaning. This requires us to extract the different features according to each class.

Class 1: **Walking**, **running**, **sitting**, and **standing** are regular types of behavior with less detailed gestures of the hands and feet than the behavior in Class 2. Walking and running behavior involves active motion, whereas sitting and standing behavior is motionless. In addition, the rough shape of the human body when standing is different from that when sitting. Therefore, the motion and rough shape of the human body are measured using the speed and ratio of height to width of the detected human box, respectively. That is, the speed and ratio are key features of the spatial and temporal information of the detected box that enable recognition of the regular types of behavior of Class 1.

Class 2: **Kicking**, **punching**, **lying down**, **waving with two hands**, and **waving with one hand** are types of behavior consisting of special gestures for hands or feet, motion, and shape except for lying down. In addition, they are characterized by very active motions of the hands or feet. Some of the behavior produces similar shapes in an image, such as waving with one hand and punching. Therefore, we need to be able to track feature points efficiently, which we achieve by using the proposed PbD method to find the tip points of legs and hands as feature points. We track the path of these tip points to recognize the behavior in this class. However, lying down can be recognized in the same way as the behavior in Class 1, because it is motionless and the ratio of height to width of the human box is different from that of other behaviors.

Class 3: **Approaching** and **leaving** are interactional types of behavior, both of which are recognized by measuring the change in the distance between the boxes of two persons.

2.2.2. Recognition of Behavior in Classes 1 and 3

In this section, we explain the proposed method in terms of the recognition of behavior in Classes 1 and 3. Behavior in Classes 1 and 3 is recognized by analyzing and comparing the detected human box in the current frame with the previous nine boxes from the previous nine frames as shown in Figure 5.

As shown in Equation (1), the sum of change in the width of detected box is calculated from 10 frames (current and previous nine frames):

$$W_v = \sum_{i=0}^{N-1} |W_B\,(t-(i+1)) - W_B\,(t-i)| \tag{1}$$

where N is 9 and $W_B\,(t)$ is the width of the detected human box at the t-th frame as shown in Figure 5. Based on W_v compared to the threshold ($\mathrm{T_{width}}$), the first classification for behavior recognition is conducted as shown in Figure 3.

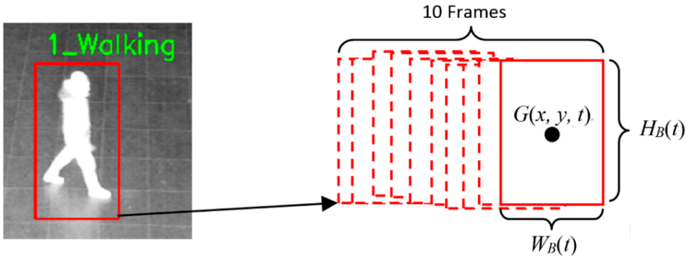


Figure 5. Example showing the parameters used to analyze the detected human box by comparing the box in the current frame with those in the previous nine frames.

If W_v is less than the threshold ($\mathrm{T_{width}}$), the sum of change in the center of the detected box in the vertical direction is calculated from 10 frames (current and previous nine frames). Based on the sum compared to the threshold ($\mathrm{T_{p\_y}}$), walking and running behaviors in the vertical or diagonal direction are distinguished from those of standing, sitting, and lying down as shown in the left-hand part of Figure 3.

The distance between the center positions ($G(x, y, t)$ of Figure 5) of two detected boxes is measured by comparing two successive frames among the 10 frames (current and previous nine frames). Then, the consequent movement velocity of the center position can be calculated as the movement speed of the detected box because the time difference between two successive frames can be obtained by our system. Based on the average speed (considering the movement direction) calculated from 10 frames (current and previous nine frames), walking and running are recognized. If the speed exceeds the specified threshold, our system determines the behavior as running, otherwise the behavior is determined as walking, as shown in Figure 3.

In general, the distance between two center positions in the image in the horizontal direction is different from that in the vertical or diagonal directions even with the same distance in 3D space. That is because the camera setup is tilted as shown in Figure 2b. Therefore, a different threshold ($\mathrm{T_{s1\_y}}$) is used for recognizing walking and running in the vertical or diagonal directions compared to the thresholds ($\mathrm{T_{s1\_x}}$ and $\mathrm{T_{s2\_x}}$) for the horizontal direction, as shown in Figure 3.

The ratio of the height to width of the detected box is calculated by Equation (2), and based on the average R from 10 frames (current and previous nine frames), standing, sitting, or lying down are recognized. As shown in Figure 3, if the average R is larger than the threshold (T_{r2}), our system determines the behavior as standing. If the average R is less than threshold (T_{r2}), but larger than threshold (T_{r3}), the behavior is recognized as sitting. If the average R is less than both threshold (T_{r2}) and threshold (T_{r3}), our system determines the behavior as lying down.

$$R = \frac{H_B(t)}{W_B(t)} \tag{2}$$

where $W_B(t)$ and $H_B(t)$ represent the width and height, respectively, of the detected human box at t-th frame as shown in Figure 5.

Approaching and leaving behavior is recognized by measuring the change in the distance between the center positions of the boxes of two persons, as shown in Equation (3):

$$D_v = \sum_{i=0}^{N-1} (b(t-(i+1)) - b(t-i)) \tag{3}$$

where N is 9 and $b(t)$ is the Euclidean distance between the two center positions of the boxes of two persons at the t-th frame. If D_v is larger than the threshold, the behavior is determined as approaching, whereas the behavior is determined as leaving if D_v is smaller than the threshold.

The optimal thresholds in the flowchart of Figure 3 were experimentally determined to minimize the error rate of behavior recognition. To ensure fairness, the images used for the determination of the thresholds were excluded from the experiments that were conducted to measure the performance of our system.

2.2.3. Recognition of Behavior in Class 2

We developed a new method, named the PbD method, to extract the features of behavior in Class 2. Firstly, X_G is calculated as the geometric center position of the binarized image of the human area as shown in Figure 6a. Then, the distances (d_r (or d_l)) between the X positions of X_G and the right-most (or left-most) white pixel of the human area are calculated at each Y position of the detected human box. The processing time was reduced by calculating the distances at each Y position for every three pixels. These distances enable us to obtain the two profile graphs (D_{left} and D_{right}) including the leg, body, arm, and head of human area as shown in Figure 6b,c. Because the profile graph is obtained based on the projection of the distance in the horizontal direction, we named this method the PbD method.

$$D_{left} = \{d_{l1}, d_{l2} \ldots d_{ln}\},\ D_{right} = \{d_{r1}, d_{r2} \ldots d_{rn}\} \tag{4}$$

where n is the number of distance.

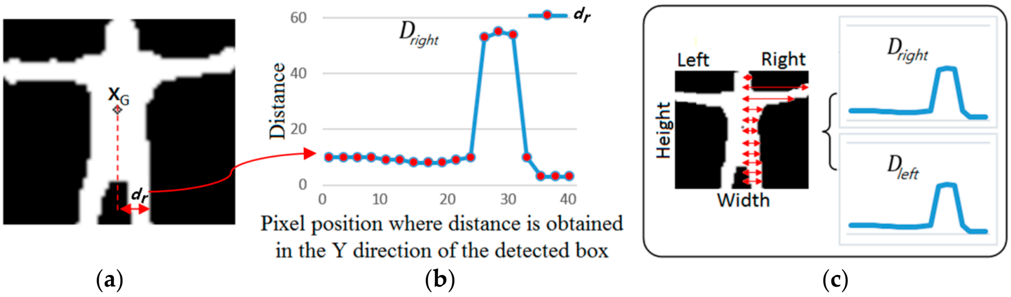


Figure 6. Example of profile graphs produced by the proposed PbD method. (**a**) Binarized image of human area in the detected box; (**b**) profile graph representing the right part of the human area; (**c**) two profile graphs representing the right and left parts of the human area, respectively.

The profile graph is used to determine the position whose distance (Y-axis value) is maximized as the tip position (M in Figure 7) of the hand or leg. The tip point is determined to either belong to the hand or the leg by obtaining the L_1 and L_2 lines from the first and last points of the profile graph with point M as shown in Figure 7.

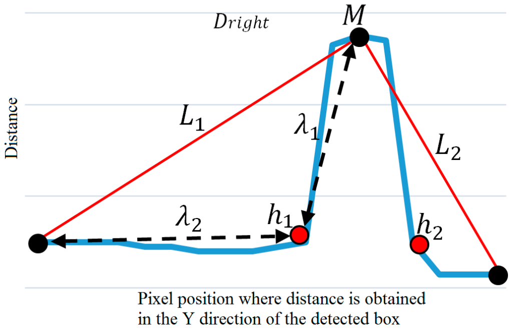
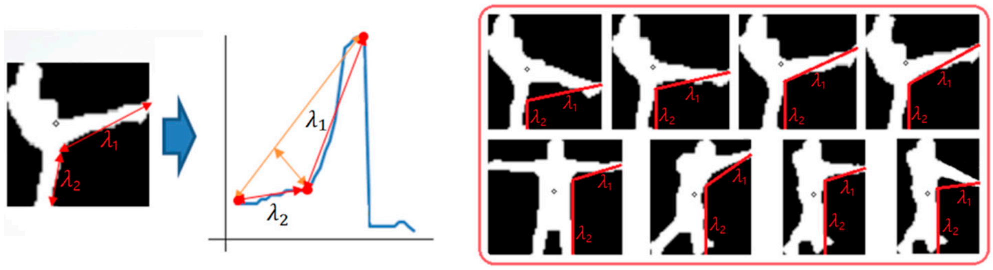


Figure 7. Example illustrating the extraction of features from the profile graph obtained by the proposed PbD method.

Then, the two points on the graph (whose distances from the L_1 and L_2 lines are maximized, respectively) are detected as the starting and ending points of the human arm, respectively, as shown in h_1 and h_2 of Figure 7. In case the distance from the L_2 line is smaller than the threshold, the ending point of the profile graph is determined as h_2. In addition, if the distance from the L_1 line is smaller than the threshold, the starting point of the profile graph is determined as h_1.

Based on the starting point, h_1, and the M points, the two distances λ_1 and λ_2 can be calculated. Because the human arm is shorter than the distance between the starting position of the arm (h_1 in Figure 7) and the foot (the starting position of the profile graph in Figure 7), λ_1 inevitably becomes smaller than λ_2 when point M is the tip position of the hand. On the other hand, because the human leg is longer than the distance between the starting position of the leg and foot, λ_1 inevitably becomes larger than λ_2 when point M is the tip position of the leg (foot) as shown in Figure 8b. Based on these observations, we can discriminate between the position of the hand tip and that of the leg tip based on the ratio of λ_1 and λ_2 as shown in Equation (5):

$$\begin{cases} M \text{ belongs to hand tip, } if\ \frac{\lambda_2}{\lambda_1} > 1 \\ M \text{ belongs to leg tip, } else \end{cases} \quad (5)$$



(**a**) (**b**) (**c**)

Figure 8. Examples of the two features λ_1 and λ_2 for kicking, and the comparison of these two features for kicking and punching. (**a**) Case of kicking; (**b**) profile graph of (a) by the proposed PbD method; (**c**) comparison of λ_1 and λ_2 for kicking and punching.

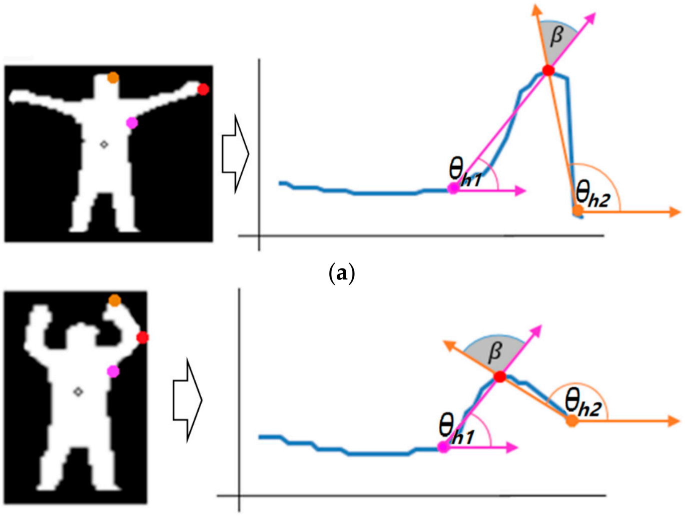
As shown in Figure 9, hand waving behavior can happen by unfolding (Figure 9a) or bending (Figure 9b) the user's arms. In the latter case (Figure 9b), point M can be the tip position of the elbow instead of that of the hand. We proposed the method expressed by Equations (6) and (7) to discriminate the former case (point M is the tip position of the hand) from the latter (point M is the tip position of the elbow).

As shown in Figure 9, the two angles θ_{h1} and θ_{h2} are calculated based on three points of the starting point of the arm, point M, and the ending point of the profile graph. Then, the difference between these two angles is calculated as β. In general, β is smaller when hand waving occurs by unfolding than when hand waving occurs by bending of the user's arm as shown in Figure 9. Therefore, based on β, the former case (point M is the tip position of the hand) is discriminated from the latter case (point M is the tip position of the elbow) as follows:

$$\begin{cases} M \text{ belongs to hand tip, } if\ \beta < Th_\theta \\ M \text{ belongs to elbow tip, } else \end{cases} \tag{6}$$

$$\beta = \text{abs}\left(\frac{\pi - arc\sin\left(\frac{Y_M - Y_{h_2}}{\sqrt{\left(X_M - X_{h_2}\right)^2 + \left(Y_M - Y_{h_2}\right)^2}}\right) - arc\sin\left(\frac{\left(Y_M - Y_{h_1}\right)}{\sqrt{\left(X_M - X_{h_1}\right)^2 + \left(Y_M - Y_{h_1}\right)^2}}\right)}{\pi/180} \right) \tag{7}$$

where (X_M, Y_M) is the coordinate of point M in Figure 7, and (X_{h_1}, Y_{h_1}) is the position of h_1. In addition, (X_{h_2}, Y_{h_2}) is that of h_2 in Figure 7. *abs* is the function for obtaining absolute value (non-negative value). Based on the extracted feature points of Figures 6–9, the types of behavior in Class 2 are recognized based on the decision rules of Table 1. Rules 1–4 of Figure 3 are shown in Table 1.



(**a**)

(**b**)

Figure 9. Examples of two types of hand waving by (**a**) unfolding; and (**b**) bending of the user's arm.

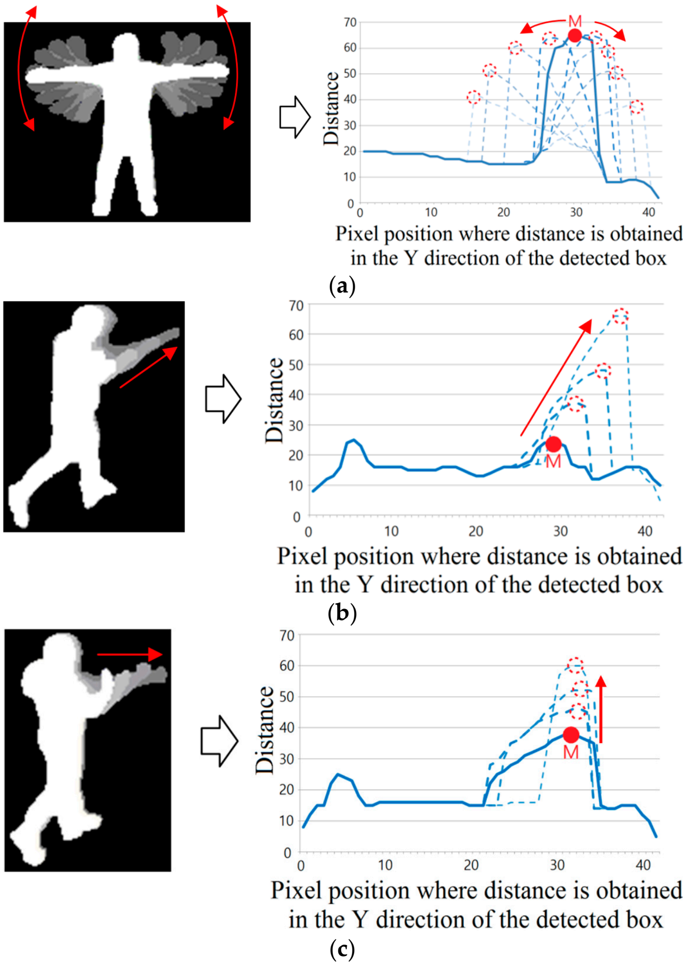
Table 1. Decision rules for recognizing the types of behavior in Class 2.

Decision rules	
Rule 1 (Lying down)	*If* ("*Condition 1* * is TRUE") : *Lying down* *Else* : *Go to Rule 2*
Rule 2 (Kicking)	*If* ("M is leg tip" *and* "*Condition 2* * is TRUE") : *Kicking* *Else* : *Go to Rule 3*

Table 1. *Cont.*

Decision rules	
Rule 3 (Hand waving)	*If* ("M is hand tip" *and* "*Condition 3* * is TRUE") : *Hand waving* *Else* : *Go to Rule 4*
Rule 4 (Punching)	*If* ("M is hand tip" *and* ("*Condition 4* * *or Condition 5* * is TRUE")) : *Punching* *Else* : *Go to the rule* (for checking running, walking or undetermined case as shown in the right lower part of Figure 3)

* ***Condition 1***: Average *R* of Equation (2) is less than two thresholds (T_{r2} and T_{r3}) as shown in Figure 3; * ***Condition 2***: both X and Y positions of M point are higher than threshold; * ***Condition 3***: The position of M is changed in both X and Y directions in recent N frames as shown in Figure 10a; * ***Condition 4***: The position of M is increased in both X and Y directions in recent N frames as shown in Figure 10b; * ***Condition 5***: The position of M is increased only in the Y directions in recent N frames as shown in Figure 10c.



(**a**)

(**b**)

(**c**)

Figure 10. Examples of changes in the position of point M in the profile graph for hand waving and punching. (**a**) Hand waving (the figure on the right shows one profile graph among the two graphs representing the waving of both left and right hands); (**b**) punching (case 1); (**c**) punching (case 2).

The proposed PbD method enables us to rapidly find the correct positions of the tips of the hand and leg without using a complicated algorithm that performs boundary tracking of binarized human area.

3. Experimental Results

3.1. Description of Database

Open databases exist for behavior recognition of visible light images [27,47,48] or that of thermal images [49]. However, there is no open database (for behavior recognition in outdoor environments) with images collected by both visible light and thermal cameras at the same time. Therefore, we used the database that was built by acquiring images with our dual camera system. Data acquisition for the experiments was performed by using a laptop computer and the dual cameras shown in Figure 1. All the images were acquired based on the simultaneous use of visible light and thermal cameras. The laptop computer was equipped with a 2.50 GHz CPU (Intel (R) Core (TM) i5-2520M, Intel Corp., Santa Clara, CA, USA) and 4 GB RAM. The proposed algorithm was implemented using a C++ program using Microsoft foundation class (MFC, Microsoft Corp., Redmond, DC, USA) and OpenCV library (version 2.3.1, Intel Corp., Santa Clara, CA, USA).

We built large datasets that include 11 different types of behavior (explained in Section 2.2.1) as shown in Table 2. Datasets were collected in six different places during the day and at night in environments with various temperatures of four seasons with the different camera setup positions. The database includes both thermal and visible images. The total number of images in the database is 194,094. The size of the human area varies from 28 to 117 pixels in width and from 50 to 265 pixels in height, respectively.

Table 2. Description of 12 datasets.

Dataset	Condition	Detail Description
I (see in Figure 11a)	1.2 °C, morning, humidity 73.0%, wind 1.6 m/s	- The intensity of background is influenced by the window of building - Human shadow is reflected on the window, which is detected as another object in thermal image
II (see in Figure 11b)	−1.0 °C, evening, humidity 73.0%, wind 1.5 m/s	- The intensity of background is influenced by the window of building - Object is not seen in visible light image - Human shadow is reflected on window, which is detected as another object in thermal image
III (see in Figure 11c)	1.0 °C, afternoon, cloudy, humidity 50.6%, wind 1.7 m/s	- The intensity of background is influenced by leaves and trees
IV (see in Figure 11d)	−2.0 °C, dark night, humidity 50.6%, wind 1.8 m/s	- The intensity of background is influenced by leaves and trees - Object is not seen in visible light image
V (see in Figure 11e)	14.0 °C, afternoon, sunny, humidity 43.4%, wind 3.1 m/s	- Difference between background and human diminishes because of the high temperature of background
VI (see in Figure 11f)	5.0 °C, dark night, humidity 43.4%, wind 3.1 m/s	- The air heating system of building increases the temperature of part of the building in background - Object is not seen in visible light image
VII (see in Figure 11g)	−6.0 °C, afternoon, cloudy, humidity 39.6%, wind 1.9 m/s	- Halo effect is shown near the human area in thermal image, which makes it difficult to detect the correct human area
VIII (see in Figure 11h)	−10.0 °C, dark night, humidity 39.6%, wind 1.7 m/s	- Halo effect is shown near the human area in thermal image, which makes it difficult to detect the correct human area - Object is not seen in visible light image
IX (see in Figure 11i)	21.9 °C, afternoon, cloudy, humidity 62.6%, wind 1.3 m/s	- Halo effect is shown near the human area in thermal image, which makes it difficult to detect the correct human area - Difference between background and human diminishes due to the high temperature of background
X (see in Figure 11j)	−10.9 °C, dark night, humidity 48.3%, wind 2.0 m/s	- The dataset was collected at night during winter. Therefore, the background in thermal image is too dark because of low temperature - Object is not seen in visible light image

Table 2. *Cont*.

Dataset	Condition	Detail Description
XI (see in Figure 11k)	27.0 °C, afternoon, sunny, humidity 60.0%, wind 1.0 m/s	- Human is darker than road because the temperature of the road is much higher than that of a human in summer - Leg is not clear when kicking behavior happens because the woman in the image wore a long skirt
XII (see in Figure 11l)	20.2 °C, dark night, humidity 58.6%, wind 1.2 m/s	- Human is darker than road because the temperature of the road is much higher than that of a human in summer - Object is not seen in visible light image

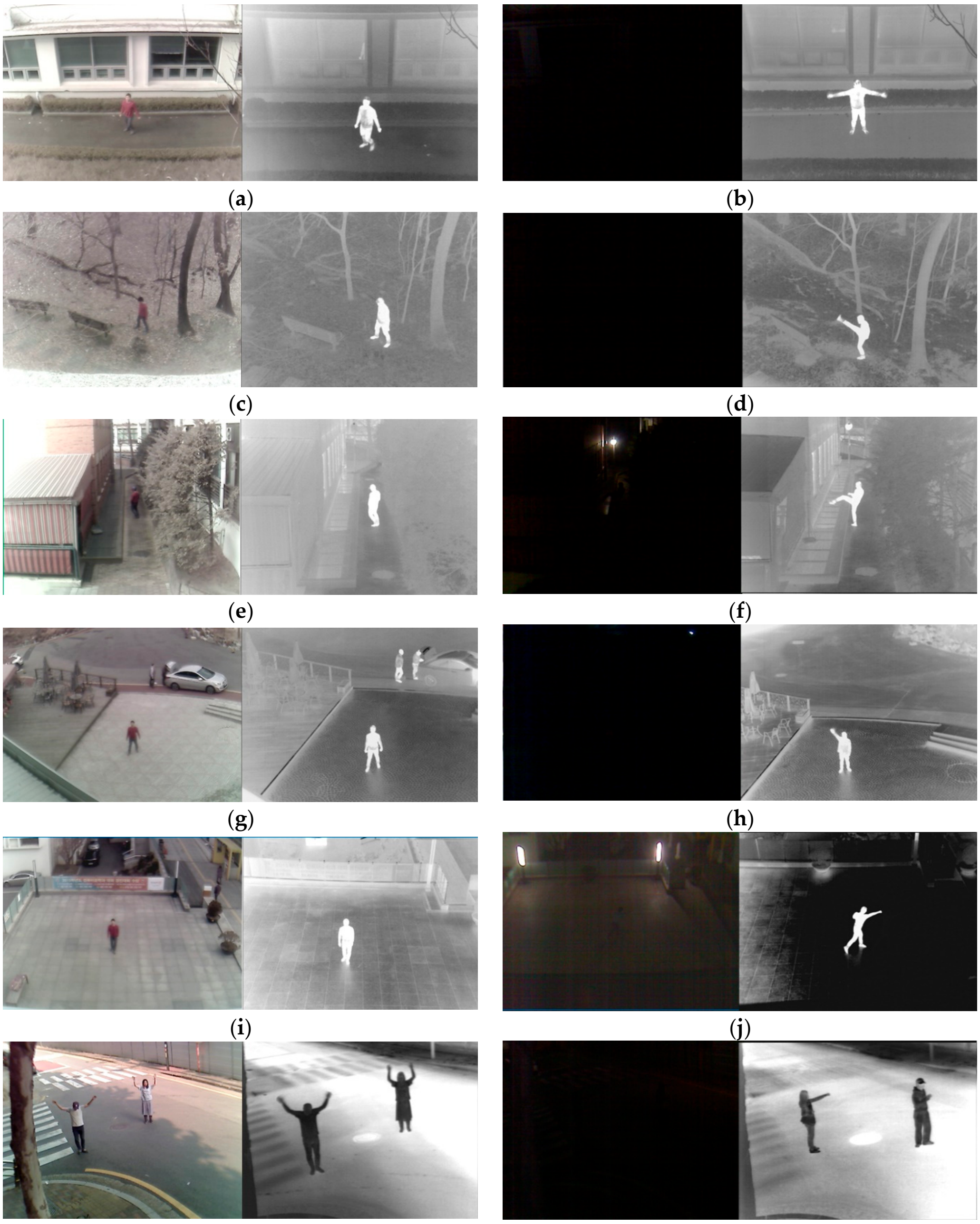

(**a**)
(**b**)
(**c**)
(**d**)
(**e**)
(**f**)
(**g**)
(**h**)
(**i**)
(**j**)
(**k**)
(**l**)

Figure 11. Examples from each of the 11 datasets. (**a**) Dataset I; (**b**) dataset II; (**c**) dataset III; (**d**) dataset IV; (**e**) dataset V; (**f**) dataset VI; (**g**) dataset VII; (**h**) dataset VIII; (**i**) dataset IX; (**j**) dataset X; (**k**) dataset XI; (**l**) dataset XII.

As shown in Figure 12 and Table 3, we collected the 11 datasets with a camera setup of various heights, horizontal distances, and Z distances. We then classified the datasets according to the kind of behavior, again, and the numbers of frames and types of behavior are shown in Table 4.

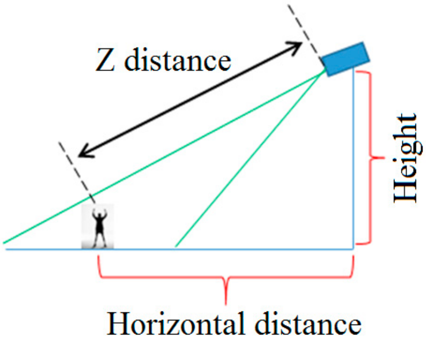


Figure 12. Example of camera setup.

Table 3. Camera setup used to collect the 11 datasets (unit: meters).

Datasets	Height	Horizontal Distance	Z Distance
Datasets I and II	8	10	12.8
Datasets III and IV	7.7	11	13.4
Datasets V and VI	5	15	15.8
Datasets VII and VIII	10	15	18
Datasets IX and X	10	15	18
Datasets XI and XII	6	11	12.5

Table 4. Numbers of frames and the types of behavior in each dataset.

	#Frame		#Behavior	
Behavior	**Day**	**Night**	**Day**	**Night**
Walking	1504	2378	763	1245
Running	608	2196	269	355
Standing	604	812	584	792
Sitting	418	488	378	468
Approaching	1072	1032	356	354
Leaving	508	558	163	188
Waving with two hands	29588	14090	1752	870
Waving with one hand	24426	15428	1209	885
Punching	21704	13438	1739	1078
Lying down	7728	5488	2621	2022
Kicking	27652	22374	2942	3018
Total	194094		24051	

3.2. *Accuracies of Behavior Recognition*

We used the datasets in Table 4 to evaluate the accuracy of behavior recognition by our method. The accuracies were measured based on the following equations [50]:

$$\text{Positive predictive value (PPV)} = \frac{\#\text{TP}}{\#\text{TP} + \#\text{FP}} \tag{8}$$

$$\text{True positive rate (TPR)} = \frac{\#\text{TP}}{\#\text{TP} + \#\text{FN}} \tag{9}$$

$$\text{Accuracy (ACC)} = \frac{\#\text{TP} + \#\text{TN}}{\#\text{TP} + \#\text{TN} + \#\text{FP} + \#\text{FN}} \tag{10}$$

$$\text{F\_score} = 2 \cdot \frac{\text{PPV} \cdot \text{TPR}}{\text{PPV} + \text{TPR}} \tag{11}$$

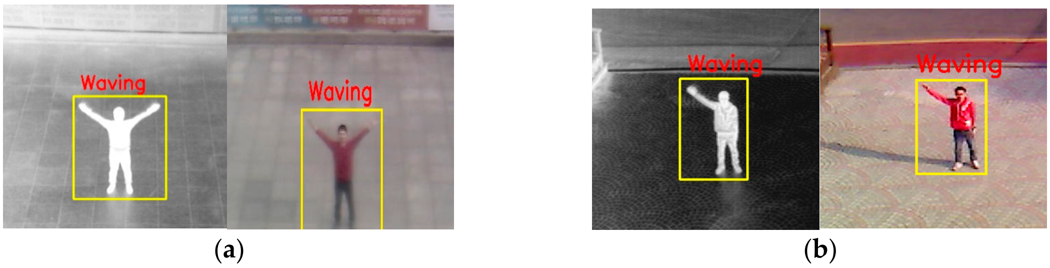
where #TP is the number of true positives (TPs), and TP represents the cases in which the behavior included in input image is correctly recognized. #TN is the number of true negatives (TNs), and TN represents the cases in which the behavior not included in input image is correctly unrecognized. In addition, #FP is the number of false positives (FPs), and FP represents the case in which the behavior not included in input image is incorrectly recognized. #FN is the number of false negatives (FNs), and FN represents the case in which the behavior included in input image is incorrectly unrecognized. For all the measures of PPV, TPR, ACC, and F_score, the maximum and minimum values are 100(%) and 0(%), respectively, where 100(%) and 0(%) are the highest and lowest accuracies, respectively.

As shown in Table 5, the accuracies of behavior recognition by our method are higher than 95% in most cases, irrespective of day and night and kinds of behaviors. In Figure 13, we show examples of correct behavior recognition with our datasets collected in various environments. The results of behavior recognition can show various information of emergency situation. For example, in Figure 13l, one human (object 2) is leaving the other human (object 1), who is lying on the ground, in which case an emergency situation would be suspected.

In the next experiment, we measured the processing time of our method. The procedure of human detection shown in Figure 1 took an average processing time of about 27.3 ms/frame (about 27.7 ms/frame for day images and about 26.9 ms/frame for night images). With the detected human area, the average processing time of behavior recognition of Figure 3 is shown in Table 6. Based on these results, we can confirm that our system can be operated at a speed of about 33.7 frames/s (1000/29.7) including the procedures of human detection and behavior recognition. Considering only the procedure of behavior recognition, this procedure can be operated at a speed of 416.7 (1000/2.4) frames/s.

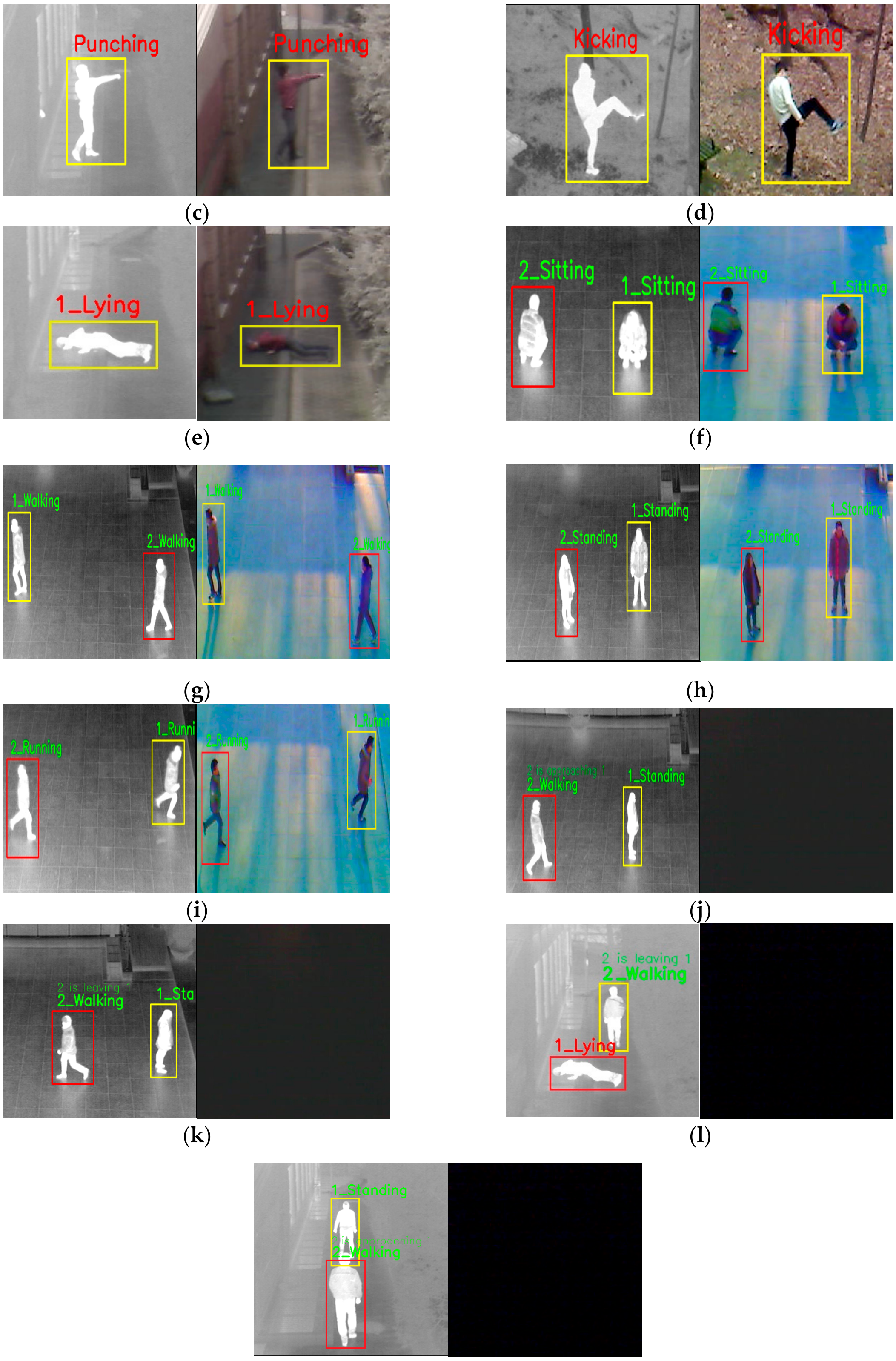
Table 5. Accuracies of behavior recognition by our method (unit: %).

	Day				Night			
Behavior	**TPR**	**PPV**	**ACC**	**F_Score**	**TPR**	**PPV**	**ACC**	**F_Score**
Walking	92.6	100	92.7	96.2	98.5	100	98.5	99.2
Running	96.6	100	96.7	98.3	94.6	100	94.7	97.2
Standing	100	100	100	100	97.3	100	97.3	98.6
Sitting	92.5	100	92.5	96.1	96.5	100	96.5	98.2
Approaching	100	100	100	100	100	100	100	100
Leaving	100	100	100	100	100	100	100	100
Waving with two hands	95.9	98.8	99.4	97.3	97.0	99.5	99.6	98.2
Waving with one hand	93.6	99.4	99.3	96.4	90.0	100	99.0	94.7
Punching	87.8	99.5	98.0	93.3	77.4	99.6	96.4	87.1
Lying down	99.4	99.0	98.9	99.2	97.3	98.0	96.3	97.6
Kicking	90.6	95.8	97.2	93.1	88.5	90.3	94.6	89.4
Average	95.4	99.3	97.7	97.3	94.3	98.9	97.5	96.4



(**a**) (**b**)

Figure 13. *Cont.*



(**c**) (**d**) (**e**) (**f**) (**g**) (**h**) (**i**) (**j**) (**k**) (**l**) (**m**)

Figure 13. Examples of correct behavior recognition. In (**a**–**m**), the images on the left and right are obtained by thermal and visible light camera, respectively. (**a**) Waving with two hands; (**b**) waving with one hand; (**c**) punching; (**d**) kicking; (**e**) lying down; (**f**) sitting; (**g**) walking; (**h**) standing; (**i**) running; (**j**) and (**m**) approaching (nighttime); (**k**) and (**l**) leaving (nighttime).

Table 6. Processing time of our method for each behavior dataset (unit: ms/frame).

	Processing Time	
Behavior	**Day**	**Night**
Walking	2.3	2.6
Running	1.5	1.5
Standing	3.2	3.1
Sitting	1.9	2.0
Approaching	3.3	2.9
Leaving	2.9	2.9
Waving with two hands	3.2	3.1
Waving with one hand	2.6	2.1
Punching	2.7	2.3
Lying down	1.2	1.0
Kicking	1.9	2.0
Average	2.4	

3.3. Comparison between the Accuracies Obtained by Our Method and Those Obtained by Previous Methods

In the next experiment, we performed comparisons between the accuracies of behavior recognition by our method and those by previous methods. These comparisons tested the following methods: Fourier descriptor-based method by Tahir et al. [22], GEI-based method by Chunli et al. [33], and convexity defect-based methods by Youssef [39]. As shown in Tables 5 and 7, our method outperforms the previous methods for datasets of all types of behavior.

Table 7. Accuracies of other methods (unit: %) *.

	Fourier Descriptor-Based				**GEI-Based**				**Convexity Defect-Based**			
Behavior	**TPR**	**PPV**	**ACC**	**F_Score**	**TPR**	**PPV**	**ACC**	**F_Score**	**TPR**	**PPV**	**ACC**	**F_Score**
Walking	0	0	1.6	-	83.9	97.7	82.4	90.3	17.4	98.4	18.4	29.6
Running	13.0	97.2	16.0	22.9	0	0	3.5	-	23.4	98.4	27.6	37.8
Standing	85.9	100	85.9	92.4	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
Sitting	59.7	100	59.7	74.8	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
Waving with two hands	81.4	13.7	47.8	23.5	96.5	19.1	59.6	31.9	89.0	68.6	94.9	77.4
Waving with one hand	78.2	13.8	50.3	23.5	21.2	10.1	73.8	13.7	34.9	73.4	92.4	47.3
Punching	27.1	20.3	74.6	23.2	62.2	16.5	50.1	26.1	39.7	55.7	86.9	46.4
Lying down	12.9	72.0	38.6	21.9	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
Kicking	61.1	42.0	78.7	49.8	24.7	80.7	86.0	37.8	29.3	47.6	82.2	36.3
Average	46.6	51.0	50.4	48.7	55.5	46.3	65.1	50.5	47.7	77.4	71.8	59.0

* n/a represents "not available" (the method was unable to produce a result for behavior recognition).

In Table 7, because all the previous methods are able to recognize individual types of behavior (rather than interaction-based behavior), interaction-based behaviors such as leaving and approaching were not compared. The reason for the difference between the ACC and F_score in some cases (for example, punching, by the Fourier descriptor-based method) is that #TN is large in these cases, which increases the ACC as shown in Equation (10).

Since the GEI-based method is based on human motion, motionless behaviors, such as standing, sitting, and lying down, cannot be recognized. The convexity defect-based method is based on a defect point, which cannot be obtained when the distance between two hands and two feet is small. In the case of standing, sitting, and lying down, the distance becomes small; hence, the defect points are not produced. Therefore, these types of behavior cannot be recognized and the recognition results are indicated as "n/a".

In Fourier descriptor-based methods, walking and lying down are recognized with very low accuracy. This is because walking, running, and kicking behaviors have a very similar pattern of the Fourier descriptor, which means that walking is recognized as kicking or running. Moreover, standing and lying down have similar patterns, such that lying down has been recognized as a standing behavior (see Figure 14a–c). In addition, because of the rotation in variant characteristics of the Fourier descriptor, the Fourier descriptor of the standing behavior in Figure 14d has a pattern similar to that of the lying down behavior in Figure 14e.

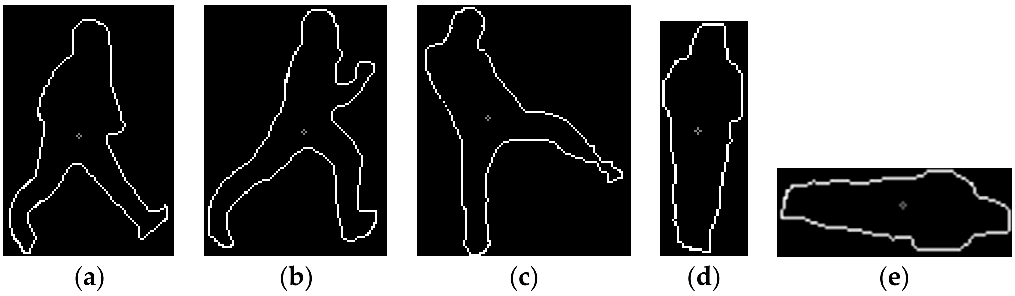


Figure 14. Examples of cases in which the Fourier descriptor-based method produced an erroneous recognition result. (**a**) Walking; (**b**) running; (**c**) kicking; (**d**) standing; (**e**) lying down.

GEI-based methods have similar patterns for walking, running, and kicking behaior. Therefore, running was recognized as walking (see Figure 15c,d). In addition, a kicking image produced by GEI is not much different from those showing walking or running as shown in Figure 15. Therefore, the PPV and consequent F_score of kicking are low as shown in Table 7.

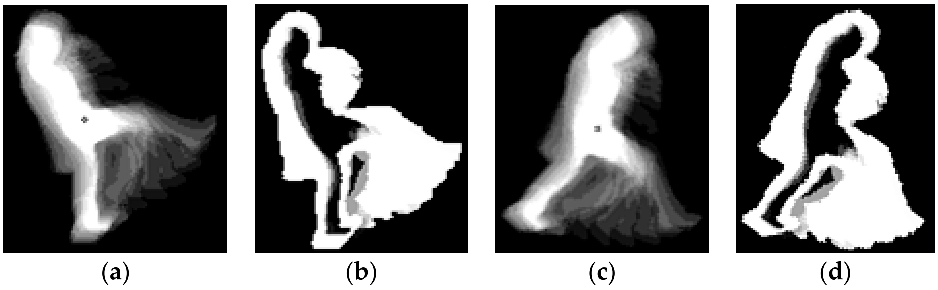


Figure 15. Examples of cases in which the GEI-based method produced an erroneous recognition result. Kicking images by (**a**) GEI; and (**b**) EGEI; running or walking image by (**c**) GEI; and (**d**) EGEI.

In the case of a small amount of hand waving (Figure 16a), GEI-based methods do not provide correct information for behavior recognition. In addition, because our camera set captures images from an elevated position (height of 5–10 m) (see Figures 12 and 16, and Table 3), the information produced by GEI-based methods in the case of hand waving is not distinctive, as shown in Figure 16b,c. Therefore, the accuracies obtained for images showing waving with one or two hands by GEI-based methods are low, as shown in Table 7.

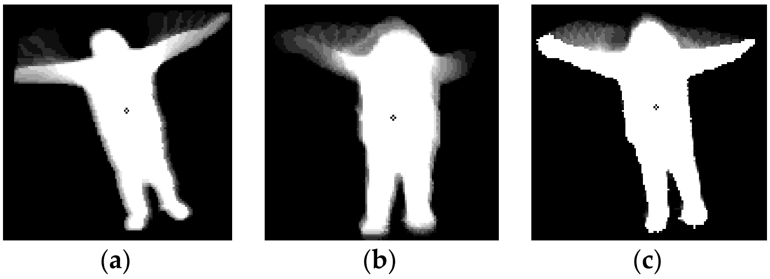


Figure 16. Examples of cases in which the GEI-based method produced an erroneous recognition result. (**a**) A small amount of waving; (**b**,**c**) cases in which information produced by GEI is not distinctive because the image is captured by our camera system installed at a height of 5–10 m (see Figure 12 and Table 3).

The convexity defect-based method produces the errors for behavior recognition of waving, punching, and kicking because of the error of detection of the positions of the hand and leg tip. As shown in Figure 17, in addition to those of the hand and leg tip, other points are incorrectly detected by the convexity defect-based method, which causes an error of behavior recognition for waving, punching, and kicking.

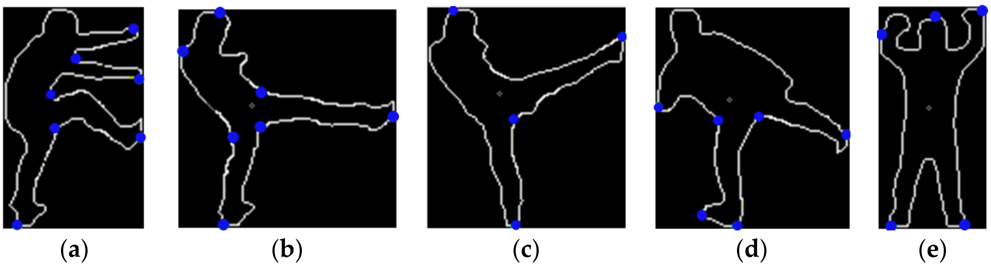


Figure 17. Examples of cases in which the convexity defect-based method produced an erroneous recognition result. (**a**–**d**) Kicking; (**e**) waving with two hands.

Table 8 presents the behavior recognition results of our method as a confusion matrix. In this table, the actual behavior means the ground-truth behavior, whereas predicted behavior represents that recognized by our method. Therefore, the higher values shown in the diagonal cell of the table (i.e., "standing" both in terms of actual and predicted behaviors) indicate the higher accuracies of behavior recognition. In some cases in Table 8, the sum of values in a row is not 100%, which means that an "undetermined case" scenario, as shown in Figure 3, occurs. For example, for Table 8, in the case of "punching", the percentage of "undetermined cases" is 15.3% (100 − (0.3 + 1.6 + 0.3 + 82.5)) (%).

Table 8. Confusion matrix of the results of behavior recognition by our method (unit: %).

Predicted Actual	Walking	Running	Standing	Sitting	Waving with two hands	Waving with one hand	Lying down	Kicking	Punching
Walking	95.5		0.3	0.4				0.5	
Running	0.4	95.6				0.3		3.3	0.4
Standing			98.5						
Sitting			0.5	94.6					
Waving with two hands					96.4			0.3	
Waving with one hand			0.2	0.1		91.8			0.1
Lying down	0.3			0.6			98.3		
Kicking	0.2			3.0		0.3		89.5	
Punching			0.3	1.6				0.3	82.5

Table 9. Summarized comparisons of accuracies and processing time obtained by previous methods and by our method.

Method	TPR (%)	PPV (%)	ACC (%)	F_score (%)	Processing Time (ms/frame)
Fourier descriptor-based	46.6	51.0	50.4	48.7	16.1
GEI-based	55.5	46.3	65.1	50.5	4.9
Convexity defect-based	47.7	77.4	71.8	59.0	5.2
Our method	94.8	99.1	97.6	96.8	2.4

In Table 9, we present a summarized comparison of the accuracies and processing time obtained by previous methods and by our method. The processing time only represents the time taken by behavior recognition, and excludes the time for detection of the human box because this is the same in each method. Based on the results in Table 9, we can confirm that our method outperforms other methods both in terms of accuracy and processing time.

Our method is mainly focused on behavioral recognition based on frontal-viewing human body, like most previous studies [17–20,26,27,30,32–36,38,39,41,42,51]. The change of view direction (from frontal view to side view) does not affect the recognition accuracy of "standing" and "sitting" of Figure 3. Because our method already considers "walking" and "running" in various directions (horizontal, vertical, or diagonal), as shown in Figure 3, the change of view direction does not have much effect on the recognition accuracy of "walking" and "running" in Figure 3, either. However, "lying down" is mainly recognized based on the average ratio of height to width of the detected human box, as shown in Figure 3. Therefore, the recognition accuracy can be highly affected by a change in view direction because the average ratio of height to width of the detected box is changed according to the change of view direction. In the case of behaviors such as "kicking", "waving", and "punching", they are recognized based on the detected tip positions of hands and feet. Therefore, the movement of these positions is difficult to detect in side-viewing images, and the recognition accuracy can be highly affected by the change of view direction. The recognition of "lying down", "kicking", waving", and "punching" only by one side-viewing camera is s challenging problem that has not been solved in previous studies. Therefore, our method is also focused on behavioral recognition based on a frontal view of the human body, like most previous studies.

4. Conclusions

In this research, we propose a robust system for recognizing behavior by fusing data captured by thermal and visible light camera sensors. We performed behavior recognition in an environment in which a surveillance camera would typically be used, i.e., where the camera is installed at a position elevated in relation to that of the human. In this case, behavior recognition usually becomes more difficult because the camera looks down on people. In order to solve the shortcomings of previous studies of behavior recognition based on GEI and a region-based method, we propose the PbD method. The PbD method employs a binarized human blob, which it uses to detect the tip positions of hands and legs. Then, various types of behavior are recognized based on the information obtained by tracking these tip positions and our decision rules. We constructed multiple datasets collected in various environments, and compared the accuracies and processing time obtained by our method to those obtained by other researchers. The experiments enabled us to confirm that our method outperforms other methods both in terms of accuracy and processing time.

In the future, we plan to increase the number of behavioral types by including those associated with emergency situations such as kidnapping, etc., for our experiments. In addition, we plan to research a method for enhancing the accuracy of determining emergency situations by combining our vision-based method with other sensor-based methods such as audio sensors.

Acknowledgments: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01056761).

Author Contributions: Ganbayar Batchuluun and Kang Ryoung Park designed the overall system and made the behavior recognition algorithm. In addition, they wrote and revised the paper. Yeong Gon Kim, Jong Hyun Kim, and Hyung Gil Hong helped to develop the method of human detection and collected our database.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Besbes, B.; Rogozan, A.; Rus, A.-M.; Bensrhair, A.; Broggi, A. Pedestrian Detection in Far-Infrared Daytime Images Using a Hierarchical Codebook of SURF. *Sensors* **2015**, *15*, 8570–8594. [CrossRef] [PubMed]
2. Zhao, X.; He, Z.; Zhang, S.; Liang, D. Robust Pedestrian Detection in Thermal Infrared Imagery Using a Shape Distribution Histogram Feature and Modified Sparse Representation Classification. *Pattern Recognit.* **2015**, *48*, 1947–1960. [CrossRef]

3. Li, Z.; Zhang, J.; Wu, Q.; Geers, G. Feature Enhancement Using Gradient Salience on Thermal Image. In Proceedings of the International Conference on Digital Image Computing: Techniques and Applications, Sydney, Australia, 1–3 December 2010; pp. 556–562.
4. Chang, S.L.; Yang, F.T.; Wu, W.P.; Cho, Y.A.; Chen, S.W. Nighttime Pedestrian Detection Using Thermal Imaging Based on HOG Feature. In Proceedings of the International Conference on System Science and Engineering, Macao, China, 8–10 June 2011; pp. 694–698.
5. Lin, C.-F.; Chen, C.-S.; Hwang, W.-J.; Chen, C.-Y.; Hwang, C.-H.; Chang, C.-L. Novel Outline Features for Pedestrian Detection System with Thermal Images. *Pattern Recognit.* **2015**, *48*, 3440–3450. [CrossRef]
6. Bertozzi, M.; Broggi, A.; Rose, M.D.; Felisa, M.; Rakotomamonjy, A.; Suard, F. A Pedestrian Detector Using Histograms of Oriented Gradients and a Support Vector Machine Classifier. In Proceedings of the IEEE Conference on Intelligent Transportation Systems, Seattle, WA, USA, 30 September–3 October 2007; pp. 143–148.
7. Li, W.; Zheng, D.; Zhao, T.; Yang, M. An Effective Approach to Pedestrian Detection in Thermal Imagery. In Proceedings of the International Conference on Natural Computation, Chongqing, China, 29–31 May 2012; pp. 325–329.
8. Wang, W.; Wang, Y.; Chen, F.; Sowmya, A. A Weakly Supervised Approach for Object Detection Based on Soft-Label Boosting. In Proceedings of the IEEE Workshop on Applications of Computer Vision, Tampa, FL, USA, 15–17 January 2013; pp. 331–338.
9. Wang, W.; Zhang, J.; Shen, C. Improved Human Detection and Classification in Thermal Images. In Proceedings of the IEEE International Conference on Image Processing, Hong Kong, China, 26–29 September 2010; pp. 2313–2316.
10. Takeda, T.; Kuramoto, K.; Kobashi, S.; Haya, Y. A Fuzzy Human Detection for Security System Using Infrared Laser Camera. In Proceedings of the IEEE International Symposium on Multiple-Valued Logic, Toyama, Japan, 22–24 May 2013; pp. 53–58.
11. Sokolova, M.V.; Serrano-Cuerda, J.; Castillo, J.C.; Fernández-Caballero, A. A Fuzzy Model for Human Fall Detection in Infrared Video. *J. Intell. Fuzzy Syst.* **2013**, *24*, 215–228.
12. Ghiass, R.S.; Arandjelović, O.; Bendada, H.; Maldague, X. Infrared Face Recognition: A Literature Review. In Proceedings of the International Joint Conference on Neural Networks, Dallas, TX, USA, 4–9 August 2013; pp. 1–10.
13. Davis, J.W.; Sharma, V. Background-subtraction Using Contour-based Fusion of Thermal and Visible Imagery. *Comput. Vis. Image Underst.* **2007**, *106*, 162–182. [CrossRef]
14. Arandjelović, O.; Hammoud, R.; Cipolla, R. Thermal and Reflectance Based Personal Identification Methodology under Variable Illumination. *Pattern Recognit.* **2010**, *43*, 1801–1813. [CrossRef]
15. Leykin, A.; Hammoud, R. Pedestrian Tracking by Fusion of Thermal-Visible Surveillance Videos. *Mach. Vis. Appl.* **2010**, *21*, 587–595. [CrossRef]
16. Kong, S.G.; Heo, J.; Boughorbel, F.; Zheng, Y.; Abidi, B.R.; Koschan, A.; Yi, M.; Abidi, M.A. Multiscale Fusion of Visible and Thermal IR Images for Illumination-Invariant Face Recognition. *Int. J. Comput. Vis.* **2007**, *71*, 215–233. [CrossRef]
17. Rahman, S.A.; Li, L.; Leung, M.K.H. Human Action Recognition by Negative Space Analysis. In Proceedings of the IEEE International Conference on Cyberworlds, Singapore, 20–22 October 2010; pp. 354–359.
18. Rahman, S.A.; Leung, M.K.H.; Cho, S.-Y. Human Action Recognition Employing Negative Space Features. *J. Vis. Commun. Image Represent.* **2013**, *24*, 217–231. [CrossRef]
19. Rahman, S.A.; Cho, S.-Y.; Leung, M.K.H. Recognising Human Actions by Analyzing Negative Spaces. *IET Comput. Vis.* **2012**, *6*, 197–213. [CrossRef]
20. Rahman, S.A.; Song, I.; Leung, M.K.H.; Lee, I.; Lee, K. Fast Action Recognition Using Negative Space Features. *Expert Syst. Appl.* **2014**, *41*, 574–587. [CrossRef]
21. Fating, K.; Ghotkar, A. Performance Analysis of Chain Code Descriptor for Hand Shape Classification. *Int. J. Comput. Graph. Animat.* **2014**, *4*, 9–19. [CrossRef]
22. Tahir, N.M.; Hussain, A.; Samad, S.A.; Husain, H.; Rahman, R.A. Human Shape Recognition Using Fourier Descriptor. *J. Electr. Electron. Syst. Res.* **2009**, *2*, 19–25.
23. Toth, D.; Aach, T. Detection and Recognition of Moving Objects Using Statistical Motion Detection and Fourier Descriptors. In Proceedings of the IEEE International Conference on Image Analysis and Processing, Mantova, Italy, 17–19 September 2003; pp. 430–435.

24. Harding, P.R.G.; Ellis, T. Recognizing Hand Gesture Using Fourier Descriptors. In Proceedings of the IEEE International Conference on Pattern Recognition, Washington, DC, USA, 23–26 August 2004; pp. 286–289.
25. Ismail, I.A.; Ramadan, M.A.; El danaf, T.S.; Samak, A.H. Signature Recognition Using Multi Scale Fourier Descriptor and Wavelet Transform. *Int. J. Comput. Sci. Inf. Secur.* **2010**, *7*, 14–19.
26. Sun, X.; Chen, M.; Hauptmann, A. Action Recognition via Local Descriptors and Holistic Features. In Proceedings of the Workshop on Computer Vision and Pattern Recognition for Human Communicative Behavior Analysis, Miami, FL, USA, 20–26 June 2009; pp. 58–65.
27. Schüldt, C.; Laptev, I.; Caputo, B. Recognizing Human Actions: A Local SVM Approach. In Proceedings of the IEEE International Conference on Pattern Recognition, Cambridge, UK, 23–26 August 2004; pp. 32–36.
28. Laptev, I.; Lindeberg, T. Space-time Interest Points. In Proceedings of the IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; pp. 432–439.
29. Laptev, I. On Space-Time Interest Points. *Int. J. Comput. Vis.* **2005**, *64*, 107–123. [CrossRef]
30. Wang, L.; Qiao, Y.; Tang, X. Motionlets: Mid-level 3D Parts for Human Motion Recognition. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Porland, OR, USA, 23–28 June 2013; pp. 2674–2681.
31. Arandjelović, O. Contextually Learnt Detection of Unusual Motion-Based Behaviour in Crowded Public Spaces. In Proceedings of the 26th International Symposium on Computer and Information Sciences, London, UK, 26–28 September 2011; pp. 403–410.
32. Eum, H.; Lee, J.; Yoon, C.; Park, M. Human Action Recognition for Night Vision Using Temporal Templates with Infrared Thermal Camera. In Proceedings of the International Conference on Ubiquitous Robots and Ambient Intelligence, Jeju Island, Korea, 30 October–2 November 2013; pp. 617–621.
33. Chunli, L.; Kejun, W. A Behavior Classification Based on Enhanced Gait Energy Image. In Proceedings of the IEEE International Conference on Network and Digital Society, Wenzhou, China, 30–31 May 2010; pp. 589–592.
34. Liu, J.; Zheng, N. Gait History Image: A Novel Temporal Template for Gait Recognition. In Proceedings of the IEEE International Conference on Multimedia and Expo, Beijing, China, 2–5 July 2007; pp. 663–666.
35. Kim, W.; Lee, J.; Kim, M.; Oh, D.; Kim, C. Human Action Recognition Using Ordinal Measure of Accumulated Motion. *Eur. J. Adv. Signal Process.* **2010**, *2010*, 1–12. [CrossRef]
36. Han, J.; Bhanu, B. Human Activity Recognition in Thermal Infrared Imagery. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, San Diego, CA, USA, 25 June 2005; pp. 17–24.
37. Lam, T.H.W.; Cheung, K.H.; Liu, J.N.K. Gait Flow Image: A Silhouette-based Gait Representation for Human Identification. *Pattern Recognit.* **2011**, *44*, 973–987. [CrossRef]
38. Wong, W.K.; Lim, H.L.; Loo, C.K.; Lim, W.S. Home Alone Faint Detection Surveillance System Using Thermal Camera. In Proceedings of International Conference on Computer Research and Development, Kuala Lumpur, Malaysia, 7–10 May 2010; pp. 747–751.
39. Youssef, M.M. Hull Convexity Defect Features for Human Action Recognition. Ph.D. Thesis, University of Dayton, Dayton, OH, USA, August 2011.
40. Zhang, D.; Wang, Y.; Bhanu, B. Ethnicity Classification Based on Gait Using Multi-view Fusion. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, San Francisco, CA, USA, 13–18 June 2010; pp. 108–115.
41. Rusu, R.B.; Bandouch, J.; Marton, Z.C.; Blodow, N.; Beetz, M. Action Recognition in Intelligent Environments Using Point Cloud Features Extracted from Silhouette Sequences. In Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication, Munich, Germany, 1–3 August 2008; pp. 267–272.
42. Kim, S.-H.; Hwang, J.-H.; Jung, I.-K. Object Tracking System with Human Extraction and Recognition of Behavior. In Proceedings of the International Conference on Management and Artificial Intelligence, Bali, Indonesia, 1–3 April 2011; pp. 11–16.
43. Lee, J.H.; Choi, J.-S.; Jeon, E.S.; Kim, Y.G.; Le, T.T.; Shin, K.Y.; Lee, H.C.; Park, K.R. Robust Pedestrian Detection by Combining Visible and Thermal Infrared Cameras. *Sensors* **2015**, *15*, 10580–10615. [CrossRef] [PubMed]
44. Tau 2. Available online: http://mds-flir.com/datasheet/FLIR_Tau2_Family_Brochure.pdf (accessed on 21 March 2016).

45. Infrared Lens. Available online: http://www.irken.co.kr/ (accessed on 31 March 2016).
46. Jeon, E.S.; Choi, J.-S.; Lee, J.H.; Shin, K.Y.; Kim, Y.G.; Le, T.T.; Park, K.R. Human Detection Based on the Generation of a Background Image by Using a Far-Infrared Light Camera. *Sensors* **2015**, *15*, 6763–6788. [CrossRef] [PubMed]
47. Gorelick, L.; Blank, M.; Shechtman, E.; Irani, M.; Basri, R. Actions as Space-Time Shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 2247–2253. [CrossRef] [PubMed]
48. Ikizler, N.; Forsyth, D.A. Searching for Complex Human Activities with No Visual Examples. *Int. J. Comput. Vis.* **2008**, *80*, 337–357. [CrossRef]
49. The LTIR Dataset v1.0. Available online: http://www.cvl.isy.liu.se/en/research/datasets/ltir/version1.0/ (accessed on 21 February 2016).
50. Precision and Recall. Available online: https://en.wikipedia.org/wiki/Precision_and_recall (accessed on 7 February 2016).
51. Gehrig, D.; Kuehne, H.; Woerner, A.; Schultz, T. HMM-based Human Motion Recognition with Optical Flow Data. In Proceedings of the IEEE-RAS International Conference on Humanoid Robots, Paris, France, 7–10 December 2009; pp. 425–430.