Visual Perception for Multiple Human–Robot Interaction From Motion Behavior

Emrah Benli, *Member, IEEE*, Yuichi Motai, *Senior Member, IEEE*, and John Rogers, *Senior Member, IEEE*

***Abstract*—Visual perception is an important component for human–robot interaction processes in robotic systems. Interaction between humans and robots depends on the reliability of the robotic vision systems. The variation of camera sensors and the capability of these sensors to detect many types of sensory inputs improve the visual perception. The analysis of activities, motions, skills, and behaviors of humans and robots have been addressed by utilizing the heat signatures of the human body. The human motion behavior is analyzed by body movement kinematics, and the trajectory of the target is used to identify the objects and the human target in the omnidirectional (O-D) thermal images. The process of human target identification and gesture recognition by traditional sensors have problem for multitarget scenarios since these sensors may not keep all targets in their narrow field of view (FOV) at the same time. O-D thermal view increases the robots' line-of-sights and ability to obtain better perception in the absence of light. The human target is informed of its position, surrounding objects and any other human targets in its proximity so that humans with limited vision or vision disability can be assisted to improve their ability in their environment. The proposed method helps to identify the human targets in a wide FOV and light independent conditions to assist the human target and improve the human–robot and robot–robot interactions. The experimental results show that the identification of the human targets is achieved with a high accuracy.**

***Index Terms*—Command cognition, human motion analysis, multiple human targets, multiple robots, omnidirectional (O-D) camera, robotic perception, target identification, thermal vision, visual perception, walking behavior.**

I. Introduction

VISUAL perception for robotic systems has been the center of attraction and utilized in many different platforms and methods. The cognition in robots is used to understand their environment, recognize the surrounding objects, and identify human targets. The interaction between robots and humans is the most significant application of visual perception for autonomous systems. In order to interact with humans and the other robots, it is very critical to obtain accurate information such as behavior, shape, and movement. A successful analysis of these features of human targets will provide a reliable identification and recognition of the targets. The current methods are focused on human–robot interaction by focusing on a single target of interest or the behavior analysis of a group of targets. The identification of human targets and surrounding objects plays a critical role before there can be interaction between humans and robots. However, traditional sensors do not provide adequate field of view (FOV) to keep multiple targets in view and analyze the body features and human motion behaviors. The process of human gesture command recognition requires sufficient space for the command gestures and multiple human targets to track. After identification of the human targets and the surrounding objects, it is important for the robot to assist the target of interest in order to fulfill the demands of humans in their environment.

The proposed method addresses the problem in human target identification and gesture recognition, and meets the needs of humans with disabilities and humans who require assistance for critical missions under limited lighting conditions such as rescue and military operations. We used an omnidirectional (O-D) thermal imager to cover a wide FOV, horizontal 360°, and utilize the heat signature of a human body and the surrounding objects. The human thermal view is analyzed to understand the behavior of its trajectory and the movement kinematics during the target's motion. We have obtained a personal behavior signature for each human target to navigate around other humans while they are informed about their surroundings. Our method identifies the human targets with high accuracy and assists them by increasing their limited vision capabilities and solves the target identification and gesture recognition problems coming from traditional methods.

The article is organized as follows. Section II discusses related works. Section III discusses visual perception for human–robot interaction from motion behavior. Section IV discusses experimental results, and Section V includes conclusion and future work.

Manuscript received February 20, 2019; revised September 9, 2019 and November 24, 2019; accepted December 5, 2019. Date of publication December 27, 2019; date of current version June 3, 2020. This work was supported in part by the U.S. Navy, Naval Surface Warfare Center Dahlgren, in part by the U.S. Army Research Laboratory, and in part by the Ministry of National Education of Turkey. *(Corresponding author: Emrah Benli.)*

E. Benli is with the Department of Electrical and Electronics Engineering, Gümüşhane University, Gümüşhane 29100, Turkey (e-mail: benlie@vcu.edu).

Y. Motai is with the Department of Electrical and Computer Engineering, Virginia Commonwealth University, Richmond, VA 23284 USA (e-mail: ymotai@vcu.edu).

J. Rogers is with the U.S. Army Research Laboratory, Adelphi, MD 20783 USA (e-mail: john.g.rogers59.civ@mail.mil).

Digital Object Identifier 10.1109/JSYST.2019.2958747

II. Related Works

The related studies for the perception of multiple robots utilizing the combination of O-D thermal visual system, stereo thermal, and monocular thermal sensor is given in two sections. First, the literature search in terms of the robotic perception to understand its environment for the human–robot interaction is given in Section II-A. Second, Section II-B examines the relevant studies in the sense of interaction between humans and robots for command cognition. Table I shows the related studies in terms of their methods of robotic perception and analysis for the interaction.

TABLE I
ROBOTIC PERCEPTION FOR COGNITION

Perception	Analysis	Interaction
Visual [8]–[11]	Human activity	Gait recognition and human identification
Vocal [19]	Speech	Localization by spoken sentence
Visual [27]	Human skill	Transfer the human's skill to a robot
EEG + Visual [21]	Brain signal	Manipulation by reading the human's mind
Visual [24, 25]	Object	Object recognition, object weight analysis
Range + Visual [27]- [29]	Human behavior	Human activity recognition, behavior analysis for human-object interaction

A. Robotic Perception

The robotic perception is a critical factor for the applications of autonomous systems, security and safety, behavior analysis, and human–robot interaction. The visual perception is the most significant element for robotics systems. Some studies have been applied to visual perception in robotic systems with respect to human activity analysis and motion-based behavior [1]–[6]. Thermal and visible sensors were combined to identify people with dual-path network so that large-crossmodality and intracrossmodality caused by multiple sensors can be handled [7]. Gait recognition and person identification have also been applied to obtain more information from human targets [8]–[11]. Thermal image-based deep convolutional neural network method has been used gait-based human identification since the human face recognition is not possible from thermal images [12]. Visual perception helps to detect the arms or hands and coordinate these parts from the inspiration of human-like behaviors as well as for hand grasp analysis [13]. In order to improve the robotic visual perception, an O-D vision system is utilized with multiple systems such as door knob hand recognition system, three-dimensional (3-D) model based tracking system, and visual-compass system [14]–[16]. These are utilized to improve the robotic visual perception for autonomous systems.

B. Interaction of Humans and Robots for Command Cognition

Human–robot interaction for command cognition is inspired from interaction among humans [17], [18]. In order to show the vocal interaction between robot and human, self-localization of a robot by analysis of spoken sentences is proposed in [19]. Language acquisition from speech signals and image-based sign language recognition by robot is another method for the interaction proposed in [20] and [21]. Visual analysis of human actions is the most significant method for the social cognition to understand human behaviors [22]–[30]. A gaze tracking system is used to predict the effects of display clutter in real time [22], [23]. Human behaviors are evaluated when a human interacts with objects [24]. The actions are analyzed to acquire information about the weight of objects during the lifting process of the robot [25]. Human activities are investigated to recognize by a sensor network [26]. RGB-D sensors are used for human activity recognition using soft labels [27], navigation assistance to guide people with vision loss [28], and human–object interaction [29]. A perception system combines detection of several features of humans for the interaction [30]. A robot is manipulated by a brain–machine interaction after reading the operator's mind, while the operator sees the robot's vision [31]. Far-infrared sensor array has been used to recognize the commands from hand gesture [32]. Since it is difficult to spot hand gesture from low-resolution thermal image sequence, a voting-based approach is used to spot the hand gesture. Human–robot interaction for gesture recognition utilizes deep learning method in some works [33], [34]. Recent conventional methods and the deep learning-based methods require training process and a training dataset. Additional advantage of our method is to maintain the operation time without light dependency, while these methods can maintain their mission under limited light conditions. Our method overcomes the training process and maintains the operation process with no light dependency.

III. VISUAL PERCEPTION FOR HUMAN–ROBOT INTERACTION FROM MOTION BEHAVIOR

We proposed a new method to identify the human targets in the thermal O-D scene by using robotic visual perception for command cognition. This method improves the visual perception ability of the human targets while they have limited vision and helps targets to understand their environment by notifying them about their surroundings from the robotic vision system. The identification of the human targets is the initial objective while concurrently understanding the surroundings and visualizing the scenario for the human target. Section III-A analyzes the human targets with respect to their body kinematics. Section III-B utilizes the trajectories of each human target to categorize the targets. Then, Section III-C identifies each human target from the information obtained through the analysis of human kinematics and trajectories. Finally, gesture analysis is given for command cognition of human–robot and robot–robot interactions in Section III-D.

A. Human Kinematics Analysis

Human kinematics analysis is the first criteria to identify a human during its motion or static scene. Human bodies are extracted from the thermal view and those regions are evaluated by separation of the human blobs. The human targets are selected from dynamic targets for analysis of kinematics and trajectories to evaluate our method. The main target knows the commands and is in an upright position. The orientation of the human body around the arm region $[X_A\ Y_A\ Z_A]$, legs $[X_{R,L}\ Y_{R,L}\ Z_{R,L}]$, and head for the upper part of the target's blob $[X_U\ Y_U\ Z_U]$ is obtained for each human. Human body gives some clues about the disposition during the movement or static state. The propensity to a direction is also derived from the evaluation of the orientation between those kinematics. If we consider the angle between head and legs and the legs are open with a specific angle, the head kinematics gives the direction of movement from the slope of head. In Fig. 1, we can see human target H_1 is being illustrated with a walking direction to the left and the head tilted toward the same direction. In case of a static state of the target, the orientation between head, legs, and body including arms may be similar to human targets $H_{3,4}$.

The images from each sensors are converted to binary images by *binarize()*, then *connected_components()* finds the candidate regions in the binary images. For the human regions, *human_detect()* uses the human body ratios (decided in Section IV-C) to height/width ratio from candidate regions. The human regions are selected by *human_detect()* and separated into five equal rows by *separate()* in Algorithm 1 as the upper

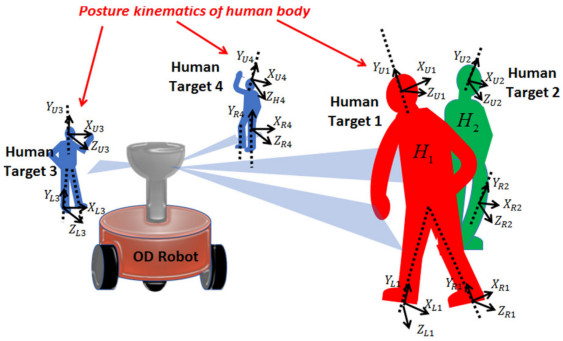


Fig. 1. Human body kinematics in a static view. Head and legs kinematics from human posture are used to identify the targets.

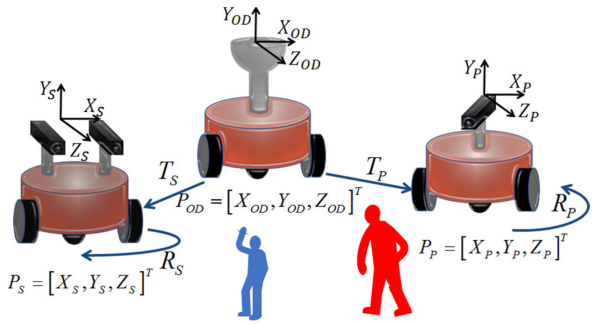


Fig. 2. Three robots equipped with O-D IR sensor, stereo thermal sensor, and single thermal camera. Translation and rotation between robots is shown with respect to the O-D robot.

Algorithm 1: Target Kinematic Analysis.

Input: O-D Thermal image I_{OD}, Stereo Thermal images I_S, Perspective Thermal image I_P, Threshold for binarization T_b
Output: Head kinematics K_U, Arm region kinematics K_A, Leg kinematics K_L
OD_b, S_b, $P_b \leftarrow$ *binarize*(I_{OD}, I_S, I_P) with Threshold T_b
for all $I = OD_b$, S_b, P_b *do*
  $TR \leftarrow$ *connected_components*(*I*) // *TR* is target region candidates
for all *TR do*
  Given human body ratio r_h, height h_h, width w_h
  $T_H \leftarrow$ *human_detect*(*TR*,r_h,h_h,w_h) // T_H is human target regions among the target regions
  U_h, A_h, $L_h \leftarrow$ *separate*(T_H) // U_h, A_h, L_h are head, arm, and leg regions for h^{th} target
for all U_h, A_h, L_h *do*
  K_U, K_A, $K_L \leftarrow$ *PCA*(U_h, A_h, L_h) // K_U, K_A, K_L are kinematics of head, arm region, and leg regions
end

part is head, the lower two rows are legs, and the middle two rows are the region including arms. Principal component analysis (PCA) is applied to the human regions in the thermal images after the whole human silhouette is separated for regions of head, arm region, and legs. The threshold value is selected from a human body heat signature for each sensor to identify the humanlike regions. The PCA method gives us the direction vector of these regions so that the orientation can be obtained from the direction vector of head, arm region, and legs. The algorithm of this process is given as follows.

After the kinematics of each human is analyzed with the orientations, the human targets are labeled with respect to their movement trends. The same process is applied from a different point of view by using another robotic visual perception. The position information of each robot, R_r, T_r, is used to transfer the human body kinematics for another view direction (see Fig. 2). The kinematics of each human target is obtained for the second visual perception and the first O-D kinematics are transferred to the second robot's position. This transfer is derived from each labeled target's kinematics, and the rotation matrix and translation vector of the transformation (1). The transferred kinematics for the second robot's point of view is given with H_S and the

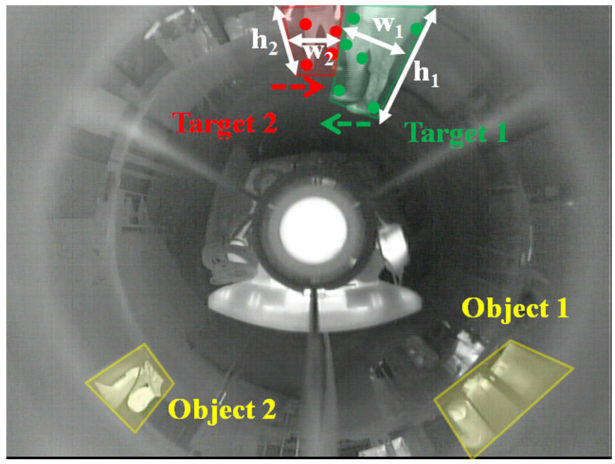


Fig. 3. Height and width ratio of each target gives the information about the targets' current movement direction. The detected and tracked feature points have the horizontal movement information. The objects are also detected and shown.

kinematics from the first robot's point of view, $[X_H\ Y_H\ Z_H]$, are multiplied with rotation between the orientation of the robots and a translation between the robots is also added

$$H_S = \begin{bmatrix} X_H & Y_H & Z_H \end{bmatrix} \begin{bmatrix} \cos\theta_r & -\sin\theta_r & 0 \\ \sin\theta_r & \cos\theta_r & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} x_r & y_r & z_r \end{bmatrix}. \quad (1)$$

The correlation between two sets of kinematics will be used to decide the labeled targets from two robotic perceptions to relate the same targets in two different thermal views.

B. Human Trajectory Analysis

The trajectories of the human targets are analyzed in the thermal scene with respect to the trajectory pattern of the corresponding target's feature points. The target regions are detected and the feature points are tracked during its motion, shown in Fig. 3. The width and height of the target region are calculated for a ratio to obtain initial information about the target's trajectory pattern. The ratio changes depending on the target's orientation during the rotation about its center. The width and height change for the target's forward and backward movement with respect

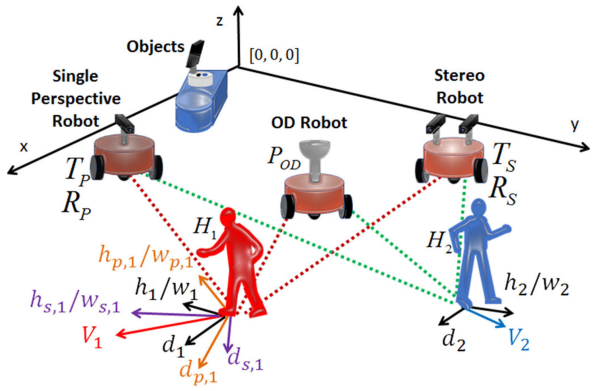


Fig. 4. Robots find the trajectory vector for each target from their height/ width ratios and horizontal movements in the image.

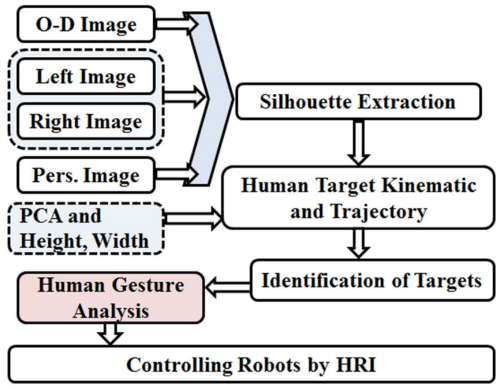


Fig. 5. Human–robot interaction algorithm using four input images from three robots.

to the robot. When the target's width increases, the target is considered to be approaching toward the robot. In Fig. 4, the forward and backward movement estimations and the magnitude of the direction vector is given for each robot. The changes of ratios from each robot's perspective are shown in Fig. 4. as well. These changes are recorded during the target's motion and updated for a time period by (2). The ratio h_s/w_s should remain stable if the target does not make a rotation around itself. In order to obtain the approach direction vector V_w, the width w_s, and the height h_s are used in the last two images

$$V_w = \begin{cases} w_{s,t} - w_{s,t-1} & \text{if } (h_s/w_s)_t - (h_s/w_s)_{t-1} = 0 \\ 0 & \text{if } (h_s/w_s)_t - (h_s/w_s)_{t-1} \neq 0. \end{cases} \tag{2}$$

If there is no change in the difference between consecutive ratios, the direction vector V_w is derived from the difference between consecutive target's width in these consecutive images. If the ratio is changing, we consider that the target is just making a rotation around itself.

The target's horizontal movement is also used for left and right directions in its trajectory. We consider the extracted feature points for a specific target is x_f, and the total number of these extracted feature point for this specific target is F. The horizontal movement of feature points x_f is tracked after calculation of their average coordinates until the total number of F. Then, the last average of horizontal positions at time $t-1$ is subtracted from the current average. This difference helps to obtain the horizontal vector magnitude and the direction d_s from the following equation:

$$d_s = \left(\sum_{f=1}^{F} x_{f,t}\right) \Big/ F - \left(\sum_{f=1}^{F} x_{f,t-1}\right) \Big/ F. \tag{3}$$

A trajectory pattern is created for each target in the O-D thermal image and it is compared with the other robots' trajectory estimations. Each robot's estimation uses horizontal and approach vectors to find the final trajectory vector

$$V_s = \sqrt{{V_w}^2 + {d_s}^2}. \tag{4}$$

The trajectories for each human target in the O-D thermal image are transformed to the other robot's coordinate system via transformation matrix.

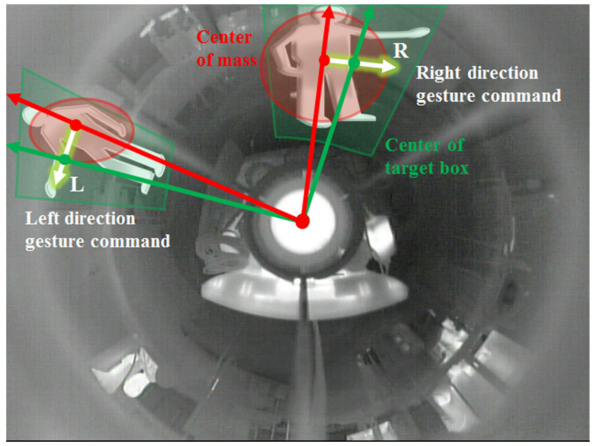


Fig. 6. Target's gesture commands are analyzed by the difference between the centers of target's body mass and box of the detected target region.

C. Identification of Human Targets

The identification of each target is done by combining the analysis of human kinematics and the trajectory of each target. The correlation of these features of the related target is derived after the analysis of kinematics and trajectories. After the targets are identified, the commands from the main target are analyzed to control robots' tasks. The algorithm of the human–robot interaction process from visual perception is given in Fig. 5. The detected target regions are labeled with respect to their identities with two criteria. The first is the mass center of the target region and the second criteria is a box drawn around the target's labeled region. A vector is obtained from the mass center to the box center to understand the human target's gesture, shown in Fig. 6. If the vector points to the right direction, the human target commands the robots to track any human target on his right. Then, O-D robot commands one of the robots to track the human target or possible object on the target's right after receiving the command from the human target.

The human target regions are detected and surrounded with a rectangle frame. The frame is divided into $S = (Hb/n_{\text{ver}})(Wb/n_{\text{hor}})$ number of cells to analyze each segment separately, shown in Fig. 7. The number of cells in vertical direction is given with n_{ver} and in horizontal direction n_{hor}. The human target is selected by using connected components

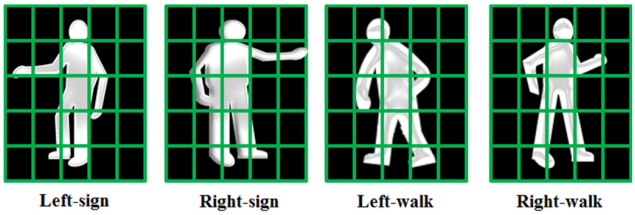


Fig. 7. Target gesture analysis from the segmented target silhouette. Target pixels are calculated for every segment for limbs and head orientation.

and separated from the other objects in the target frame. Then, the pixels in every cell are summed by (5) and a total coverage of human target is obtained for the corresponding cell. B_s is the total pixel value of the cell, s is the corresponding cell, Wb and Hb are width and height of the target frame, respectively. Every column and the row of the cells are also summed to have bird's-eye view and perspective view spectrum of the target as an additional target identical signature

$$B_s = \sum_{x=B_x}^{x_l} \sum_{y=B_y}^{y_l} p(x,y). \tag{5}$$

Every pixel in this cell is given with the pixel p and the coordinates of this pixel in the cell is (x, y). The last pixel coordinates from the left and bottom of every cell can be calculated from the following equation:

$$\begin{aligned} x_l &= B_x + s_x\,(Wb/n_{\text{hor}}) \\ y_l &= B_y + s_y\,(Hb/n_{\text{ver}})\,. \end{aligned} \tag{6}$$

The beginning coordinates of each cell B_x and B_y are obtained from the current coordinates of the target frame, and width and height of the frame from the following equation:

$$\begin{aligned} B_x &= \frac{Wb}{n_{\text{hor}}}(s_x - 1) \\ B_y &= \frac{Hb}{n_{\text{ver}}}(s_y - 1). \end{aligned} \tag{7}$$

The index of each cell is calculated from $s_y = \lfloor (s-1) /n_{\text{ver}} \rfloor + 1$ by using floor function after division and $s_x = s - s_y n_{\text{hor}}$. After obtaining the kinematics for head, arm, and leg regions, *corr()* decides the correlation of these region kinematics for every thermal image of the same target. Then, movement decision of the corresponding target is made. Another correlation result comes from trajectories by using the target trajectory directions with *corr()* to decide the final trajectory direction with the highest correlation result. The final decision is made by using two correlation result to identify human targets and label them with a number in step 7, *identify()*, of Algorithm 2. The algorithm to identify human targets from each robot according to their kinematics and features is given as follows.

The target can command a robot without any voice commands such as an environment that requires silence during the human–robot interaction process. The robot gives the improved vision to the human in limited lighting conditions while visibility is low and controls the other robots according to the human's command for robot–robot interaction.

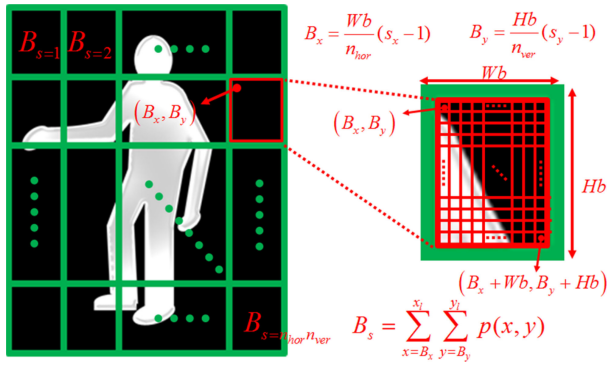


Fig. 8. Human gesture analysis by separating the body into cells and calculating the cells to read the command which matches with the predefined commands.

Algorithm 2: Target Identification.

1: *Input*: Kinematics of head K_U^h, kinematics of arm region K_A^h and kinematics of legs K_L^h for h^{th} target
Trajectory vectors of each target V_s^h
2: *Output*: Label of each target L_T^h
3: *for* all H (targets) *do*
4: $C_K \leftarrow corr(K_U^h, K_A^h, K_L^h)$ // C_K is correlation result for kinematics
5: $C_V \leftarrow corr(V_s^h)$ // C_V is correlation result for trajectory vectors
6: *end*
7: $L_T^h \leftarrow identify(C_K, C_V)$ // L_T^h is labeled targets
8: *end*

D. Command Cognition for Human–Robot Interaction

Command cognition is one of the supportive methods for human–robot interaction when human needs any assistance for vision, tracking, and navigation. Since humans have limited abilities such as a narrow FOV and a limited visible band, we use O-D IR sensor to increase the tracking of surrounding targets to assist humans by compensating this disadvantage. Thermal stereo and single thermal perspective camera mounted mobile robots are also used to improve the visual perception via multisensory setup and collaboration among the robots. Robot–robot interaction utilizes the multisensory data by collecting and processing the images from all robots. Second part of this interaction is the main robot leads the other robots after receiving the command from the main target. The communication between human–robot and robot–robot provides us an environment which consists of human and various machine systems.

The orientations of human body, head, and the limbs are mapped by using the values of $S = n_{\text{ver}} n_{\text{hor}}$ cells in the target frame while we assign the number of cells in vertical n_{ver} and horizontal direction n_{hor}, experimentally, shown in Fig. 8. We also used the bird's-eye view and perspective view spectrum vectors to validate the correlation by using (8). The correlation of three equal vertical column is used to correlate the spectrum vectors for the bird's-eye view. For perspective spectrum is analyzed by separating the region below the neck after finding the minimum of the neck row. Then, we checked the correlation of the specific orientation of human body with our command

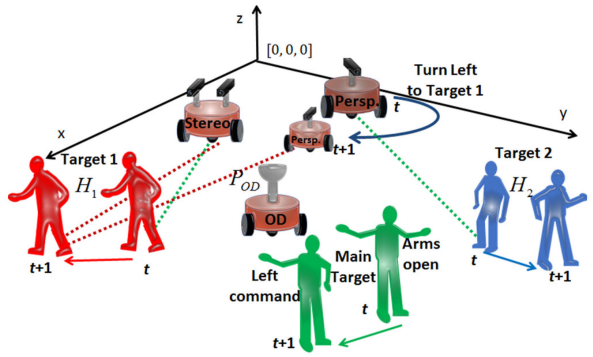


Fig. 9. Command cognition space illustrates the human–robot interaction while the main target commands the robots to focus on specific target.

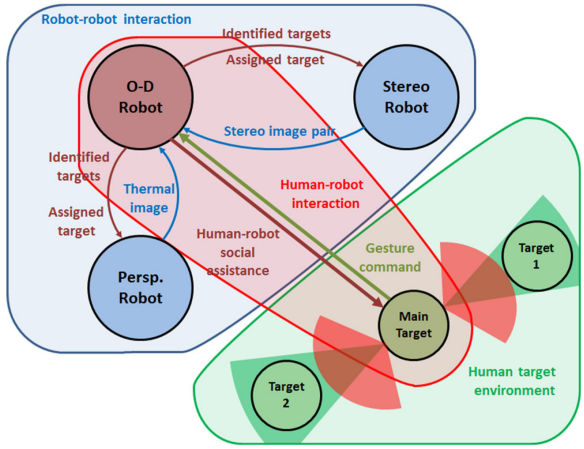


Fig. 10. Map of the command cognition process to inform main target about the surroundings. O-D robot collects the images to recognize the commands and identifies the targets. The main target is informed of surroundings and assisted by the collaboration.

cognition dictionary to command the robots with the corresponding gesture and posture

$$S_p(c) = \sum_{c=1}^{Wb} \sum_{r=1}^{Hb} p(x, y)$$

$$S_p(r) = \sum_{r=1}^{Hb} \sum_{c=1}^{Wb} p(x, y). \tag{8}$$

The algorithm to understand human target's commands from the O-D robot according to human gesture cognition process is given as follows.

The separated cells of target region helps to increase the processing time of every target region from multiple robots' perspective with respect to the gradient information-based HOG methods. This method also reduces the possible error caused by the noise in thermal images. Gesture cognition algorithm decides the command via correlation between the vertical and horizontal spectrum and specific gesture template after the cell filter allows to pass.

The human–machine environment is obtained by classification of the targets as main target and the targets which are desired to be tracked. The scenarios of cognition and interaction depend on the main target's commands by using key gesture. We utilize the target's arm movements as the commands. The target that opens both arms is assigned as the main target and the other targets are selected to be tracked. After the main target lifts the left or right arm, the robots track and follow the target on the side of the main target's command direction. Fig. 9 illustrates the human–robot environment with three targets and three robots. Main target is selected at time t by detecting the target's arms both opened in the images. The robots were tracking different targets to cover larger area by utilizing O-D view and stereo-perspective thermal view. The targets are mapped in 3-D space from the multisensory visual perception. An O-D robot detected the main target's command at time $t + 1$ and commands the other robots to focus on the target on its left as part of command cognition. The stereo robot was already tracking the target 1 but the robot equipped with single perspective thermal camera changed its direction to the interested target as part of human–machine collaboration mission. This assignment is done by the gesture command of the human, while O-D robot decides the selection of the interested target. The task can be changed by assigning different interactions from recognized commands from robots.

Algorithm 3: Gesture Cognition.

Input: Label of each target L_T^h
Bird's-eye view, $S_p^h(c)$, and perspective view, $S_p^h(r)$, spectrum vectors
Output: Gesture Cognition G_T^h
$L_T^h \leftarrow$ *identify*(C_K, C_V) // L_T^h is labeled targets
for all L_T^h (identified targets) *do*
$C_S \leftarrow$ *corr*$(S_p^h(c), S_p^h(r))$ // C_S is correlation result for bird's-eye view and perspective view vectors
$G_T^h \leftarrow$ *dictionary*(C_S) // G_T^h is matched gesture
end

Algorithm 4: Command Cognition.

Input: Gesture command G_T^h
Bird's-eye view, $S_p^h(c)$, and perspective view, $S_p^h(r)$, spectrum vectors
Output: Assignment of targets; main target M_T, main target's interest I_T^h, other targets O_T^h
Command from O-D robot, S_{OD}, to assign the closest robot S_C, to I_T^h
if both arms $\leftarrow$ *evaluate*(G_T^h)
$M_T \leftarrow$ *assign*(S_C) // M_T is main target
if left or right arm $\leftarrow$ *evaluate*(G_T^h)
$S_C \leftarrow$ *command*(M_T)
$I_T^h \leftarrow$ *assign*(S_C)
end
end

The process of the human–robot command cognition for the collaboration is shown in Fig. 10. O-D robot collects the thermal images from stereo and perspective robots as inputs. The main target commands via gesture signal and O-D robot recognizes these commands. In order to assist the main target as part of target's need in a human–robot environment, the output is sent to the robots as processed thermal view with identified targets and assigned targets for each robot along with the positioning

TABLE II
DATASETS

Imaging	Number of Targets	Covered Area (degree/m^2)	Total Frame Size (images)
O-D Lab Room	3	360 / 62.40	6000
O-D Hallway	3	360 / 122.88	1000
Stereo Lab Room	2-3	60 / 24.31	12000
Stereo Hallway	2	60 / 46.80	2000
Single Persp. Lab Room	1-2	50 / 20.26	6000
Single Persp. Hallway	1	50 / 76.05	1000

in a 3-D map. The main human target is also notified for the surroundings while we use a 3-D map with the positions of robots and targets.

The algorithm of the command cognition for interaction between human–robot and robot–robot shows that the gesture command assigns the main target when both the arms are raised. The main target is watched for the second command to appoint the interested target to be tracked. The gesture command assigns the closest robot S_C to the interested target I_T^h. Then the robots collaborate to follow the humans by interaction between humans and robots.

IV. EXPERIMENTS

The experimental results are organized into four sections. First, human kinematics analysis is discussed in Section IV-A. Second, human trajectory analysis is discussed in Section IV-B. Then, the human targets are identified in the multiple images in Section IV-C. Lastly, in Section IV-D the gesture commands from a human are analyzed and the robots assist the human target in its environment.

A. Human Kinematics Analysis

Human target regions are detected and analyzed in this section. The thermal images from O-D sensor, stereo sensor, and the perspective single sensor are converted to binary images with a threshold pixel value that corresponds to human body temperature. Table II shows the detail about the imaging conditions corresponding to different environment and sensor with the number of targets, covered angle/area, and total number of images.

The connected components in the images are obtained for analysis to find the regions which have human body features such as the size of region and size ratio of human body. The final regions that correspond to human body features are separated into three parts: head, arm, and leg regions. PCA analysis is applied to every part of each target region to determine the orientation of human head, arm region, and legs. PCA provides the angle of the interested region along with the major and minor axes. These components help us to decide the movement tendency of the human target.

The process of obtaining orientations of the human body parts is shown in Fig. 11. First, a gray level image is received from the sensors. Then, the gray level images are converted to binary and based on this binary conversion, a target region is detected. The target region of interest is separated into its parts and the PCA draws circles to every part of the body. In order to find the human body parts, we used the vertical spectrum of the target regions. This spectrum consists of the sum of the pixel values of every row in human target region, plotted in Fig. 11. Local minimum method is applied to find the neck region of the human body

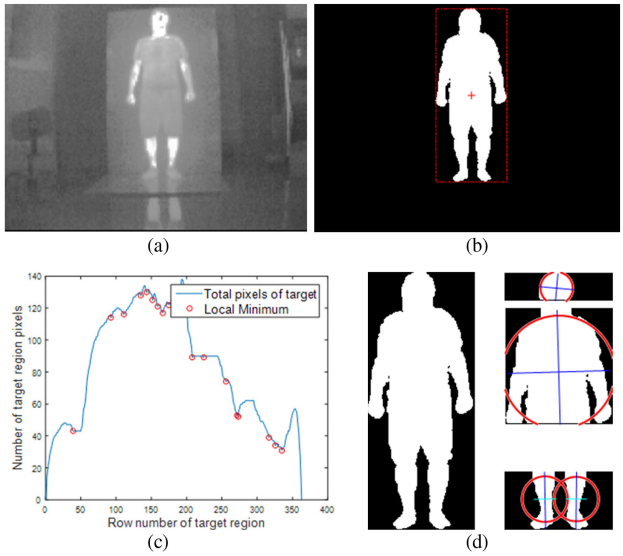


Fig. 11. Target body kinematics. (a) Original thermal image. (b) Detected target region. (c) Number of pixels that covers target region is calculated for every row in the target region plot, and the first local minimum is used to decide head part of the target region. (d) Head, arm, and leg regions are separated, and the orientation of the head, arm, and leg regions is used to obtain the kinematics of the regions.

TABLE III
IDENTIFICATION OF HUMAN TARGETS

Method	Target Trajectory Estimation Error (%)	Target Kinematics Estimation Error (%)	Target Identification Error (%)
O-D Sensor	11.50	22.23	15.63
Stereo Sensor	13.46	11.42	9.38
Perspective Single	7.69	11.12	18.75
Final decision of the sensors	11.53	13.88	6.25

to separate the head part. The first local minimum corresponds to the coordinate of the human neck and provides the border between human head and body regions. Then, the remaining part of the image is divided into four equal pieces horizontally and the first two pieces give the arm region, while the last piece gives the leg regions. These regions shown in circles are also provided with the orientation properties individually. The target's movement tendencies are obtained depending on the orientation of these components, the upper body and the average angle of the legs. If the angles were larger than 75°, the horizontal movement of the corresponding targets was decided. After the analysis of the target kinematics, we labeled them with five different classes with respect to target's movement tendency. Target can be moved to the right or left. Target can move away from the main robot or move closer to the robot. The last label is a stable target (right-left, forward-backward directions). Target may be labeled with two of these labels such as moving away and moving to the right or moving closer and moving to the left. The accuracy is compared from the images and given in Table III. O-D sensor provided 22.23% error, while the stereo and single perspective sensors gave 11.42% and 11.12% error, respectively. The error and accuracy are calculated by using 28 000 images, details are given in Table II, actual and estimated

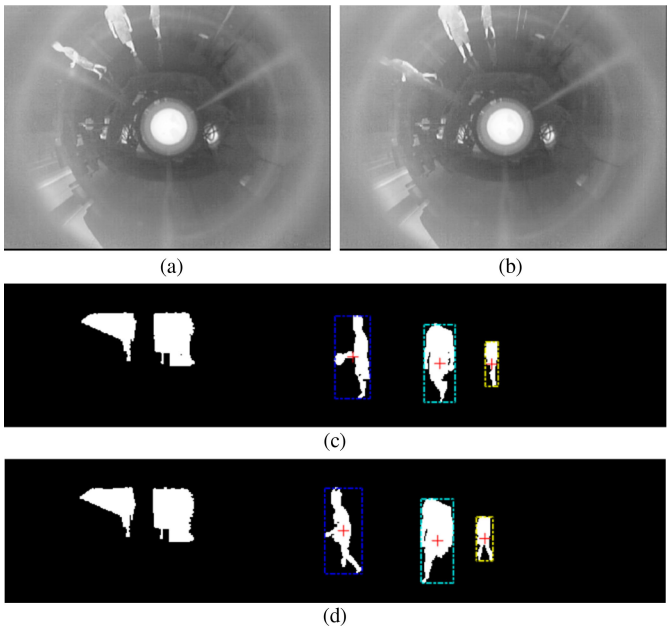


Fig. 12. Target region boxes were tracked in the latest image sequence to have an initial estimation of target trajectories. (a) Original O-D image. (b) Consecutive image. (c) Three targets are detected in unwrapped binary version of first image. (d) Detected targets in unwrapped binary version of consecutive image.

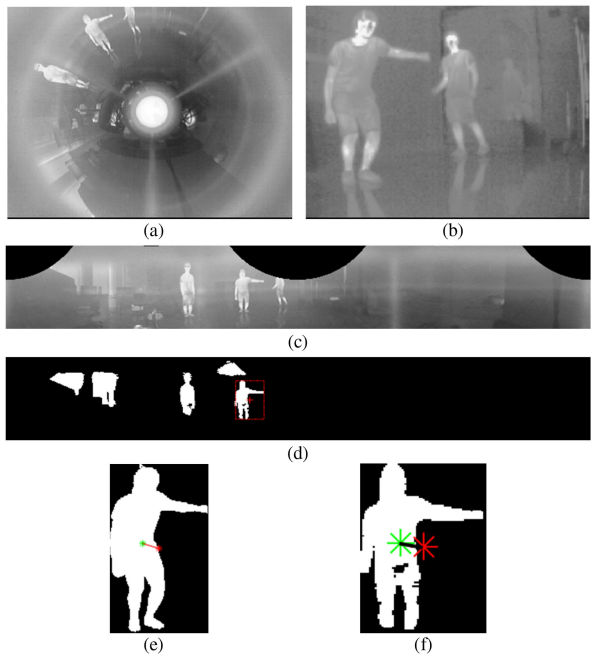


Fig. 13. Initial decision of target's gesture command from the center of target box and center of the region mass in O-D and stereo images. (a) Original O-D thermal image. (b) Original thermal image from stereo sensor. (c) Unwrapped O-D thermal image. (d) Detected target in unwrapped binary O-D image. (e) Initial gesture estimation in the image from stereo. (f) Initial gesture estimation from in the O-D image.

target movement tendencies are compared to obtain the wrong estimations. The reason for high error with O-D sensor was the small size of the targets in the images and sudden changes. Stereo and single perspective sensors offered an advantage to reduce this error with the additional views.

B. Human Trajectory Analysis

The target regions were detected and separate boxes were fitted to these regions. The target boxes were tracked in consecutive image frames from every sensor. Two consecutive O-D thermal images are shown in Fig. 12(a) and (b). The targets were detected and shown in the unwrapped binary images for both images [see Fig. 12(c) and (d)]. The change of the tracked boxes can be seen for each thermal sensor in Fig. 12(c) and (d). Two different positions of the target box provided the size, the ratio of height and width, and horizontal movement of the box in the image to create the vectors for each target. After obtaining vectors of the target boxes from each point of view, the boxes were matched on every image by labeling them.

The labeled boxes had a final vector on the plane by adding all matched vectors from every point of view. The magnitude and the angles of the final vectors are used to decide the direction of the targets. The error was calculated from the difference between the final vector and each sensor's estimation by using 28 000 images in the datasets. From Table III, we can see that perspective single sensor gave a better result by 7.69% error, while the stereo provided 13.46% error for the final target's direction, which were obtained from the magnitude and the angle accuracies. O-D sensor offers 11.50% error for the decision of the target's direction while increasing the possible number of targets in a wider FOV.

C. Identification of Human Targets

The acquired kinematics of the targets and the initial trajectories were utilized from every image, and initial target's gesture estimation assisted them to identify the human targets from each sensors. The initial gesture command used two criteria. The first, the target region components were extracted for every image. The original O-D and single image from stereo thermal sensor are shown on the top in Fig. 13. O-D images were converted to panoramic images and the targets were detected by using the human body characteristics such as the width/height ratio. We used 0.20 for the lower ratio threshold, while the higher threshold was 0.75. The detected target regions can be seen in Fig. 13. The binary target regions were bounded with a box and the center of this box compared with the center of the target region mass. We obtained a vector from the center of the mass to the center of the box. The direction and magnitude of this vector provided the possible gesture comment to identify targets. The center of the region mass is shown with green star and the center of the target box is shown with a red star in Fig. 13.

The targets from four different perspective, stereo sensors, single sensor, and O-D sensor, were detected and selected with the corresponding color to the identity of the target. The final decision of the target trajectory from every sensor considered as ground truth and error of each sensor were calculated with respect to this ground truth. The accuracy of the target kinematics were obtained from the difference between the kinematics estimation and final decision of target trajectory direction. The initial gesture estimation of the target also associated with the kinematics to match gesture and center body kinematics. A gesture command changed the angle of the arm region kinematics

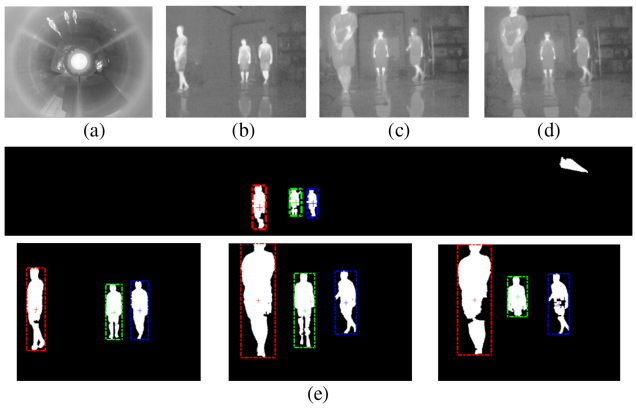


Fig. 14. Identified targets from (a) original O-D image, (b) original image from perspective single sensor, (c) right image from stereo sensor, (d) left image from stereo sensor, (e) identified targets from O-D sensors, perspective single, stereo right, and stereo left, respectively.

according to the direction of the raised arm. To avoid this angle change caused by gesture command, we set up a 10% threshold of initial gesture vector. If the vector is larger than this threshold, the angle of upper body kinematics was considered as not giving movement information. The accuracy from each sensor is given in Table III with respect to each sensor's trajectory and kinematic analysis after utilizing the 28 000 images in the datasets. The targets were selected with the specific color with respect to the corresponding target from every view, shown in Fig. 14. The stereo sensor gave the minimum error for the identification of the targets by 9.38% after utilizing trajectory analysis and kinematic analysis. These two analysis reduced the identification error when they were used together. The stereo sensors utilized the advantage of close image pair with a consistent view and improved the O-D robot's view and decision by increasing the accuracy. O-D sensor had higher error than the error with only trajectory analysis but provided lower identification error than the error with the kinematics analysis only. Final target identification from every sensor decreased the error around 50% from trajectory or kinematics analysis only and provided 6.25% error for identification of targets.

D. Command Cognition for Human–Robot Interaction

The robots were commanded by use of gesture controls from the main target which was identified in the previous step. The human target region was analyzed by using the vertical and horizontal target region spectrum for the decision of the commands. The variations of possible gesture commands from human target are shown in Fig. 15. First, a map was constructed by using five horizontal and five vertical cells in the box of target region. This provided a separate head region from arm region with the next two rows and the leg region with the last two rows in order to analyze these regions separately. The cell was turned ON if 1/6 of the cell had part of target, shown in Fig. 16. Fig. 16(a) illustrates the target when the arms stands still and do not show any direction. The map of the same target showed a more compact white cells. Fig. 16(b) shows the target with the raised left arm, while the map has four cells on the left in the second and third rows. After the target region passed from this filter, a spectrum was obtained from each rows and columns.

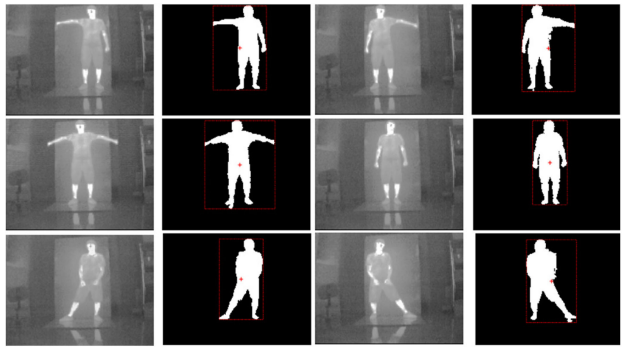


Fig. 15. Human gesture variations. Original images are on the left, detected target regions on the right.

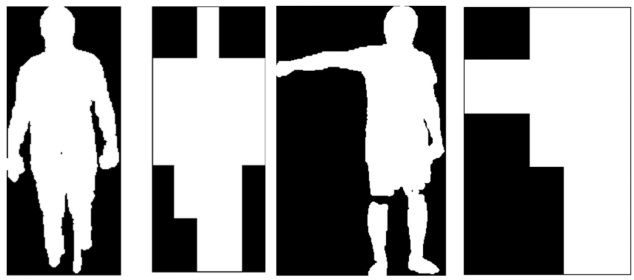


Fig. 16. Human target region is separated into cells and calculated cells to read the command which matches with the predefined commands.

Fig. 17 shows the sum of the pixel corresponding to the target for every rows in the detected box. The first local minimum provided us the neck region between head and body. The local minimum for the neck is shown with a green bold circle in Fig. 17. The remaining part of the plot provided the arms and legs. The change of the human body silhouette gives information to decide the commands from the main target. If we consider two-dimensional change in the silhouette, it provides unique signatures from the spectrum. While the main target raised an arm, a peak occurred in the spectrum right after the head region, shown in Fig. 17(a)–(c). If both arms were raised, the difference between the peak and right side of the peak increased since the arm regions moved from body to shoulder region. We divided the after head part to four equal pieces and analyzed the first two to decide if both arms were raised or only one. Once we decide it is only one arm raised, we used the horizontal spectrum to decide if the low pixel values were on the left or right side on the spectrum, shown in Fig. 18(a) and (b). Figs. 17(a) and 18(a) matches ensured the left command, and Figs. 17(b) and 18(b) ensured the right command. Figs. 17(c) and 18(c) show the raising of both arms since the horizontal spectrum had low pixel regions on both sides. Figs. 17(d) and 18(d) show a more grouped pattern and this represents a still standing human body. The horizontal spectrums were analyzed as three pieces so that the difference between these pieces could be more informative about the status of the arms. In order to show the effect of using legs to point left and right or opening both legs, we used vertical spectrums, [see Fig. 17(e) and (f)] and horizontal spectrums [see Fig. 18(e) and (f)]. Since the total pixel value of leg regions did not change for the target region rows, the spectrum appeared similar to the still standing. Leg movements did not give any

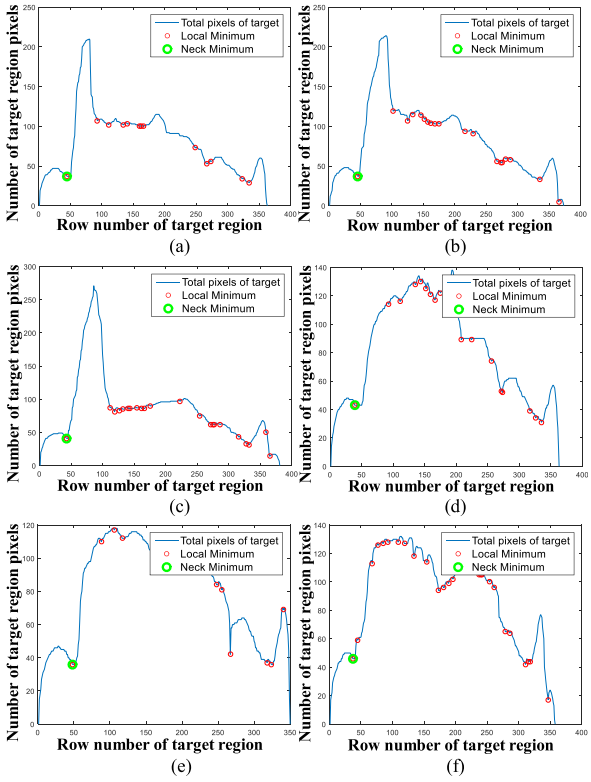


Fig. 17. Number of pixels that covers target region is calculated for every row in the target region. The first local minima is used to decide head part of the target region. (a) Left arm is raised, (b) right arm is raised, (c) both arms are spread, (d) both arms are down, (e) left leg is shown, (f) right leg is shown.

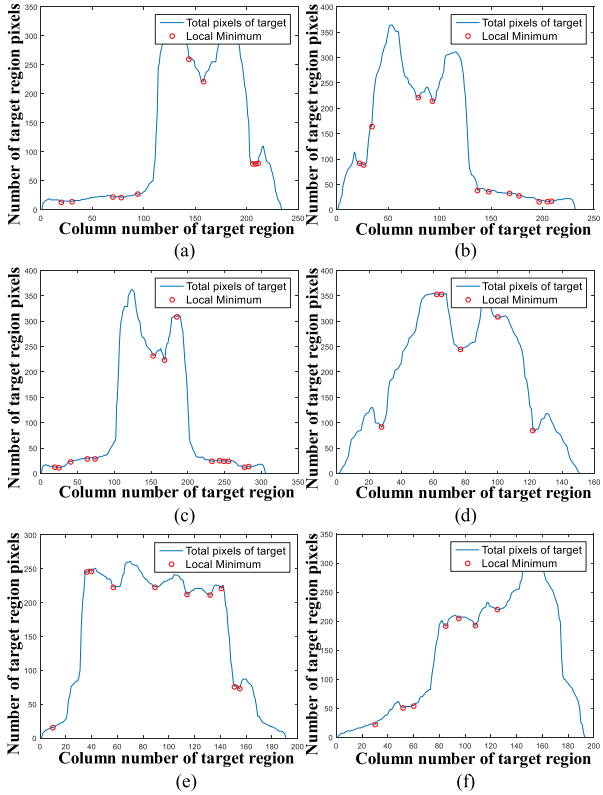


Fig. 18. Number of pixels that covers target region is calculated for every column in the target region. The first and last local minima are used to decide arm parts of the target region. The length of these regions helps to decide gesture command. (a) Left arm is raised, (b) right arm is raised, (c) both arms are spread, (d) both arms are down, (e) left leg is shown, and (f) right leg is shown.

TABLE IV
GESTURE ANALYSIS FOR COMMAND COGNITION

Command	Left	Right	Spread	Standing	Error (%)
Lab Room					
Left	**98.50**	1.33	0.60	1.06	1.50
Right	0.00	**94.00**	2.41	5.30	6.00
Spread	0.80	3.00	**96.99**	4.95	3.01
Standing	0.70	1.67	0.00	**89.04**	10.96
Hallway					
Left	**98.36**	0.00	2.67	5.68	1.63
Right	0.00	**90.56**	12.00	3.41	9.43
Spread	1.63	0.00	**84.00**	0	16.00
Standing	0.00	9.43	1.33	**90.91**	9.09

The percentage (%) of the detected gesture commands are given in the table.
The bold values are the accuracy of correct classification for each gesture command.

confusing pattern for the sum of the columns in the horizontal spectrum as well, shown in Fig. 18(e) and (f).

Two different datasets, from the lab room and the hallway, were used for gesture analysis experiments. Since the thermal imager captures only heat from the objects, the lighting condition of the dataset environment did not affect the thermal images. The images were taken when there was no light source in the indoor environment. The decision of the main target and the selection of interested target is shown in Fig. 19, while the green human is the main target and the red human is interested target depending on the command of the main target. Table IV shows the accuracy percentage of the detected commands from vertical, horizontal spectrums combination for the lab room and the hallway datasets. In case of using lab room images, the accuracy was 98.50% for the left gesture dataset, 985 images out of 1000 images were detected succesfully, and 94.00% for right gesture command. Horizontal spectrum helped to increase the accuracy for the detection of the spread arms command up to 96.99 and 89.04% for the standing posture. The decision of all robots gave 3.01% error to detect the main target, while the connection strength depends on the distance and decision accuracy. Then, the O-D robot commands the final result of combination to the related robot with 1.50% error for detection of left target and 6.00% error for detection of the right target. The hall way dataset gave a high accuracy with 1.63% error for the left gesture command and 9.43% error for the right gesture command. Since the confusion of right command was with standing movement without any command, this does not cause any problem for the manipulation of the robots. Both datasets provided a high accuracy for making the critical decision of selection of main target with spread arm gesture command. 0.80% confusion for the right command 3.00% for the left command with the spread arm gesture command in the lab room dataset. The main target communicated with the robot and commanded the direction with 96.99% accuracy in the wide lab room conditions. Accuracy of the target identification for every individual robots in hallway dataset is shown in Fig. 20(a), and lab room dataset with the common decision as a final result is given in Fig. 20(b). Stereo and perspective single sensors

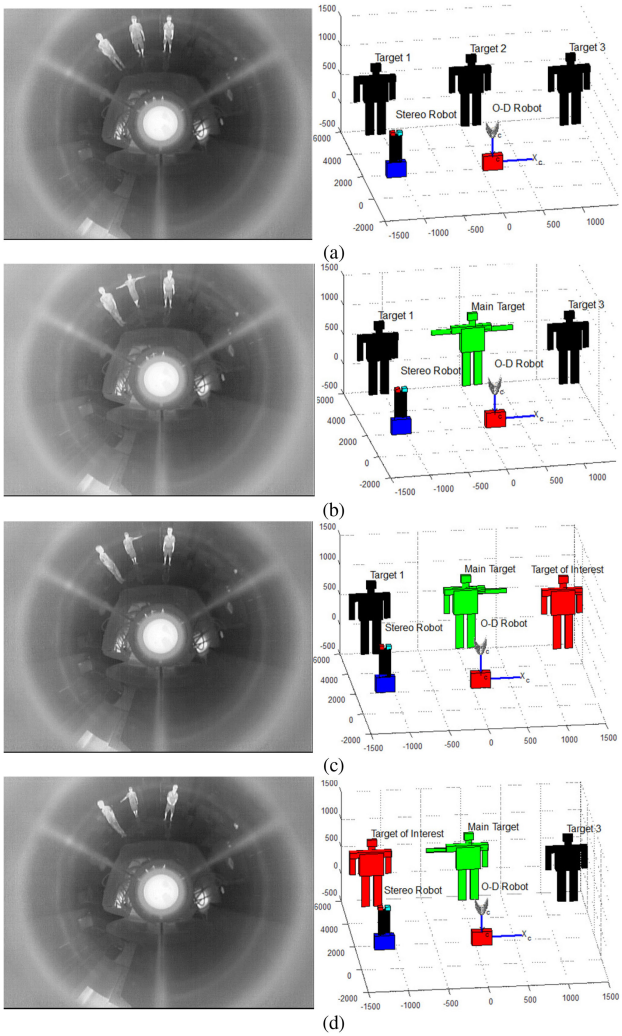


Fig. 19. Classification of the targets from gesture command cognition. (a) Gesture command does not present, (b) both arms are spread, main target is selected, (c) right arm is raised, the target on the right is selected as target of interest, (d) left arm is raised, the target on the left is selected as target of interest.

are shown in blue circles with blue links to targets in their FOV. O-D sensor is shown in red circle with the red links to the targets to show the human–robot interaction. Our proposed method provided a wide area, from 20.26 to 122.88 m^2 to understand gesture commands for human–robot interaction. It can be seen in Fig. 20 that our sensors offered 360° FOV from 2.0 to 15.0 m of visual perception for interaction space, while the work in [30] offers from 1.7 to 4 m with maximum 3.35 m^2 interaction area. The final target identification provided 93.75% accuracy. The accuracy of gesture command decision is also given along with the average accuracy of lab room and hallway datasets, 98.43% for left, 92.28% for right, and 90.49% for the main target. AlexNet method has 76.96% accuracy, a deep learning-based method using VGG16 has 93.07% accuracy while utilizing 55 000 images for training [33], and a deep convolutional neural network (DCNN) method utilized 9360 images and 94.57% accuracy [34]. Our method had 93.75% accuracy for gesture command recognition without a training process and training data while deep learning-based methods require a training process.

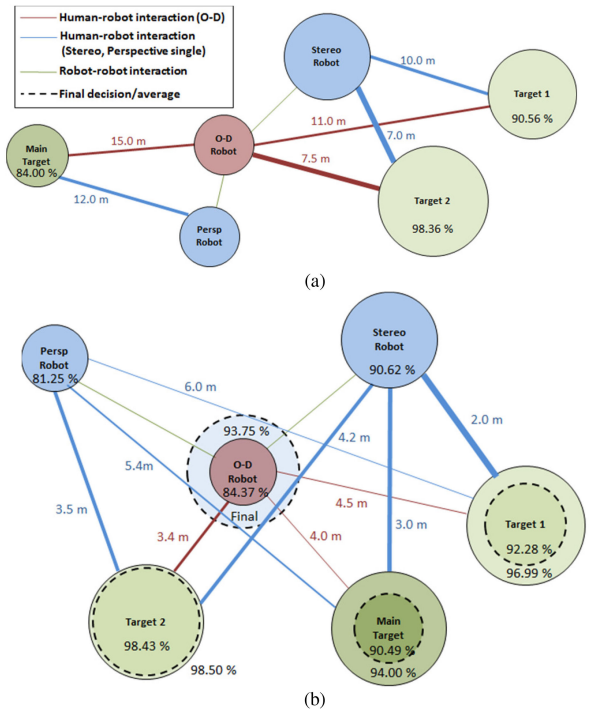


Fig. 20. Human–robot interaction with the accuracy of target identification and gesture command detection. (a) Hallway dataset and (b) lab room dataset with the final decision and the average accuracy gesture detection. Gesture command accuracy for the main target, left target, and the right target is shown with Target 1 as target on the right, and Target 2 as target on the left.

V. Conclusion and Future Work

We proposed human–robot and robot–robot interaction based command cognition by using visual perception. Our method identified the targets from multiple sensors and assisted the human target according the target's command. The gesture commands are acquired from identified main target in multiple images. The commands from the main target were interpreted by means of the human–robot interaction, and the other robots were commanded by the O-D robot. 360° thermal view of the O-D robot helps to collaborate multiple robots from any perspective. The fusion of O-D, stereo, and single thermal camera improved the visual perception to track human targets in wider FOV and under low lighting environments. O-D sensor provided an advantage to create multiview perception to any direction with the other sensors.

We plan to utilize the identified multiple human targets and the gesture recognition to predict targets' next moves. This will allow us to track targets more precisely and estimate their behavior trends.

References

[1] E. Benli, Y. Motai, and J. Rogers, "Human behavior-based target tracking with an omni-directional thermal camera," *IEEE Trans. Cogn. Develop. Syst.*, vol. 11, no. 1, pp. 36–50, Mar. 2019.

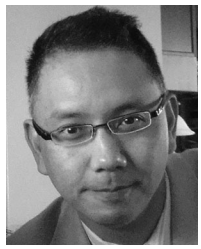
[2] D. Cavaliere, S. Senatore, and V. Loia, "Proactive UAVs for cognitive contextual awareness," *IEEE Syst. J.*, vol. 13, no. 3, pp. 3568–3579, Sep. 2019.

[3] Z. Ju, X. Ji, J. Li, and H. Liu, "An Integrative framework of human hand gesture segmentation for human–robot interaction," *IEEE Syst. J.*, vol. 11, no. 3, pp. 1326–1336, Sep. 2017.

[4] C. Hwang, B. Chen, H. Syu, C. Wang, and M. Karkoub, "Humanoid robot's visual imitation of 3-D motion of a human subject using neural-network-based inverse kinematics," *IEEE Syst. J.*, vol. 10, no. 2, pp. 685–696, Jun. 2016.
[5] M. Gupta, S. Kumar, L. Behera, and V. K. Subramanian, "A novel vision-based tracking algorithm for a human-following mobile robot," *IEEE Trans. Syst. Man Cybern., Syst.*, vol. 47, no. 7, pp. 1415–1427, Jul. 2017.
[6] S. Gaglio, G. L. Re, and M. Morana, "Human activity recognition process using 3-D posture data," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 5, pp. 586–597, Oct. 2015.
[7] M. Ye, X. Lan, Z. Wang, and P. C. Yuen, "Bi-directional center-constrained top-ranking for visible thermal person re-identification," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 407–419, Jun. 2019.
[8] L. Wang, T. Tan, H. Ning, and W. Hu, "Silhouette analysis-based gait recognition for human identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1505–1518, Dec. 2003.
[9] C. Bauckhage, J. K. Tsotsos, and F. E. Bunn, "Automatic detection of abnormal gait," *Image Vis. Comput.*, vol. 27, no. 1, pp. 108–115, 2009.
[10] J. Cashbaugh and C. Kitts, "Vision-based object tracking using an optimally positioned cluster of mobile tracking stations," *IEEE Syst. J.*, vol. 12, no. 2, pp. 1423–1434, Jun. 2018.
[11] C. A. Cifuentes, C. Rodriguez, A. Frizera-Neto, T. F. Bastos-Filho, and R. Carelli, "Multimodal human–robot interaction for walker-assisted Gait," *IEEE Syst. J.*, vol. 10, no. 3, pp. 933–943, Sep. 2016.
[12] G. Batchuluun, H. S. Yoon, J. K. Kang, and K. R. Park, "Gait-based human identification by combining shallow convolutional neural network-stacked long short-term memory and deep convolutional neural network," *IEEE Access*, vol. 6, pp. 63164–63186, 2018.
[13] Y. Hu, Z. Li, G. Li, P. Yuan, C. Yang, and R. Song, "Development of sensory-motor fusion-based manipulation and grasping control for a robotic hand-eye system," *IEEE Trans. Syst. Man Cybern., Syst.*, vol. 47, no. 7, pp. 1169–1180, Jul. 2017.
[14] X. Qu, D. Zhang, G. Lu, and Z. Guo, "Door knob hand recognition system," *IEEE Trans. Syst. Man Cybern., Syst.*, vol. 47, no. 11, pp. 2870–2881, Nov. 2017.
[15] G. Caron, E. M. Mouaddib, and E. Marchand, "3D model based tracking for omnidirectional vision: A new spherical approach," *Robot. Auton. Syst.*, vol. 60, no. 8, pp. 1056–1068, Aug. 2012.
[16] G. L. Mariottini, S. Scheggi, F. Morbidi, and D. Prattichizzo, "An accurate and robust visual-compass algorithm for robot-mounted omnidirectional cameras," *Robot. Auton. Syst.*, vol. 60, no. 9, pp. 1179–1190, Sep. 2012.
[17] E. Bauzano, B. Estebanez, I. Garcia-Morales, and V. F. Muñoz, "Collaborative human–robot system for HALS suture procedures," *IEEE Syst. J.*, vol. 10, no. 3, pp. 957–966, Sep. 2016.
[18] T. Tung, R. Gomez, T. Kawahara, and T. Matsuyama, "Multiparty interaction understanding using smart multimodal digital signage," *IEEE Trans. Human-Mach. Syst.*, vol. 44, no. 5, pp. 625–637, Oct. 2014.
[19] A. Taniguchi, T. Taniguchi, and I. Tetsunari, "Spatial concept acquisition for a mobile robot that integrates self-localization and unsupervised word discovery from spoken sentences," *IEEE Trans. Cogn. Develop. Syst.*, vol. 8, no. 4, pp. 285–297, Dec. 2016.
[20] M. Mohandes, M. Deriche, and J. Liu, "Image-based and sensor-based approaches to Arabic sign language recognition," *IEEE Trans. Human-Mach. Syst.*, vol. 44, no. 4, pp. 551–557, Aug. 2014.
[21] T. Taniguchi, S. Nagasaka, and R. Nakashima, "Nonparametric bayesian double articulation analyzer for direct language acquisition from continuous speech signals," *IEEE Trans. Cogn. Develop. Syst.*, vol. 8, no. 3, pp. 171–185, Sep. 2016.
[22] H. Cai, B. Liu, J. Zhang, S. Chen, and H. Liu, "Visual focus of attention estimation using eye center localization," *IEEE Syst. J.*, vol. 11, no. 3, pp. 1320–1325, Sep. 2017.
[23] N. M. Moacdieh and N. Sarter, "Using eye tracking to detect the effects of clutter on visual search in real time," *IEEE Trans. Human-Mach. Syst.*, vol. 47, no. 6, pp. 896–902, Dec. 2017.
[24] C. Castellini, T. Tommasi, N. Noceti, F. Odone, and B. Caputo, "Using object affordances to improve object recognition," *IEEE Trans. Auton. Mental Develop.*, vol. 3, no. 3, pp. 207–215, Sep. 2011.
[25] A. Sciutti, L. Patanè, F. Nori, and G. Sandini, "Understanding object weight from human and humanoid lifting actions," *IEEE Trans. Auton. Mental Develop.*, vol. 6, no. 2, pp. 80–92, Jun. 2014.
[26] M. U. S. Khan *et al.*, "On the correlation of sensor location and human activity recognition in body area networks (BANs)," *IEEE Syst. J.*, vol. 12, no. 1, pp. 82–91, Mar. 2018.
[27] N. Hu, G. Englebienne, Z. Lou, and B. Krose, "Learning to recognize human activities using soft labels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 10, pp. 1973–1984, Oct. 2017.
[28] A. Aladrén, G. López-Nicolás, L. Puig, and J. J. Guerrero, "Navigation assistance for the visually impaired using RGB-D sensor with range expansion," *IEEE Syst. J.*, vol. 10, no. 3, pp. 922–932, Sep. 2016.
[29] P. Wei, Y. Zhao, N. Zheng, and S. C. Zhu, "Modeling 4D human-object interactions for joint event segmentation, recognition, and object localization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1165–1179, 1 Jun. 2017.
[30] A. Zaraki *et al.*, "Design and evaluation of a unique social perception system for human-robot interaction," *IEEE Trans. Cogn. Develop. Syst.*, vol. 9, no. 4, pp. 341–355, Dec. 2017.
[31] S. Qiu, Z. Li, W. He, L. Zhang, C. Yang, and C. Y. Su, "Brain-machine interface and visual compressive sensing-based teleoperation control of an exoskeleton robot," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 1, pp. 58–69, Feb. 2017.
[32] Y. Kawanishi *et al.*, "Voting-based hand-waving gesture spotting from a low-resolution far-infrared image sequence," in *Proc. IEEE Visual Commun. Image Process.*, 2018, pp. 1–4.
[33] S. Hussain, R. Saxena, X. Han, J. A. Khan, and H. Shin, "Hand gesture recognition using deep learning," in *Proc. Int. SoC Design Conf.*, 2017, pp. 48–49.
[34] M. R. Islam, U. K. Mitu, R. A. Bhuiyan, and J. Shin, "Hand gesture feature extraction using deep convolutional neural network for recognizing american sign language," in *Proc. 4th Int. Conf. Frontiers Signal Process.*, 2018, pp. 115–119.

Emrah Benli (M'17) received the B.Sc. degree in electronics and telecommunication engineering from Kocaeli University, Kocaeli, Turkey, in 2009, the M.Sc. degree in electrical and computer engineering from Clemson University, Clemson, SC, USA, in 2013, and the Ph.D. degree in electrical and computer engineering from Virginia Commonwealth University, Richmond, VA, USA, in 2017.

He worked as a Postdoctoral Fellow specializing in autonomous mobile robotics with the U.S. Army Research Laboratory's Computational and Information Sciences Directorate, Adelphi, MD, USA, in 2018. He is currently an Academician of electrical and electronics engineering with Gümüshane University, Gümüshane, Turkey. His research interests include intelligent systems, computer vision, artificial intelligence, multimodal sensory, robotic system design and control, and human–robot interaction.



Yuichi Motai (SM'12) received the B.Eng. degree in instrumentation engineering from Keio University, Tokyo, Japan, in 1991, the M.Eng. degree in applied systems science from Kyoto University, Kyoto, Japan, in 1993, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2002.

He is currently an Associate Professor of electrical and computer engineering at Virginia Commonwealth University, Richmond, VA, USA. His research interests include the broad area of sensory intelligence, particularly in medical imaging, pattern recognition, computer vision, and sensory-based robotics.

John Rogers (SM'17) received the B.S. and M.S. degrees in electrical and computer engineering from Carnegie Mellon University, Pittsburg, PA, USA, in 2002, the M.S. degree in computer science from Stanford University, Stanford, CA, USA, in 2006, and the Ph.D. degree from the Georgia Institute of Technology, Atlanta, GA, USA, in 2012.

He is a Research Scientist specializing in autonomous mobile robotics with the Army Research Laboratory's Computational and Information Sciences Directorate. His current research interests include automatic exploration and mapping of large-scale indoor and outdoor environments, place recognition in austere locations, and semantic scene understanding and probabilistic reasoning for autonomous mobile robots.