Jurnal
Teknologi

Full Paper

ROBUST HUMAN DETECTION WITH OCCLUSION HANDLING BY FUSION OF THERMAL AND DEPTH IMAGES FROM MOBILE ROBOT

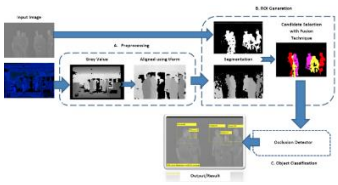
Saipol Hadi Hasim[a*], Rosbi Mamat[a], Usman Ullah Sheikh[b], Shamsuddin Mohd Amin[a],

[a]Dept. of Mechatronics and Robotics, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia
[b]Dept. of Electronic and Computer Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

Article history
Received
11 November 2015
Received in revised form
25 March 2016
Accepted
11 April 2016

*Corresponding author
shshadi2@live.utm.my

Graphical abstract



Abstract

In this paper, a robust surveillance system to enable robots to detect humans in indoor environments is proposed. The proposed method is based on fusing information from thermal and depth images which allows the detection of human even under occlusion. The proposed method consists of three stages; pre-processing, ROI generation and object classification. A new dataset was developed to evaluate the performance of the proposed method. The experimental results show that the proposed method is able to detect multiple humans under occlusions and illumination variations.

Keywords: Human detection; occlusion handling; mobile robot; depth imaging; thermal imaging

1.0 INTRODUCTION

The main function of a surveillance system is to monitor an area of interest from any incident. Basically, it involves guarding a property against invaders. Unfortunately existing surveillance systems rely too much on human operators and this is unacceptable if the area to be monitored is large.

Previous surveillance systems are based on face recognition [1,2,3,4], motion [5], body detection [6,7,8] or a combination of all three [7,9]. Color-based segmentation methods have limited applications in changing illumination [10], thus some works have proposed the use of infrared spectrum [11,12,13,14].

The work presented here is part of an automated surveillance system involving mobile robots, where one of the tasks of a mobile robot is to know the exact location of people in its surroundings. The platform used in this work is the UTMbot mobile robot. A Kinect sensor and a thermal camera are placed on top of the robot for sensing the environment (see Figure 1).

This paper is organized as follows. Related works are described in Section II. In Section III, the proposed methodology is explained. Section IV describes the experimental setup and the results are in Section V, the paper is concluded.

Figure 1 UTMbot-Mobile Robot used in this work

2.0 RELATED WORKS

There are several different approaches to human detection such as using monocular vision [15][16], stereo vision [17][18], sonar + vision [2], thermal vision [19][20][21][22], vision + laser [23] and depth sensor such as Kinect [24]. Infrared camera has the advantage of simplifying the task of detection as it is an illumination invariant. Thermal vision simplifies the segmentation of human bodies or human body parts from the background. Some works utilize more than one sensor [25], however the main issue is to determine optimal method on how to incorporate information from multiple sensors.

Detection of multiple humans is one of the main element of a machine/robot especially for effective human-machine interaction. Most previous efforts on overcoming occlusion in detecting multiple objects are based on combining some camera inputs such as in [26][27][28]. One method is by using mask and appearance model to deal with shape changes and large occlusions [29]. Works in [30] developed Bayesian segmentation approach that combines the region-based background subtraction and human shape model for detecting humans under occlusion. In [31], the authors proposed a dynamic Bayesian network to overcome occlusion while in [32], an appearance model is used for detecting objects under occlusion. A template matching technique using appearance features was proposed by [33]. Works by [34][35][36] proposed techniques that combine the outputs of part detectors to compute the likelihood presence of multiple humans while considering possible occlusions.

Recently, [13] proposed a method of human detection using 2D head contour model and a 3D head surface model to detect seed regions and then detect humans by using a region growing algorithm. Whereas, work by [37] presented a human detection method aimed at handling occlusions using depth data obtained from 3D imaging methods. The proposed method uses a split-merge approach, over-segmentation and clustering on foreground regions followed by height validation. Consequently, evidence has shown that, if the occlusion is too severe, most of the current methods will fail [14],[38].

3.0 SYSTEM OVERVIEW

This paper presents an approach using thermal and depth images for human detection with occlusion handling. The overall block diagram is shown in Figure 2. The proposed method is based on the algorithm proposed in [39].

3.1 Pre-processing Module

First, image registration is performed on both depth and thermal images. This process is carried out offline to obtain camera calibration parameters so that the two images (depth and thermal) are properly registered producing an output fused image. The process is performed only once, manually by an operator by selecting similar control points in both images.

Next is the preprocessing stage whereby the depth information is normalized between 0 – 255. The input depth information, I_D has a resolution of 13-bits, and is normalized using $I'_D = \frac{I_D}{2^{13}} * 255$. Next is to register the thermal image, I_T with the depth image, I'_D using the camera parameters obtained from the registration process.

Viola-Jones cascade object detector is modified for the purpose of detecting upper human body structure. The cascade object detector is trained using a new set of samples gathered from thermal images of human upper body. This training consists of positive and negative images. The trained detector is then applied to the input thermal image, I_T to obtain the bounding box coordinates of persons and these coordinates are stored in a matrix which is called BBC (bounding box coordinates) which is then passed to the ROI generation process.

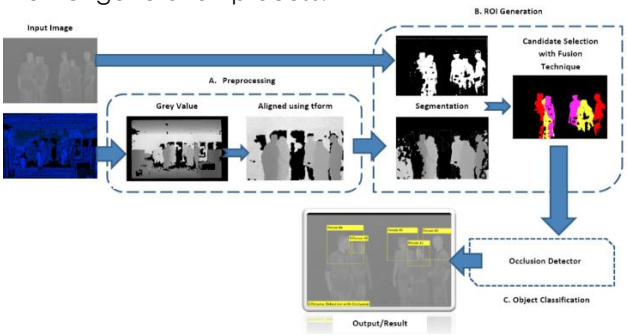


Figure 2 Overview of the human detection system with occlusion handling

3.2 ROI Generation

The purpose of ROI generation process is to extract regions of interest from the input images (depth and thermal) in the presence of background clutter. First, the depth image is processed to select ROI with the existence of people in the range of depth sensor. Due to sensor limitation, it is important to maintain the depth information in the range of 3 to 8 meters. Depth range below 3 meters and above 8 meters is inaccurate [40]. Thermal image is then filtered for temperature between 18-37°C for finding the existence of humans.

3.3 Object Classification

Object classification is conducted in two stages; person prediction and verification. In the first stage (person prediction), coordinates obtained from the pre-processing stage (BBC) is used to find a new person based on the distances in each x-axis and y-axis of the bounding box. A new matrix is created called NPC to collect coordinates and distances of

these new persons. This procedure is presented in Algorithm 1. In the second stage (verification), NPC is compared against BBC to eliminate the coordinates of the same person. Finally, the final matrix (RBB) of the bounding box is produced. This procedure is presented in Algorithm 2.

4.0 EVALUATION OF THE PROPOSED METHOD

Experiments were conducted in bright illumination as well as under no illumination (total darkness). Ground truth is used to evaluate the performance of the proposed method.

Algorithm 1: Person Prediction
Initialize the total number of BBCs
- *BBCj* = [*xj*, *yj*, *hj*, *wj*], *j*=1,……, N is total number of bounding box

For each BBC
- *j* is no. of current BBC
- Compute the height and width of the current bounding box based on min x and min *y*.
- Initialize the reference distance (Z^r).
- Compare distance in every pixel on bounding box against Z^r :
  - if distance is 0 and same Z^r, ignored and go to next pixel
  - if distance is not equal to Z^r, go to storage array
- Matrix of new person: NPC = [distance, x, y, j]. Store coordinates and distance of new person.

Algorithm 2: Verification
a. Initialize the total number of NPCs and BBCs
b. Occlusion handling: Eliminate coordinates of same person
   For each NPC:
   - *k* is no. of current NPC and *j* is no. of BBC in NPC

   For each BBC:
   - *d* is no. of current BBC.
   - if *j* equal to d, ignore and go to next d.
   - if *j* not equal to d, compare the value of x in current NPC to value of x-min and x-max in current BBC.
   - $X_{NPC} \geq Xmin_{BBC}$ && $X_{NPC} \leq Xmax_{BBC}$ is true, then all values of current NPC are removed.
   - $X_{NPC} \geq Xmin_{BBC}$ && $X_{NPC} \leq Xmax_{BBC}$ is false, then all values of current NPC are retained.
   - go to next *k*

Matrix of real person: update final matrix of bounding box, RBB = [BBC; NPC]

4.1 Dataset Collection

A new dataset was developed to assess the performance of the proposed method in handling occlusions for human detection. Existing human detection datasets are not specifically designed for evaluating occlusion handling. To specifically compare human detection algorithm under occlusions, occlusion dataset (MRUL) consisting of occluded persons was constructed. The dataset consists of thermal and depth images taken using a thermal camera (Raven 384) and Microsoft Kinect sensor mounted on a mobile platform in a laboratory environment [40]. The setup is to imitate the characteristics of human sight.

The proposed algorithm was tested on the UTMBot-mobile robot through experimental tests performed in indoor environment under two illumination conditions; bright illumination and no illumination (total darkness).

Table 1 shows the detail specification of the collected dataset. A total of 1,225 persons were labelled in the dataset manually as ground truth.

Table 1 Specification of the developed dataset

	MRUL-D1 (Bright)	MRUL-D2 (Dark)
Total Frame	320	270
Selected Frame	122	53
Max Person	10	10
Ground Truth, GT	776	340
GT Occlusion, O_{GT}	78	31
Total Ground Truth, T_{GT}	854	371

4.2 Ground Truth

To evaluate detection performance, the output of the proposed method is compared with the ground truth (GT). The performance is evaluated using the following metrics;

a) Number of hits (true positive - *TP*)
b) Number of wrong detection (false positive - *FP*)
c) Number of missed detection(false negative - *FN*)

Precision, recall and accuracy based on three parameters above are also computed.

a) Precision (*P*) to indicate the level of false positives (*FP*):

$$P = \frac{TP}{TP + FP} \quad (1)$$

b) Recall (*R*) to indicate true positives (*TP*):

$$R = \frac{TP}{TP + FN} \quad (2)$$

c) Accuracy (*A*):

$$A = \frac{TP + N_{GT}}{TP + N_{GT} + FP + FN} \quad (3)$$

5.0 RESULTS AND DISCUSSION

Performance is evaluated on the dataset shown in Table 1. 100 images are selected

Table 2 Results of occlusion handling for human detection

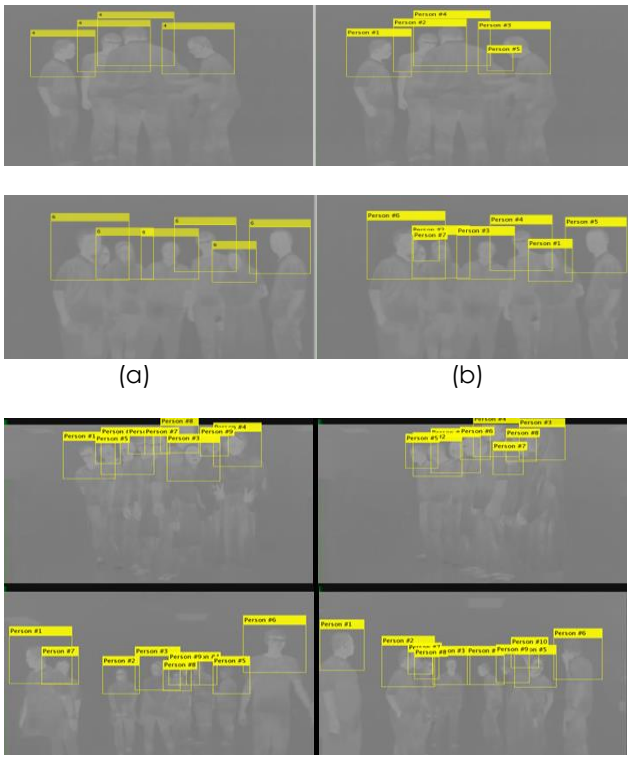
Dataset & Parameter / Item	MRUL-D1 (Bright)			MRUL-D2 (Dark)		
	TP	**FP**	**FN**	**TP**	**FP**	**FN**
Pre-Detection, BBC	765	1	11	331	0	9
Occlusion Detection, NPC_R	70	5	8	28	4	3
Total Detection, RBB	835	6	19	359	4	12

Table 3 Accuracy of the proposed method

	Percentage	
	MRUL-D1 (Bright)	MRUL-D2 (Dark)
Precision	99.28%	98.9%
Recall	97.78%	96.77%
Accuracy	98.54%	97.86%

Tables 2 and 3 show the results of human detection under two different illumination conditions (bright and total darkness). For the first dataset MRUL-D1, a total of 835 true positives were obtained resulting in a precision of 99.28%.

The second dataset MRUL-D2 is used to test the proposed method's applicability in total darkness. The dataset contains 371 persons whereby 53 frames consist of occluded persons. 359 true positives were detected as achieving precision of 98.9%. The average processing time for both datasets was found to be at 9.52 seconds per frame. Figure 3 shows sample output images of the proposed method on the dataset.



(a) (b)

(b)

Figure 3 Several test images used for performance evaluation. (a) Output of pre-detection process, (b) output after improved human detection and (c) several examples of successful multiple human detection.

6.0 CONCLUSION

In this paper, a human detection algorithm with occlusion handling for surveillance robots was proposed. The proposed method utilizes thermal and depth sensors to detect humans and resolve occlusions. The proposed algorithm was evaluated on a dataset developed specifically to evaluate occlusion detection. Some of the advantages of the proposed method are that it is computationally inexpensive and performs well even with occlusion or poor illumination. The current work is limited for indoor applications and future works should focus on addressing this limitation.

Acknowledgement

The authors would like to thank the Faculty of Electrical Engineering, Universiti Teknologi Malaysia for providing technical support for this research work. The authors are also grateful to Centre for Artificial Intelligence & Robotics (CAIRO) and Computer Vision, Video and Image Processing Research Lab (CVVIP) for providing financial support.

References

[1] M. T. Ahmed and S. H. M. Amin, . 2015. Comparison of Face Recognition Algorithms for Human-Robot Interactions. J. Teknol. 2: 73–78.

[2] T. Wilhelm, H.-J. Böhme, and H.-M. Gross, . 2004. A multi-Modal System for Tracking and Analyzing Faces on a Mobile Robot. Rob. Auton. Syst. Aug. 2004. 48(1): 31–40.

[3] R. C. Luo, A. C. Tsai, and C. T. Liao, . 2007. Face Detection and Tracking for Human Robot Interaction through Service Robot. The 33rd Annual Conference of the IEEE Industrial Electronics Society (IECON). 2818–2823.

[4] W. R. Schwartz, R. Gopalan, R. Chellappa, and L. S. Davis,. 2009. Robust Human Detection Under Occlusion by Integrating Face and Person Detectors. Lect. Notes Comput. Sci. Int. Conf. Biometrics. 5558: 970–979.

[5] M. Bennewitz, G. Cielniak, and S. Thrun. 2005. Learning Motion Patterns of People for Compliant Robot Motion. Int. J. Rob. Res. 24(3): 31–48.

[6] A. Møgelmose, C. Bahnsen, T. B. Moeslund, A. Clapés, and S. Escalera, . Tri-modal Person Re-identification with RGB . Depth and Thermal Features.

[7] W. Choi, C. Pantofaru, and S. Savarese, . 2011. Detecting and Tracking People Using an RGB-D Camera Via Multiple Detector Fusion. Proc. IEEE Int. Conf. Comput. Vis. 1076–1083.

[8] H. Zhang, C. Reardon, and L. E. Parker, . 2013. Real-time Multiple Human Perception with Color-depth Cameras on a Mobile Robot. IEEE Trans. Cybern. 43(5): 1429–1441.

[9] W. Choi, C. Pantofaru, and S. Savarese, . 2013. A General Framework for Tracking Multiple People from a Moving Camera. IEEE Trans. Pattern Anal. Mach. Intell. 35(7): 1577–1591.

[10] M. Correa, G. Hermosilla, R. Verschae, and J. Ruiz-del-Solar, . 2011. Human Detection and Identification by Robots Using Thermal and Visual Information in Domestic Environments. J. Intell. Robot. Syst. Jul. 2011. 66(1–2): 223–243.

[11] A. Treptow, G. Cielniak, and T. Duckett, . 2006. Real-time People Tracking for Mobile Robots Using Thermal Vision. Rob. Auton. Syst. 54: 729–739.

[12] F. Guan, L. Y. Li, S. S. Ge, and A. P. Loh, . 2007. Robust Human Detection And Identification by using Stereo and Thermal Images In . Human 1 St Reading. 65.

[13] L. Xia, C. Chen, and J. K. Aggarwal, . 2011. Human Detection Using Depth Information by Kinect. Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE. 15–22.

[14] S. Ikemura and H. Fujiyoshi, . 2011. Real-Time Human Detection using Relational Depth Similarity Features. Computer Vision – ACCV 2010, Lecture Notes in Computer Science . 6495: 25–38.
[15] G. Medioni, A. R. J. François, M. Siddiqui, K. Kim, and H. Yoon, . 2007. Robust Real-time Vision for a Personal Service Robot. Comput. Vis. Image Underst. Oct. 2007. 108(1–2): 196–203.
[16] H.-J. Böhme, T. Wilhelm, J. Key, C. Schauer, C. Schröter, H.-M. Groß, and T. Hempel, . 2003. An Approach to Multi-modal Human–machine Interaction for Intelligent Service Robots. Rob. Auton. Syst. Jul. 2003. 44(1): 83–96.
[17] J. Satake and J. Miura, . 2009. Robust Stereo-Based Person Detection and Tracking for a Person Following Robot. IEEE International Conference on Robotics and Automation. May.
[18] L. Lit and W. Huangt, . 2004. Stereo-Based Human Detection For Mobile Service Robots. 8th Internasional Conference on Control, Automation, Robotics and Vision. December. 6–9.
[19] U. Meis, M. Oberlander, and W. Ritter, . 2004. Reinforcing the Reliability of Pedestrian Detection in Far-infrared Sensing. IEEE Intelligent Vehicles Symposium. 779–783.
[20] S. S. Mudaly, . Novel Computer-Based Infrared Pedestrian Data-Acquisition. Electronics Letters. 371–372.
[21] H. Nanda and C. Park, . 2002. Probabilistic Template Based Pedestrian Detection in Infrared Videos. IEEE Intelligent Vehicles Symposium2002. 15–20.
[22] M. Bertozzi, A. Member, A. Broggi, A. Fascioli, T. Graf, and M. Meinecke, . 2004. Pedestrian Detection for Driver Assistance Using Multiresolution Infrared Vision. IEEE Trans. Veh. Technol. 53(6): 1666–1678.
[23] N. Bellotto, S. Member, H. Hu, and S. Member, . 2009. Multisensor-Based Human Detection and Tracking for Mobile Service Robots. IEEE Trans. Syst. ManCybern. - Part B Cybern. 39(1): 167–181.
[24] F. Jurado, G. Palacios, F. Flores, and H. M. Becerra, . 2014. Vision-Based Trajectory Tracking System for an Emulated Quadrotor UAV. Asian J. Control. 16(3): 729–741.
[25] D. Y. Gared and X. Ding, . Image Fusion for Concealed Weapon Detection. Int. J. Eng. Res. Technol. 2(2): 1–4.
[26] T.-H. Chang, S. Gong, and E.-J. Ong, . 2000. Tracking Multiple People under Occlusion Using Multiple Cameras. Proc. 11th British Machine Vision Conference.
[27] S. L. Dockstader and A. M. Tekalp,. 2001. Multiple camera fusion for multi-object tracking. Proceedings 2001 IEEE Workshop on Multi-Object Tracking. 95–102.
[28] S. L. Dockstader and A. M. Tekalp, . 2001. Multiple Camera Tracking of Interacting and Occluded Human Motion. Proceedings of the IEEE . 89(10): 1441–1455.
[29] R. Cucchiara, C. Grana, G. Tardini, R. Vezzani, and R. Emilia, . 2004. Probabilistic People Tracking for Occlusion Handling. Proceedings of the 17th International Conference on ICPR 2004. 132 – 135.
[30] H. Eng, J. Wang, A. H. Kam, and W. Yau, . 2004. A Bayesian Framework for Robust Human Detection and Occlusion Handling using Human Shape Model. Proceedings of the 17th International Conference on ICPR 2004. 2: 257–260.
[31] Y. Wu, T. Yu, and G. Hua, . 2003. Tracking appearances with occlusions. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 789–795.
[32] A. Senior, A. Hampapur, Y.-L. Tian, L. Brown, S. Pankanti, and R. Bolle, . 2006. Appearance Models for Occlusion Handling. Image Vis. Comput. Nov. 2006. 24(11): 1233–1243.
[33] H. T. Nguyen and A. W. M. Smeulders, . 2004. Fast Occluded Object Tracking by a Robust Appearance Filter. IEEE Trans. Pattern Anal. Mach. Intell. Aug. 2004. 26(8): 1099–104.
[34] B. Wu and R. Nevatia, . 2005. Detection of Multiple Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectorsin . Tenth IEEE International Conference on Computer Vision 2005. 1: 90 – 97 .
[35] B. Wu and R. Nevatia, . 2006. Tracking of Multiple Partially Occluded Humans based on Static Body Part Detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1(0–7): .
[36] B. Wu and R. Nevatia, . 2007. Detection and Tracking of Multiple Partially Occluded Humans by Bayesian Combination of Edgelet based Part Detectors. Int. J. Comput. Vis. 75(2): 247–266.
[37] L. Wang, K. L. Chan, and G. Wang, . 2013. Human Detection with Occlusion Handling by Over-Segmentation and Clustering on Foreground Regions. The 11th Asian Conference on Computer Vision (ACCV2012). 197–208.
[38] O. Camps and M. Sznaier, . 2001. Segmentation for Robust Tracking in the Presence of Severe Occlusion. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 483–489.
[39] P. Viola and M. Jones, . 2001. Rapid Object Detection Using a Boosted Cascade of Simple Features. Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition. 1: 511-518.
[40] H. S. Hadi, M. Rosbi, U. U. Sheikh, and S. H. M. Amin,. 2015. Fusion of Thermal and Depth Images for Occlusion Handling for Human Detection from Mobile Robot. The 10th Asian Control Conference 2015 (ASCC 2015).