Thermal Multisensor Fusion for Collaborative Robotics

Emrah Benli, *Member, IEEE*, Richard Lee Spidalieri, and Yuichi Motai, *Senior Member, IEEE*

***Abstract*—Collaborative robotic configurations for monitoring and tracking human targets have attracted interest in the fourth industrial revolution. The fusion of different types of sensors embedded in collaborative robotic systems achieves high-quality information and contributes to significantly improve robotic perception. However, current methods have not deeply explored the capabilities of thermal multisensory configurations in human-oriented tasks. In this paper, we propose thermal multisensor fusion (TMF) for collaborative robots to overcome the limitations of stand-alone robots. Thermal vision helps to utilize the heat signature of the human body for human-oriented tracking. An omnidirectional (O-D) infrared (IR) sensor provides a wide field of view (FOV) to detect human targets, and Stereo IR helps determine the distance of the human target in the oriented direction. The fusion of O-D IR and Stereo IR also creates a multisensor stereo for an additional determination of the distance to the target. The fusion of thermal and O-D sensors brings their limited prediction accuracy with their advantages. The maximum *a posteriori* method is used to predict the distance of the target with high accuracy by using the distance results of TMF stereo from multiple platforms according to the reliability of the sensors rather than its usage of visible-band-based tracking methods. The proposed method tracks the distance calculation of each sensor instead of target trajectory tracking as in visible-band methods. We proved that TMF increases the perception of robots by offering a wide FOV and provides precise target localization for collaborative robots.**

***Index Terms*—Collaborative robotics, far infrared (IR) camera, mobile robot, multisensor, omnidirectional (O-D) camera, sensor fusion, stereo IR, target tracking, thermal vision.**

Manuscript received August 31, 2017; revised January 6, 2018 and September 17, 2018; accepted March 17, 2019. Date of publication April 1, 2019; date of current version July 3, 2019. 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. Paper no. TII-17-2046. *(Corresponding author: Yuichi Motai.)*

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

R. L. Spidalieri is with Electric Power, Inc., Chester, VA 23836 USA (e-mail: spidalierirl@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).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2019.2908626

I. Introduction

NETWORK-BASED robotics has been a center of interest with different configurations of sensors, and its importance is increasing as intelligent systems become more prevalent with the fourth industrial revolution. As sensors become easier to access, it is important to look beyond their standalone capabilities and explore the synergistic combinations between different types of sensors. Thermal images are used for human-oriented detection and tracking by using a human's thermal signature. In recent years, there has been a rapidly growing interest in using teams of mobile robots for autonomous systems. Collaborative systems are also increasingly being incorporated into commercial cars and vehicles. The systems equipped with multisensors enhance the results of prediction and tracking targets [1]–[3]. The common tracking methods use the visible-band camera sensor, which has a disadvantage of detecting targets in the absence of light. The infrared (IR) sensors provide light-independent detection of human targets with continuous tracking while the position is estimated by these sensors. There are however disadvantages since the resolution of these images and the information utilized from them are limited compared to visual sensors. Traditional perspective cameras are limited to a narrow field of view (FOV), which may cause it to lose the target easily. We use the term, perspective camera, to define the traditional directed cameras that do not use any increased FOV such as O-D spherical mirrors or fisheye lenses. O-D cameras utilize a wide FOV to maintain the targets within the robot's line of sight. O-D sensors offer limited prediction accuracy compared to perspective sensors due to a nonlinear structure of their mirror. A multisensor fusion of O-D, IR, and Stereo IR sensors creates a new multistereo view for target tracking available in a wide FOV for light-independent conditions for focusing on human targets. However, the fusion of a perspective sensor and an O-D sensor causes higher position estimation error in some cases.

The collaborative robots require more precise results for target positioning since the prediction of a target's position is a key element in developing the field of intelligent sensory perception. The common tracking and prediction methods have widely covered the particle filter method [4]–[7], Kalman filter method [8]–[10], and the maximum *a posteriori* (MAP) method [11]–[14]. The MAP method has an advantage of faster performance and higher accuracy than Kalman and particle filter methods. These methods utilize various sensors and platforms including perspective cameras, stereo systems, laser rangefinders, thermal sensors, and omnidirectional (O-D) cameras on mobile robots, unmanned ground, aerial, and underwater vehicles as well as

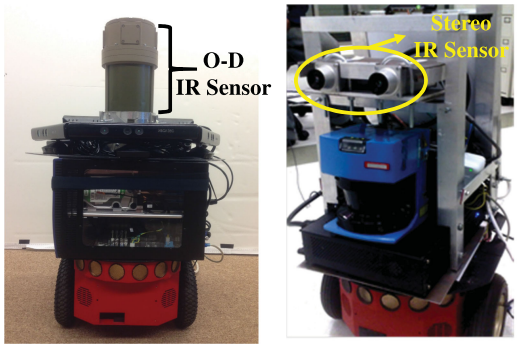


Fig. 1. Mobile robots equipped with O-D IR and Stereo IR sensors.

static platforms. The MAP method bolsters the prediction with respect to their reliability from the prediction history of sensors.

We propose the thermal multisensor fusion (TMF) method by using an O-D IR camera and a Stereo IR camera system mounted on the collaborative mobile robots for predicting and tracking the positions of targets by the MAP method for precise localization. The MAP-based estimation method is used to decide the reliability of the sensor fusion from their previous estimation states rather than using it to track a target, similar to visible-band MAP-based tracking methods. The conventional visible-band MAP-based tracking method tracks the human position along its trajectory and predicts the next position of the target. The proposed method tracks the sensors results and estimates the current target position. The multisensor fusion of thermal cameras, O-D, and perspective stereo strengthen the perceptive capabilities of the robot as well as increase the tracking duration in a large FOV with dynamic geometry for collaborative robots while the MAP method precisely predicts the final positions of targets based on the reliability of the sensors. Fig. 1 shows the experimental robots equipped with an O-D IR sensor and Stereo IR cameras.

This paper organization is given as follows. Section II presents the related works, followed by geometry of Stereo and O-D sensors in Section III, TMF for collaborative robotics in Section IV; the experiments are presented in Section V and finally, Section VI gives the conclusion.

II. Related

The relevant studies on target tracking via a 360° thermal imager mounted on a mobile robot are covered in this section. The cooperative human tracking methods based on different configurations of robots and target assignments in order to improve the tracking results are given in Table I. Collaborative tracking of human targets using mobile robots is an emerging field of science that can be implemented in both civilian and military applications. A single sensor/robot performs tracking of multiple targets in [7], [12], [14], and [15]. Standard perspective cameras are used as sensors in [12] and [14], whereas [7] and [15] use an O-D camera. Underwater target tracking is implemented with an O-D camera in [15]. The fusion of multiple sensors in [2]–[4], [8], [10], and [16]–[18] provides collaborative tracking examples. Multisensory process via Kalman [16] offers robust human tracking. In studies [19]–[21] multiple robots track

TABLE I
Configurations of Robots With Target Tracking Methods

Sensor Fusion	Camera Sensor	Method	Prediction Error
Rangefinder, Camera, Stereo [4]–[7]	IR or Color	Particle Filter	Moderate
Stereo, SFM [8]–[10]	Color	Kalman Filter	High
SFM, Rangefinder [11]–[15]	Color	MAP	Low

TABLE II
Comparison of Sensors

Sensor	FoV	Cost	Performance Indoor/Outdoor	Freq. Rate
Perspective Camera	50°	Low	Good/Moderate	30/sec+
O-D Thermal Camera	360°	High	Good/Good	30/sec
Kinect	45°	Low	Good/Poor	30/sec+
Lidar	180°	High	Moderate/Good	5/sec
RGB-D	70°	Low	Moderate/Good	30/sec+

multiple or single dynamic targets. Several sensors are used with a perspective camera, an ultrasound altimeter, an inertial measurement unit combination in [10], and a camera rangefinder combination in [4] and [8]. Another robust control method is applied as a multiple robots case consisting of unmanned aerial vehicles in [21] and multiple unmanned ground vehicles in [19] and [20]. The multisensory process requires a good evaluation for the sensor selection aimed at the specific missions during the hardware design. Table II compares the specifications of various sensors to show their advantages and disadvantages. Low-cost sensors such as perspective cameras, Kinect, and RGB-D has narrow FOV. Lidar and O-D help to increase the FOV while the cost increases. High-cost thermal sensors have an advantage of operation for a light-independent operation, whereas the low-cost Kinect has the disadantage of operating under direct sunlight. In the case, when the imaging frequency is more important, perspective camera Kinect and RGB-D have the advantage over high-cost and low-resolution thermal, O-D, and Lidar sensors. The proposed method utilizes the O-D thermal and Stereo IR from perspective sensors to increase the operation time and covered area with the minimum number of sensors.

Considerable attention is given to static sensors used in target tracking, and an emphasis is placed on the ability of perspective camera sensors to optimize the processing of available information. Numerous studies [1], [4]–[7], [22] have investigated the target localization problem using the particle filter, and then proposed the use of active target tracking for trajectory observation. The fusion of a perspective camera and a rangefinder is used in [4] and [22], and a combination of rangefinders is used in [5]. The stereo configuration [6] and the O-D camera configuration [1], [7] are implemented in the particle filter method to enhance the robots' perception. The Kalman filter is also a common method used for multiple target tracking in [2] and [8]–[10]. Observation based on distance and bearing for

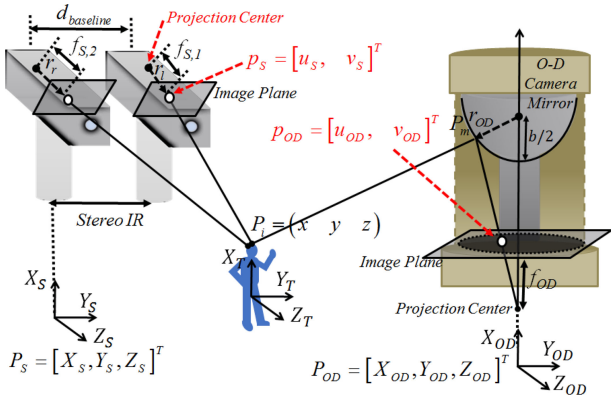


Fig. 2. Geometry of O-D IR and Stereo IR fusion.

active target tracking is investigated in [8] using target-position prediction and combination of the fuzzy compensation and Kalman Filter methods in [9]. Another Kalman filter based method with combinations of stereo and ultrasonic sensor is proposed in [10]. MAP-based methods are discussed in [11]–[14]. Huang *et al*. [11] introduces MAP-based cooperative localization and target tracking, [13] examines a MAP and particle filter combination for multiple targets, and [12] offers an approach to the batch-MAP estimation problem, whereas Choi *et al*. [14] uses an RGB-D sensor with MAP to track multiple people. By covering a larger area without increasing the number of sensors, the deployment of mobile O-D sensors with mobile platforms for tracking offers significant advantages over static and perspective sensors.

III. Geometry of Stereo and O-D Sensors

Multisensor camera systems utilize different configurations of multiple IR sensors. The first IR sensor is the O-D thermal imager, and the second sensor type is a traditional perspective camera with a thermal sensor. The O-D IR sensors inner structure is shown on the right-hand side of Fig. 2. The placement geometry of the IR stereo cameras is also given on the left-hand side of Fig. 2. Multisensor Stereo vision obtained from the O-D camera and one of the Stereo IR cameras is shown with the position coordinates of the cameras P_{OD}, P_S, and detected target point P_i in the target coordinate system P_T.

Our proposed multisensor fusion system has a more dynamic geometry in comparison to the fixed multisensor fusion of O-D perspective camera setups [18], [23]. The positions of the robots can be adjusted with respect to the positions of targets by using the transformation of each robot. The position of the stereo robot helps to maintain its ray vectors relative to the O-D robot. We will consider P_{OD} position as the origin of the three-dimensional (3-D) space and the reference for the perspective camera position P_S.

Two ray vectors from the camera projection centers to the image plane provide the direction of 3-D coordinates for the real P_i target point. The ray vector of the O-D camera is given by P_m mirror coordinates (1) of the corresponding feature point [24]

$$P_m = \begin{bmatrix} x_m & y_m & 1 \end{bmatrix}^T = R_c^T K_{\mathrm{OD}}^{-1} p_{\mathrm{OD}}. \tag{1}$$

The camera's intrinsic matrix is K_{OD}^{-1} and the transformation matrix between the camera and mirror is R_c^T. If we assume the distance between the mirror focal point and the mirror vertex is $b/2$, the third coordinate of the mirror point can be calculated as $z_m = ((x_m^2 + y_m^2)/2b) - (b/2)$. The mirror coordinates of the feature point will give us the first ray vector. The starting point of the first ray vector, $r_{\mathrm{OD}} = [\, x_m \;\; y_m \;\; z_m \,]$, is the coordinates of the O-D robot, $P_{\mathrm{OD}} = [X_{\mathrm{OD}}, Y_{\mathrm{OD}}, Z_{\mathrm{OD}}]^T$; however, we will change the X_{OD} with the mirror's focal point coordinate. Then, the second ray vector, $r_l = [\, u_s \;\; v_s \;\; f_{s,1} \,]$, from the perspective camera will be given from the projection center of the perspective camera, $f_{S,1}$, to image point, $p_S = [\, u_S \;\; v_S \,]^T$, on the image plane of the perspective camera. The position of the O-D robot is used as the reference coordinate for the Stereo IR robot. The rotation, R_S, and the translation, t_S, of the Stereo IR robot helps to obtain the orientation, $r_l = [\, u_s \;\; v_s \;\; f_{s,1} \,]$, with respect to the OD robot's position. The multisensor calculation utilizes the updated ray vector, r_l^{multi} given by (2), from the left stereo camera

$$r_l^{\mathrm{multi}} = R_S \begin{bmatrix} u_s & v_s & f_{s,1} \end{bmatrix}^T + t_S. \tag{2}$$

The transformation between the O-D IR robot and Stereo IR robot is given by the rotation and translation obtained from the current position of the robots. In the case of multiplication by the ray vector for the perspective camera, it will give the correct direction of the ray vector with respect to the O-D IR robot's coordinate space. In case any change in the position of the robots, the positions are read from the robots' odometers, and the rotation and translation vectors are updated depending on the possible movement of robots. Then, the ray vectors from each sensor are updated by translation and rotation of the robots respectively. This makes our algorithm feasible for static and dynamic sensors.

The position of robots is chosen to be static in order to evaluate the accuracy of our estimation method. Any possible error of the robot position would affect our target position estimation result and the accuracy would be unreliable to evaluate the method. For this reason, the evaluation process is done by a static robot position. Our MAP-based estimation method overcomes the position error of any robot in a later process and generates higher target position estimation depending on the more reliable robot.

Stereo IR uses two images, left and right, from two IR cameras. The images from both cameras provide the feature points of the targets in the robot's FOV. The feature point-matching algorithm returns the corresponding feature points in both images. Every matched feature point is used in the 3-D coordinate calculation process. Since we know the distance between the two cameras and their intrinsic and extrinsic parameters, they provide the ability to calculate the distances of the feature points from the robot. Because of their ability to detect the human body's heat signature, mounted IR sensors provide the advantage of precisely tracking human targets under low lighting conditions. This means that the robot's enhanced detection of a target's shape and distance can also be extended into night-time conditions. Equation (3) calculates the target's coordinates and its distance d from Stereo IR

$$d_{1,z} = d_c d_{\mathrm{baseline}} / (P_l - P_r). \tag{3}$$

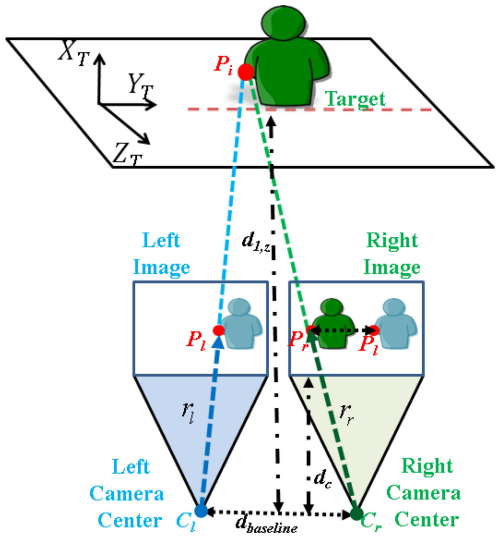


Fig. 3. Stereo IR target distance calculation.

The distance between the two cameras is expressed by d_{baseline}, the distance from the image plane to the camera center is denoted by d_c, and the feature points seen in the right and left images are represented by P_r and P_l. Fig. 3 illustrates the robot's Stereo IR view used in calculating the matched feature points of the targets.

IV. TMF for Collaborative Robotics

We implemented TMF for collaborative robotics based on the images from the fusion of the O-D IR and Stereo IR sensors. The O-D IR and Stereo IR combination provides an easier way to detect a target in a wider FOV while offering a higher degree of precision in target tracking. However, IR sensor-based methods result in higher prediction error for target tracking due to its low image resolution. Our MAP-based approach aims to predict target-position using our multisensor fusion, and improve the robots' ability to track targets. We utilized a MAP-based tracking method in order to decide which sensor was generating a more accurate result by searching their accuracy history. The difference between visual MAP-based methods and our method is that we used it to obtain a weight to increase the importance of a specific sensor, instead of tracking a target and increase the accuracy of tracking. This process helped us to improve the localization of the human target in 3-D space. Application of the O-D IR and Stereo IR fusion is organized into three steps: first, 3-D coordinate estimation of the target from O-D IR and Stereo IR multisensor fusion in Section IV-A. Second, in Section IV-B, the retrieval of occluded trajectory of the target, and finally, the localization and tracking of the target by multiple robots in Section IV-C. The flowchart of these steps is given in Fig. 4.

A. 3-D Coordinate Estimation of the Target From O-D IR and Stereo IR Multisensor Fusion

Combining data from both the Stereo IR and O-D IR sensors improves the 3-D coordinate estimation compared to results obtained from only one robot view. Utilizing sensors from two

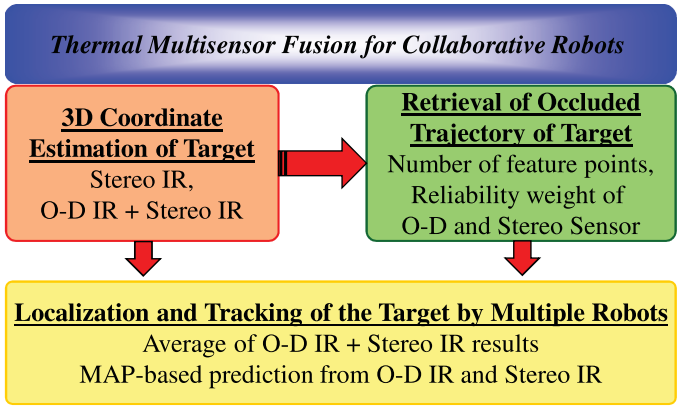


Fig. 4. Flowchart of the proposed method.

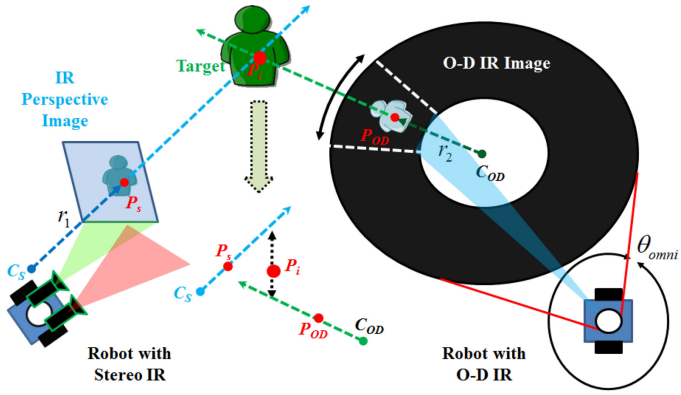


Fig. 5. O-D IR and perspective IR fusion for target position calculation in the common FOV.

robots produces stereo vision which improves the ability to calculate the depth value of each feature point. This value is calculated by using the disparity equation, $d = x_1 - x_2$, between corresponding feature points of images taken by Robot 1 and Robot 2.

The O-D IR camera increases the robots' FOV for detecting targets and the Stereo IR camera system improves the precision of acquiring the target's 3-D coordinates. The O-D thermal sensor makes use of the stereo view to create a collaborative multiview stereo vision from heterogeneous fusion, as shown in Fig. 5. Three IR cameras obtain two different sets of 3-D coordinates for both targets. The Stereo IR acquires the first set of each target's coordinates, and the O-D IR fused with one of the IR perspective cameras provides the second set of 3-D coordinates. These values are tracked by the MAP estimation algorithm, which predicts the subsequent positions of the targets. Then, the predicted 3-D coordinates are used for the final prediction of the targets by the MAP algorithm. Multiple robots decide the assignment of targets to robots.

The fusion of the stereo perspective IR camera with the O-D IR camera provides additional 3-D coordinates of the targets. The O-D camera utilizes a 360° field of vision to detect targets and aims a ray vector toward the target from the center of the O-D camera C_{OD}. A target's feature points are used to find the center of the target P_{OD}, and the ray vector from the camera center is

generated. The perspective camera generates an additional ray vector from its camera center C_s to the target's feature point P_s. The intersection of the O-D IR and perspective IR cameras' ray vectors provide the target's set of 3-D coordinates P_i. We use the closest point of these ray vectors since the feature point selection and camera ray vectors are not perfect in practice; intersection may not occur for every case. In this case, the closest points from the two ray vectors are found, and the average of those coordinates is used as an intersection point. The ray vectors from both O-D and perspective IR cameras are shown in Fig. 5. A closer look at Fig. 5 illustrates the process of finding the closest point of the ray vectors r_1 and r_2 if they do not intersect. The 3-D coordinates of the target point P_i is calculated as follows:

$$\begin{aligned} r_1 &= C_s + uP_s \\ r_2 &= C_{\mathrm{OD}} + vP_{\mathrm{OD}} \\ n &= \sqrt{(r_1 \times r_2) \cdot (r_1 \times r_2)} \\ P_i &= \left[(C_s + nP_s) + (C_{\mathrm{OD}} + nP_{\mathrm{OD}})\right]/2 \end{aligned} \tag{4}$$

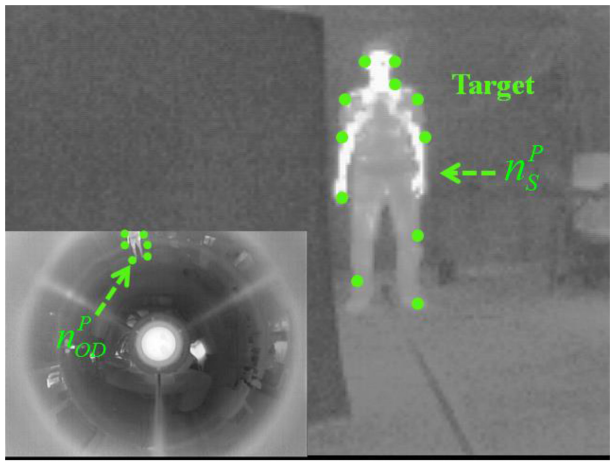
where u and v are scalars that extend the ray vectors until they intersect or make contact with the closest set of coordinates. Final target coordinates P_i will provide the distance of the target from the reference robot. The target's position for each captured moment is calculated by (5) while the target is moving on its trajectory

$$\begin{aligned} n_t &= \sqrt{\left(r_{l,t}^{\mathrm{multi}} \times r_{\mathrm{OD},t}\right) \cdot \left(r_{l,t}^{\mathrm{multi}} \times r_{\mathrm{OD},t}\right)} \\ P_t &= n_t r_{\mathrm{OD},t}. \end{aligned} \tag{5}$$

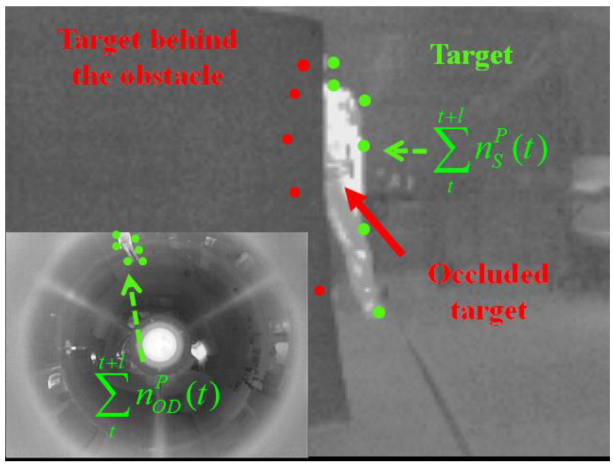
The identified targets in every frame at time t are separately used in the 3-D coordinate estimation process. The O-D IR and perspective IR cameras utilize both of the targets' thermal signatures to intersect the cameras' ray vectors by using the cross product $''\times''$ and the dot product $''\cdot''$ to obtain a scalarn_t, giving the target's position on this ray vector. The ray vectors from the robots' point of view $r_{l,t}^{\mathrm{multi}}$ and $r_{\mathrm{OD},t}$ are used in the form of a multisensor ray vector of IR Stereo's left perspective camera and an O-D sensor ray vector, respectively. The target's coordinates P_t, at time t, are calculated from the intersection of the ray vectors.

B. Retrieval of Occluded Trajectory of the Target

In the event that a target is occluded by an object during the real-time measurements for long term tracking, the short term position prediction and tracking processes will be unreliable. The multisensor fusion tracking method will offer long term tracking information about the target aided by the wide FOV from the O-D sensor. The position of the target will be chosen based on the visibility of feature points obtained from the multisensor fusion data (see Fig. 6). The weight of reliability, (6), will be proportional to the number of features obtained for a detected target and the detection time so that the target's position can be estimated accurately. The weight for the O-D sensorw_{OD}^r is derived from the proportion of the total feature points of the target in the O-D image, $\sum_t^{t+l} n_{\mathrm{OD}}^P(t)$, to the total feature points of the



(a)



(b)

Fig. 6. Occlusion causes a decrease in the number of feature points. (a) Nonoccluded target. (b) Feature points of the occluded target are not entirely detected.

target in the stereo and O-D sensors,$\sum_t^{t+l}\left(n_{\mathrm{OD}}^P(t) + n_S^P(t)\right)$, during the occlusion time l. Thus, two sensors provide the position of each target depending on the corresponding sensor's visibility

$$w_{\mathrm{OD}}^r = \frac{\sum_t^{t+l} n_{\mathrm{OD}}^P(t)}{\sum_t^{t+l}\left(n_{\mathrm{OD}}^P(t) + n_S^P(t)\right)}. \tag{6}$$

After accurately finding the features of each target, the visual reliability weight method is applied to the estimation result of the corresponding sensor with respect to the duration of occlusion, given by the algorithm in Fig. 7. In finding the precise position of a target, the MAP estimation will also be proportionally applied to the visual weighted target position to increase the accuracy of prediction. This proportion is attained from the success of each sensor in localizing the target according to the calculation of the target's coordinates.

C. Localization and Tracking of the Target by Multiple Robots

The O-D and Stereo sensors of TMF provide two different coordinates of the target. Two 3-D reconstruction method results,

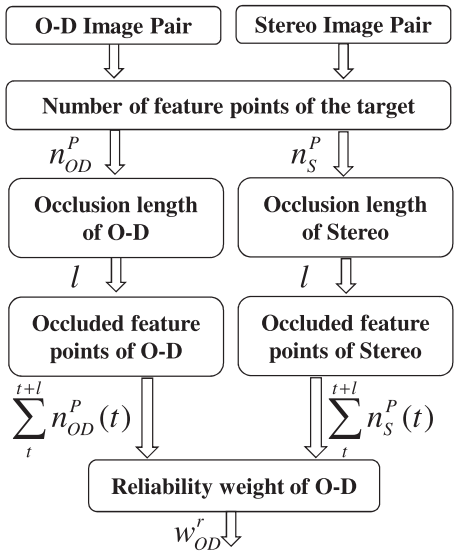


Fig. 7. Algorithm of the reliability weights for O-D and Stereo cameras when the target is not completely visible by any of the sensors.

obtained from the IR perspective Stereo $P_{t,S}$ and the multisensor fusion of IR O-D and perspective cameras $P_{t,\mathrm{OD}}$, are used in Section IV-A to find the accurate positions of the two targets by using the fusion of the reconstructed points. When the coordinates of the targets are available, the MAP estimation method utilizes these coordinate measurements and predicts the position of the target with high accuracy.

The assignment of targets to robots is decided by the MAP estimation result that receives inputs of the predicted human target position H from every mobile robot given the symbol of the sensor S. In this case, we use our configuration of two mobile robots that can be extended to K number of mobile robots. The MAP estimation applied to the target's trajectory using multiple robots is given by

$$\mathbb{P}(H|S_{1:K}) = \mathbb{P}(H_0)\prod_{k=1}^{K}\mathbb{P}(S_k|H). \tag{7}$$

When the posterior probability $\mathbb{P}(H|S_{1:K})$ is available, the maximum of its values will give us the best probable positional information regarding the target of interest. The maximum value of $\mathbb{P}(H|S_{1:K})$ is found by

$$\hat{H} = \underset{H}{\operatorname{argmax}}\, P(H|S_k). \tag{8}$$

The MAP estimation for the target's next position obtained from more than one sensor, given for K sensors, can be written in the following (9) in terms of all the sensors

$$\mathbb{P}(H|S_{1:K}) = \mathbb{P}(H_0)\mathbb{P}(S_1, S_2 \ldots S_K|H) \tag{9}$$

where $\mathbb{P}(H_0) \sim N\left(\theta_H, \sigma_H^2\right)$ is the prior information and $\mathbb{P}(S_1, S_2 \ldots S_K|H) = \mathbb{P}(S_1|H)(S_2|H)\ldots\mathbb{P}(S_K|H)$ is the joint likelihood whose elements are given by $\mathbb{P}(S_1|H) \sim N\left(\theta_H, \sigma_1^2\right)$, $\mathbb{P}(S_2|H) \sim N\left(\theta_H, \sigma_2^2\right)$ and $\mathbb{P}(S_K|H) \sim N(\theta_H, \sigma_K^2)$. The prior information H_0 is calculated from the average of all sensor's prediction.

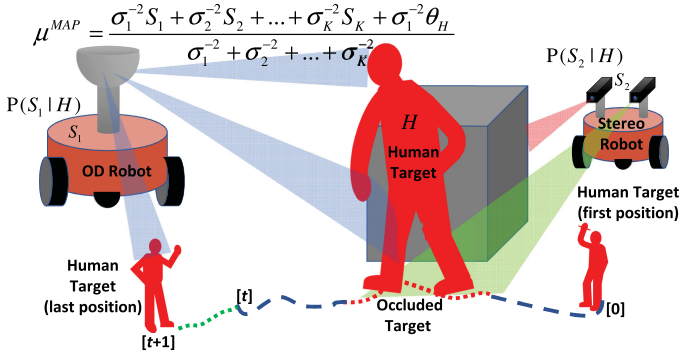


Fig. 8. MAP estimation utilizes the target positions from each sensor as part of the TMF. The occlusion of target demonstrates how Stereo can fail and O-D can be used for recovery of the occluded target.

Thus, $\mathbb{P}(H|S_{1:K}) \sim N\left(\mu_{\mathrm{MAP}}, \sigma_{\mathrm{MAP}}^2\right)$ is utilized to derive the target's coordinate where $\sigma_{\mathrm{MAP}}^{-2} = \sigma_1^{-2} + \sigma_2^{-2} + \cdots + \sigma_K^{-2} + \sigma_H^{-2}$

$$\mu^{\mathrm{MAP}} = \frac{\sigma_1^{-2}S_1 + \sigma_2^{-2}S_2 + \cdots + \sigma_K^{-2}S_K + \sigma_1^{-2}\theta_H}{\sigma_1^{-2} + \sigma_2^{-2} + \cdots + \sigma_K^{-2}}. \tag{10}$$

The standard deviation is calculated from each sensor's previous prediction accuracy of the entire trajectory, and θ is obtained from the previous state of the corresponding sensor's prediction. The standard deviation of prior belief is obtained by the error between the average result of the sensors and the final decision for the target's position. The target's most probable position coordinates are predicted from (11) and (12) for both distance as well as bearing values, which are calculated by (x, y) coordinates of the targets on the plane

$$\begin{aligned}\hat{x}^{\mathrm{MAP}} &= \mu_x^{\mathrm{MAP}} \\ &= \frac{\sigma_{1,x}^{-2}S_{1,x} + \sigma_{2,x}^{-2}S_{2,x} + \cdots + \sigma_{K,x}^{-2}S_{K,x} + \sigma_{1,x}^{-2}\theta_{H,x}}{\sigma_{1,x}^{-2} + \sigma_{2,x}^{-2} + \cdots + \sigma_{K,x}^{-2} + \sigma_{H,x}^{-2}}\end{aligned} \tag{11}$$

$$\begin{aligned}\hat{y}^{\mathrm{MAP}} &= \mu_y^{\mathrm{MAP}} \\ &= \frac{\sigma_{1,y}^{-2}S_{1,y} + \sigma_2^{-2}S_{2,y} + \cdots + \sigma_{K,y}^{-2}S_{2,y} + \sigma_{1,y}^{-2}\theta_{H,y}}{\sigma_{1,y}^{-2} + \sigma_{2,y}^{-2} + \cdots + \sigma_{K,y}^{-2} + \sigma_{H,y}^{-2}}\end{aligned} \tag{12}$$

where the predicted position of the target is at $\hat{H}^{\mathrm{MAP}}$ $(\hat{x}^{\mathrm{MAP}}, \hat{y}^{\mathrm{MAP}})$. For multiple robots, the position estimation for a target can be calculated from the equations above by adding all the sensor information about the corresponding target $\mathbb{P}(S_1, S_2 \ldots S_K|H)$. Fig. 8 demonstrates how the O-D sensor's blue line of sight combined with the stereo sensor's green line of sight finds the first 3-D position of the target. Each robot obtains its own set of predicted coordinates of the target. The O-D robot provides one set of the target's coordinates S_1 and the Stereo robot provides another set of the target's coordinates S_2 to the MAP (10) for the final localization of the target by TMF. MAP estimation is applied to obtain the precise target position by merging all the received information about the target. The distance of the target position from each robot is obtained in

TABLE III
HARDWARE AND DATASETS

Sensor	Sensor	Field of View (degree)	Mobility / Optic	Resolution/ Power
O-D IR	Remote-Reality	360°	Static / Fixed lens	640x480 / DC
Stereo IR	Raytheon Thermal	50°	Static / Fixed lens	640x480 / AC
Mobile Robot	Pioneer DX-3	NA	Dynamic	DC

(13) for the assignment of the closest robot

$$\text{Dist}_{k,H} = \sqrt{\left(\hat{H}^{\text{MAP}}(\hat{x}^{\text{MAP}}) - S_k(x)\right)^2 + \left(\hat{H}^{\text{MAP}}(\hat{y}^{\text{MAP}}) - S_k(y)\right)^2} \tag{13}$$

where $S_{k,H}$ is the robot k with the assigned target position H, shown in Fig. 8.

V. EXPERIMENTS

The experiment is divided into four sections and each one examines the results of the methodology sections. The first one gives details of our experimental setup: Section V-A presents hardware and datasets; Section V-B presents 3-D coordinate estimation of the target from O-D IR and Stereo IR multisensor fusion; Section V-C presents retrieval of occluded trajectory of the target; and Section V-D presents localization and tracking of the target by multiple robots. We demonstrated the improvement of TMF results on finding the target's position by the MAP estimation in this section.

A. Hardware and Datasets

The sensory configuration of our system consists of two Pioneer 3-DX mobile robots equipped with multiple sensors. One of the robots has an O-D thermal remote reality sensor. This sensor captures 360° bagel-shaped images and offers a wide FOV for the robot. The middle portion of the images contains a white circular area due to the camera's sensor capturing itself, and the high temperature of the working sensor is seen in white. The second robot has a Stereo vision setup that utilizes two Raytheon thermal-eye cameras mounted on top of the mobile platform. These sensors capture thermal perspective images and have a 50°FOV. Both robots are equipped with Windows computers and an Intel Xeon processor. The robots and their corresponding sensors are shown in Fig. 1 as well as the Hardware and Datasets are given with sensory details in Table III. We used 1600 different images with different surrounding setups and targets to compare the O-D Stereo fusion in addition to the occlusion to compare the change of feature points while the occlusion occurs. The objective of the proposed method is to improve the position calculation of the human target rather than detection and classification of multiple targets. For this reason, we evaluate the accuracy of the target's position, tracked and localized only one human target in the images. The number of targets can be increased by the advanced thermal target detection method. The environment of the target is the critical component for target detection methods; however, the detection part is simplified for the proposed method to show the effects of reconstruction and MAP-based estimation method on a detected target.

(a) (b)

Fig. 9. (a) Printed grid pattern in an image from visible-band camera. (b) Printed grid pattern in a thermal image.

The datasets for our experiments were collected by the O-D IR and Stereo IR sensors. The images have 640 × 480 pixel resolutions with a gray level color map. The color value of the pixels depends on the temperature of the objects captured, if the temperature is high, the color of the pixel is bright. The objects with a lower temperature have a color pattern close to black. The Thermal Stereo sensor was calibrated using the Matlab's computer vision toolbox for the stereo camera calibration method. The calibration process provided the intrinsic and extrinsic parameters of the Thermal Stereo sensor. We have also used the O-D camera calibration toolbox for our Thermal O-D sensor. However, this method was designed for visual cameras and are not sufficient to calibrate the thermal sensors so it would not work for thermal cameras. Therefore, we have designed our calibration hardware and used a custom made calibration checkerboard which is a grid pattern made out of square-shaped aluminum-based tape and a white paper-based background for the calibration. In order to utilize these visible-band-based calibration methods with our thermal sensors, we heated the checkerboard; because the printed black and white checkerboard patterns are not visible in thermal images and do not help to calibrate the sensors. Fig. 9(a) shows a grid pattern in an image from the visible-band camera, and Fig. 9(b) shows the same grid pattern in a thermal image. It is easy to see that the grid pattern is not visible in the thermal image and the calibration is not possible with conventional methods. The aluminum-based grid pattern used for the calibration process is shown in Fig. 10(a). We can also see the grid pattern in the thermal O-D image, and on the left and right images of the thermal Stereo sensors in Fig. 10(b)–(d). The possible error caused due to the thermal calibration process by using the heated chessboard is handled via our MAP-based approach, and image registration is synchronized by taking the images at the same time from both robots.

A human target is captured with varying distances from the robots in the O-D IR and the Stereo images. The human target is selected from the O-D and the Stereo thermal images with respect to the human body features such as size, height to width ratio, and temperature. Both images had specific filters for the

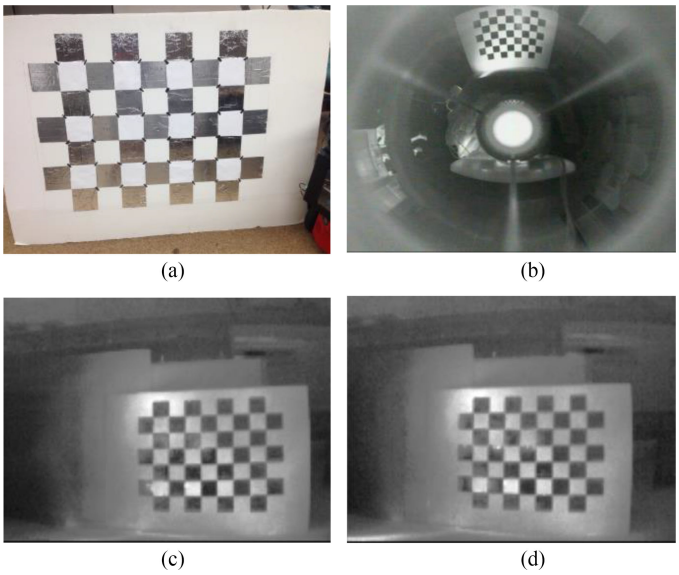


Fig. 10. (a) Heated grid pattern used for the calibration of the Thermal Stereo sensors and Thermal O-D sensor. (b) Heated grid pattern in the thermal O-D image (c) in the left image and (d) in the right image of thermal Stereo.

human region's size, which is adjusted according to the preferred target's distance from the robots. The human body ratio is used between the values of 1.7 to 2.0. The heat signature of the human body is the key feature to increase its visibility while the other objects around the human have similar temperatures that are not easy to distinguish. The mineigenfeature detection method from Matlab's computer vision toolbox is utilized to detect human feature points [25]. The feature points in the stereo image pairs are then matched from the detected feature points. This feature detection method provided better results for feature point matching for the human regions in our experiments. However, the feature extraction and matching between thermal O-D and perspective image is challenging since the Thermal O-D and Stereo sensors are mounted on separate mobile platforms and the format of these images are different. Therefore, the feature matching method is used for only the thermal Stereo image pair. The human target region is obtained by the O-D sensor using the human target region detection with the feature points and human body characteristics matching in the image. Then, another human target region is acquired from the Stereo sensor after the first 3-D coordinate calculation of the human target with the stereo image pairs. The limited resolution of the thermal images, feature extraction, matching, and data association difficulties between the O-D and perspective sensors lower the accuracy of target localization. The proposed MAP-based method improves the accuracy by selecting the more reliable sensor results from the collaborative robots.

B. 3-D Coordinate Estimation of the Target From O-D IR and Stereo IR Multisensor Fusion

The multisensor fusion of O-D IR and the Stereo IR use 3-D coordinate estimation that is implemented and the distance results were obtained from each sensory setting since the distance value provides the average of total error. The proposed method used the distance value to evaluate the general performance of the system. The main mission for human target tracking is maintaining a specific distance between human and robots which makes distance the most important component for human tracking and our experiment. The human target was recorded from five different ranges using both robots. First, the target was at 1701 mm range for the long-range experiment; second, the target was recorded from around 1500, 1300, and 900 mm as part of medium-range tests; then, the target was at a 762 mm range for a close-range investigation. For the comparison we used three distances from the nearest to farthest including the middle point. Fig. 11(a) shows the human target from the O-D sensor, and Fig. 11(b) shows the human target with the matched feature points from the Stereo image pair.

TABLE IV
O-D IR AND STEREO IR MULTISENSOR FUSION

Range	Distance of Target (mm)	Average Error of Stereo IR (mm)	Average Error of O-D and Prsp. IR Fusion (mm)	Estimation Error (%)
Long	1701	75.89	59.70	3.98
Medium	1300	84.19	78.30	6.24
Close	762	79.00	96.25	11.49
Avg. Error	0.00	79.69	78.08	7.23

The experiment results showed that the O-D IR and Stereo IR sensor fusion provided a lower average target position prediction error when the Stereo IR sensor predicted the target position with higher error [see Fig. 11(c)]. The error was calculated from the predicted result and the calculated result for each step when both values were obtained. The prediction used the normalized error estimation squared to find the estimation consistency of each sensor. The experimental results can be seen in Table IV. The Stereo IR gave a better prediction for the close-range tests as seen a lower average prediction error of 17.25 mm. When the target was in close range to the sensors, the Stereo IR provided a 2.3% lower prediction error, which increased the accuracy with respect to the O-D IR fusion prediction. The prediction error of the O-D IR fusion was 12.63% for the close-range dataset and decreased the error to 3.5% while the target moved away from the robots. The O-D IR fusion obtained a similar error percentage as Stereo IR by approximately 6.01% and had better accuracy after that range. The multisensor fusion offered us the advantage of higher accuracy for the long-range target position prediction by the Stereo sensor while taking advantage of higher accuracy from O-D IR fusion for the targets in a close range of the robots.

C. Retrieval of Occluded Trajectory of the Target

The occlusion dataset was recorded when there was an object between the Stereo robot and the target. The object blocked the target's view mostly from the left camera of the Stereo sensor. While the target was moving behind the obstacle, the feature points were tracked by the stereo sensor. Some of the matched feature points were lost since the corresponding features were

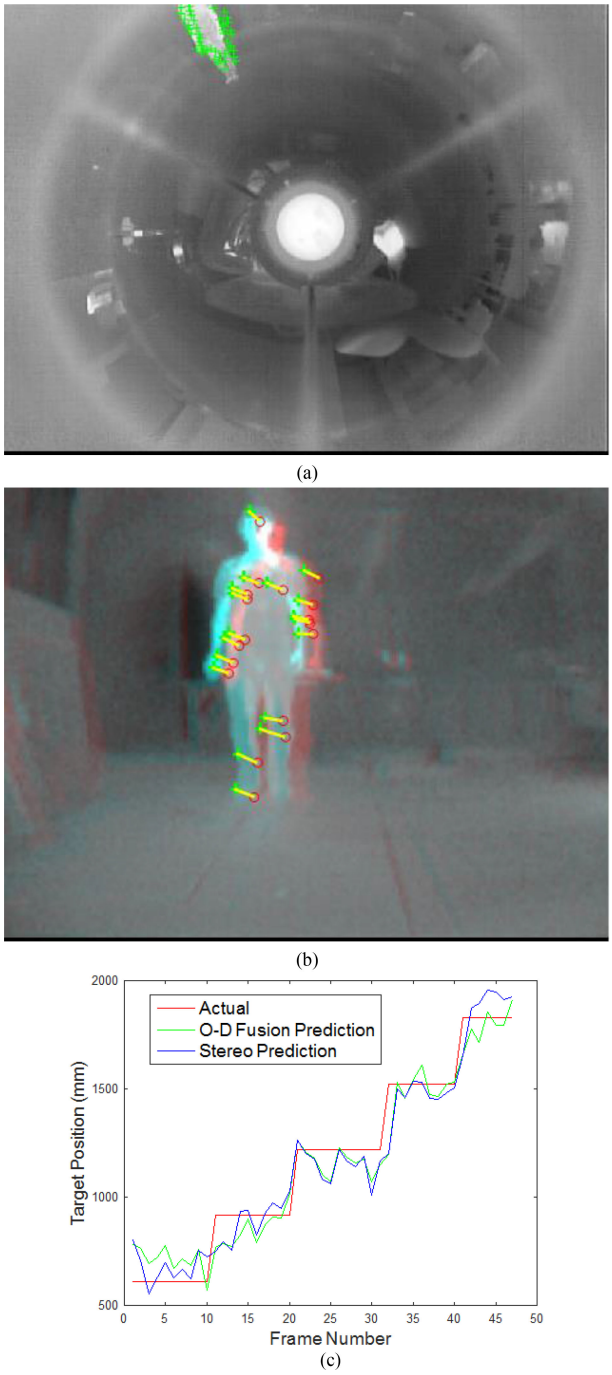


Fig. 11. Stereo IR, O-D IR, and perspective IR fusion for target position estimation. (a) Detected human target in the O-D IR image. (b) Human target's matched feature points from the left image and the right image of Stereo IR sensor. (c) Actual position of the target, prediction of the target's position from O-D IR fusion, and Stereo IR are given together.

not detected by each camera in the Stereo IR setup. Fig. 12(a) and (b) show the lost feature points when the target is going behind the obstacle. The second stereo image pair detected and matched a lower number of feature points in Fig. 12(b). The O-D and perspective camera fusion became more reliable in calculating the target's positions when the occlusion started. We calculated the reliability weight with respect to the change in the number of feature points detected from the fusion of the O-D and

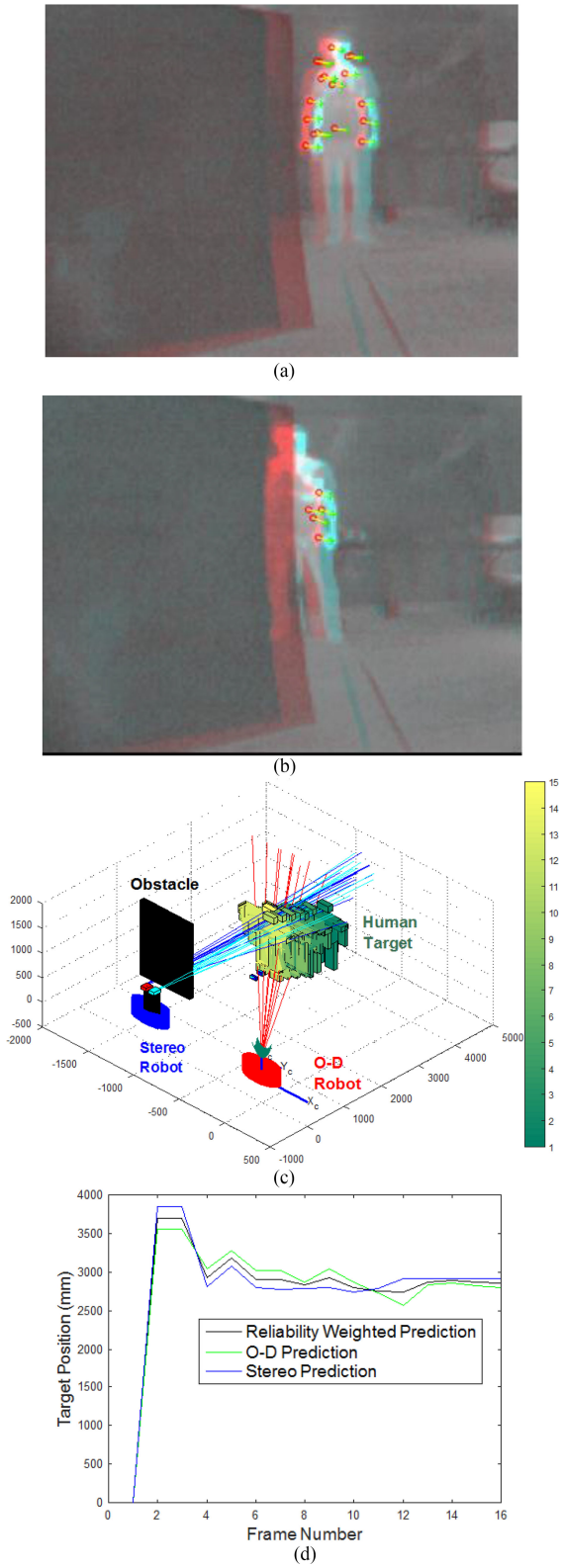


Fig. 12. Retrieval of occluded target trajectory by Stereo IR and O-D IR TMF. (a) Nonoccluded image shows the matched feature points of the human target. (b) Matched feature points of the occluded human target from the left image and the right image show that more than half of the feature points were not detected. (c) Brightening green target positions show that the target is becoming more occluded. (d) Actual position of the target, prediction of the target's position from O-D IR fusion and Stereo IR during the occlusion.

TABLE V
RETRIEVAL OF OCCLUDED TRAJECTORY OF THE TARGET

Method	Prediction Error	Average Number of Features	Retrieved Trajectory (%)
Stereo IR	61.42	10	2.04
O-D and Prsp. IR Fusion	42.42	15	1.41
Weighted Position	42.35	25	1.41

perspective camera in addition to Stereo sensors. This reliability weight applied to both the target position results from the Stereo IR and O-D IR fusion. The obstacle between the left Stereo IR camera and the target is shown with black in Fig. 12(c). The O-D sensor is given by the red box with its coordinate system. The second robot is shown with the blue box and the Stereo sensors on top. The ray vectors from those sensors are also given with an O-D ray vector in red, ray vectors of the left sensor from the Stereo in dark blue, and ray vectors of the right sensor from the Stereo in cyan.

The target position was predicted from each sensor fusion and the calculated position results were used to find the weighted final target position. The weighted target position utilized the ratio from the total feature point before and after occlusion in the Stereo image pair. A lower number of feature points from the Stereo sensor decreased the reliability of accuracy from this sensor. After the reliability weight was derived from the occluded feature points during the occlusion, each target position was weighed with respect to the reliability of the corresponding sensor. If the Stereo sensor starts losing matched feature points, then the O-D Stereo Fusion result is weighted by increased reliability, thus the final localization of the human target is closer to the O-D Stereo Fusion result. We can see that the brightening green target positions in Fig. 12(c) was predicted accurately even though the left stereo camera did not provide any information about the target. The prediction error of the target position decreased by 31.04% after the occlusion started, as shown in Fig. 12(d). In Table V, we can see that the prediction error did not provide a close error to the 61.42-mm Stereo IR prediction and it gave a close result to the more reliable O-D IR error result of 42.35 mm.

D. Localization and Tracking of the Target by Multiple Robots

The final localization of the target was conducted by the MAP estimation. This method collected the target's positions along the target's trajectory and made a final decision according to the corresponding sensor's prediction accuracy. When the image frames from each sensor are available, the target position is calculated for that instant by using the ray vectors from each sensor. Then, the MAP estimation is applied to the results of the O-D IR Fusion and Stereo IR to localize the final position of the target, as shown in green in Fig. 13(a).

The proposed method is compared amongst the Stereo IR, O-D IR fusion, and MAP estimation results. The compari-

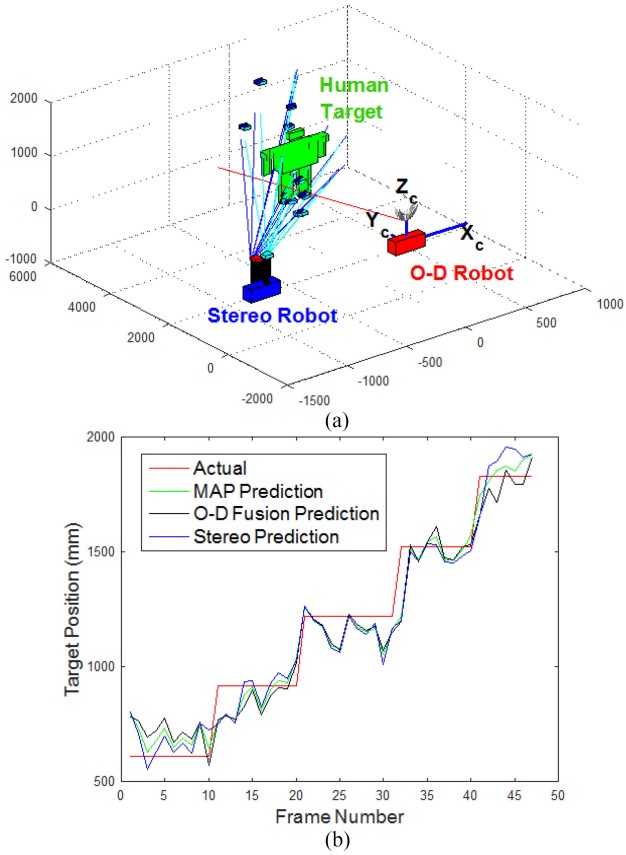


Fig. 13. Target localization by the MAP estimation. (a) Final position of the target from TMF by the MAP estimation given in green. (b) Target position follows the actual position by the MAP estimation with a green line.

TABLE VI
PERFORMANCE OF THE THERMAL STEREO

	Close Range	Medium Range	Long Range	Average
Stereo IR Error	(±)79.00 (mm)	(±)84.19 (mm)	(±)75.89 (mm)	(±)79.69 (mm)
Actual Distance	762 (mm)	1300 (mm)	1701 (mm)	1254 (mm)
Accuracy	89.63 (%)	93.52 (%)	95.54 (%)	93.64 (%)

TABLE VII
LOCALIZATION OF TARGETS BY TMF WITH THE MAP ESTIMATION

Method	Close Range (mm)	Long Range (mm)	Prediction Error (mm)	Long Range Prediction Error (%)
Stereo IR	79.0071	75.8986	80.0783	4.46
O-D and Prsp. IR Fusion	96.2575	59.7014	83.5565	3.51
Average of Two Sensors	79.7874	51.7935	73.8225	3.04
MAP Prediction	80.3720	48.3027	67.5892	2.83

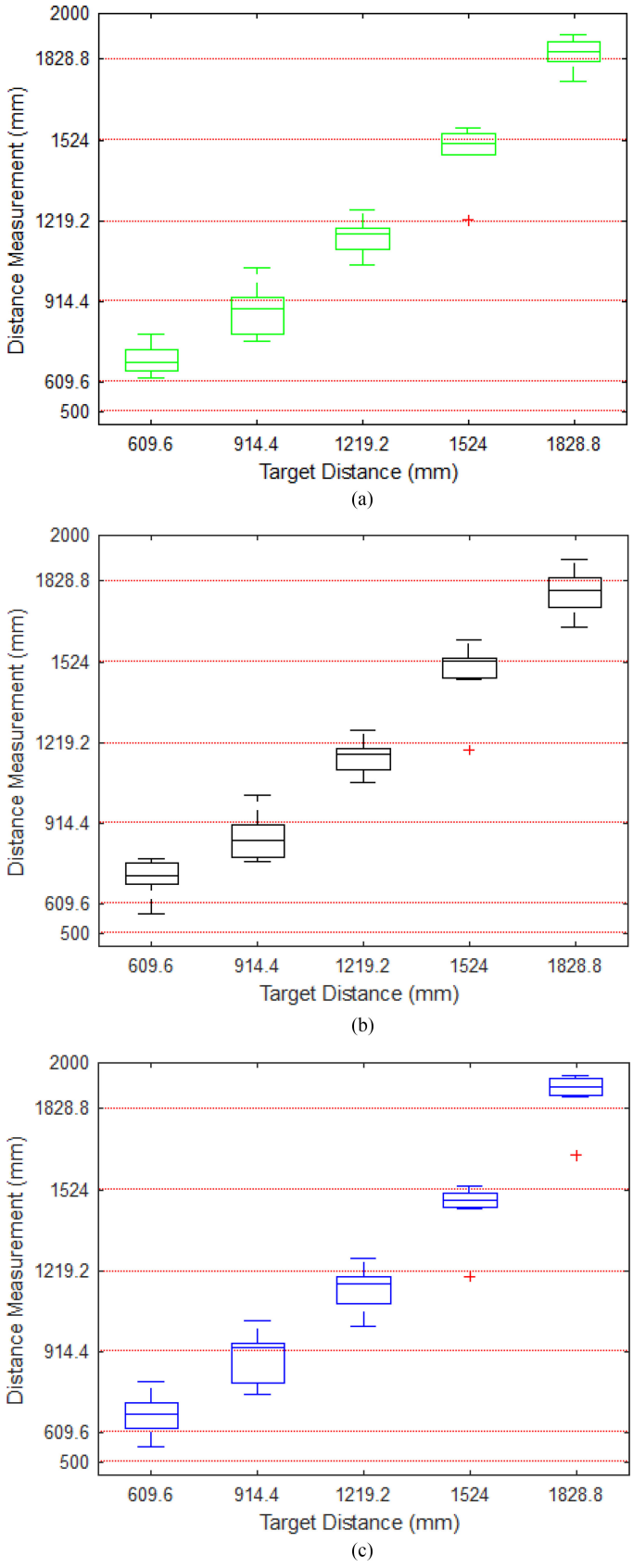


Fig. 14. Statistics of multiple measurements with respect to different target distances of (a) the MAP estimation. (b) O-D IR Fusion. (c) Stereo IR.

son between the visible-band MAP estimation methods is not equivalent to our method since the sensors have different features. Thermal sensors have the advantage to track the targets for a longer time under any lighting condition while visible-band sensors cannot generate any result. Thermal images have lower resolution and this causes difficulty on position calculation. The MAP estimation predicted the position of the target with a 19.09% higher accuracy than the O-D IR Fusion prediction and with a 36.35% higher accuracy than the Stereo IR prediction for the long range of the human target. The trajectory of the target is given with the prediction of the final localization in green [see Fig. 13(b)]. It can be seen that the final localization follows the actual target position with higher accuracy than both sensors. Table VI shows the performance of the Stereo IR sensor only and when the target distance increases it has better performance. The accuracy of the Stereo IR method for close range is 89.63% with 79.00 mm average distance prediction. The Stereo IR provides 93.52% accuracy for the medium range of targets and 84.19 mm average prediction error in distance. The highest accuracy of Stereo IR is for longer distance targets with 95.54% accuracy and 75.89 mm average error. However, Stereo IR prediction for close range gave the highest accuracy among all methods and our MAP-based method provided lower accuracy with a close prediction to the Stereo method. The target's position was predicted with a 59.7014 mm error by the O-D IR Fusion for the long range of the target's distance, and this was the lowest error among the predictions. This high prediction accuracy was improved and the 59.7014 mm error was reduced to 48.3027 mm with the MAP estimation (see Table VII). The statistics of multiple measurements, as shown in Fig. 14, were analyzed to show the performance difference between each method with the standard deviation and median around the actual target distance. Fig 14(a) shows that the MAP prediction improved the statistical accuracy for the close and long-range distances of the human target compared to O-D Fusion and Stereo IR methods. The statistics of these three methods also show that O-D Fusion predicts the distances given by the red line with higher accuracy for long range and lower accuracy for the close range [see Fig 14(b)]. The Stereo IR method provides opposite results to the O-D Fusion [see Fig 14(c)]. Thus, the TMF improved the prediction accuracy of the human target position when the Stereo IR setup was not successful by itself. The TMF provided a desirable sensor prediction result and maintained a low prediction error by using the prediction history of the sensors.

VI. Conclusion

In this paper, we implemented TMF for target tracking using mobile robots. The target position was precisely calculated by Stereo IR vision fused with O-D IR. The position obtained from the sensor fusion allowed the MAP estimation method to predict the target position with improved accuracy. The prediction by the MAP estimation method is used to decide the importance of each sensor instead of the traditional MAP-based visual tracking methods. The proposed method is compared with respect to

Stereo IR, O-D IR and MAP estimation results since the visible-band and thermal sensor properties and performance are not comparable. We utilized the thermal signature of the human target for light-independent tracking and O-D vision to track the human target in a wider FOV. Our experimental results showed that using the TMF with the MAP estimation enhanced the human localization by 36.35% with respect to the Stereo sensor from a single robot. The proposed method used one target since the purpose is to evaluate the MAP-based target localization. Our future work will focus on tracking multiple targets rather than using a single human target. The proposed method will increase the accuracy for the detected targets in future works. The number of collaborative robots will be increased as well as being equipped with TMF of O-D IR, Stereo IR, and single perspective IR sensors.

ACKNOWLEDGMENT

The authors would like to thank F. Fanary and A. Huynh for proofreading to help improve the writing quality of this paper.

REFERENCES

[1] V. D. Hoang and K. H. Jo, "A simplified solution to motion estimation using an omnidirectional camera and a 2-D LRF sensor," *IEEE Trans. Ind. Inform.*, vol. 12, no. 3, pp. 1064–1073, Jun. 2016.
[2] R. C. Luo and C. C. Chang, "Multisensor fusion and integration: A review on approaches and its applications in mechatronics," *IEEE Trans. Ind. Inform.*, vol. 8, no. 1, pp. 49–60, Feb. 2012.
[3] D. You, X. Gao, and S. Katayama, "Multisensor fusion system for monitoring high-power disk laser welding using support vector machine," *IEEE Trans. Ind. Inform.*, vol. 10, no. 2, pp. 1285–1295, May 2014.
[4] R. Mottaghi and R. Vaughan, "An integrated particle filter and potential field method applied to cooperative multi-robot target tracking," *Auton. Robots*, vol. 23, no. 1, pp. 19–35, Jul. 2007.
[5] A. Ibarguren, I. Maurtua, M. A. Perez, and B. Sierra, "Multiple target tracking based on particle filtering for safety in industrial robotic cells," *Robot. Auton. Syst.*, vol. 72, pp. 105–113, Oct. 2015.
[6] M. Marron-Romera *et al.*, "Stereo vision tracking of multiple objects in complex indoor environments," *Sensors*, vol. 10, no. 10, pp. 8865–8887, Oct. 2010.
[7] N. Koyama, R. Tajima, N. Hirose, and K. Sukigara, "IR tag detection and tracking with omnidirectional camera using track-before-detect particle filter," *Adv. Robot.*, vol. 30, no. 13, pp. 877–888, 2016.
[8] N. A. Tsokas and K. J. Kyriakopoulos, "Multi-robot multiple hypothesis tracking for pedestrian tracking," *Auton. Robots*, vol. 32, no. 1, pp. 63–79, Jan. 2012.
[9] H. Shi, X. Li, K. S. Hwang, W. Pan, and G. Xu, "Decoupled visual servoing with fuzzy Q-learning," *IEEE Trans. Ind. Inform.*, vol. 14, no. 1, pp. 241–252, Jan. 2018.
[10] K. Hausman, J. Müller, A. Hariharan, N. Ayanian, and G. S. Sukhatme, "Cooperative multi-robot control for target tracking with onboard sensing," *Int. J. Robot. Res.*, vol. 34, no. 13, pp. 1660–1677, Nov. 2015.
[11] G. Huang, M. Kaess, and J. J. Leonard, "Consistent unscented incremental smoothing for multi-robot cooperative target tracking," *Robot. Auton. Syst.*, vol. 69, pp. 52–67, Jul. 2015.
[12] G. Huang, K. Zhou, N. Trawny, and S. I. Roumeliotis, "A bank of maximum a posteriori (MAP) estimators for target tracking," *IEEE Trans. Robot.*, vol. 31, no. 1, pp. 85–103, Feb. 2015.
[13] K. L. Bell and R. Pitre, "MAP-PF 3D position tracking using multiple sensor array," in *Proc. 5th IEEE Sensor Array Multichannel Signal Process.*, 2008, pp. 238–242.
[14] W. Choi, C. Pantofaru, and S. Savarese, "A general framework for tracking multiple people from a moving camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1577–1591, Jul. 2013.
[15] J. Bosch, N. Gracias, P. Ridao, K. Istenic, and D. Ribas, "Close-range tracking of underwater vehicles using light beacons," *Sensors*, vol. 16, no. 4, p. 429, Apr. 2016, pages 26.
[16] G. Du, P. Zhang, and D. Li, "Human–manipulator interface based on multisensory process via Kalman filters," *IEEE Trans. Ind. Electron.*, vol. 61, no. 10, pp. 5411–5418, Oct. 2014.
[17] L. Wang, M. Liu, and M. Q.-H. Meng, "Real-time multisensor data retrieval for cloud robotic systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 2, pp. 507–518, Apr. 2015.
[18] C.-H. Chen, Y. Yao, D. Page, B. Abidi, A. Koschan, and M. Abidi, "Heterogeneous fusion of omnidirectional and PTZ cameras for multiple object tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 8, pp. 1052–1063, Aug. 2008.
[19] K. Zhou and S. I. Roumeliotis, "Multirobot active target tracking with combinations of relative observations," *IEEE Trans. Robot.*, vol. 27, no. 4, pp. 678–695, Aug. 2011.
[20] J. P. Barreto, L. Perdigoto, R. Caseiro, and H. Araujo, "Active stereo tracking of $N <= 3$ targets using line scan cameras," *IEEE Trans. Robot.*, vol. 26, no. 3, pp. 442–457, Jun. 2010.
[21] A. T. Hafez, A. J. Marasco, S. N. Givigi, M. Iskandarani, S. Yousefi, and C. A. Rabbath, "Solving multi-UAV dynamic encirclement via model predictive control," *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 6, pp. 2251–2265, Nov. 2015.
[22] G. Du, P. Zhang, and X. Liu, "Markerless human-manipulator interface using leap motion with interval Kalman filter and improved particle filter," *IEEE Trans. Ind. Inform.*, vol. 12, no. 2, pp. 694–704, Apr. 2016.
[23] M. Lauer, M. Schönbein, S. Lange, and S. Welker, "3D-objecttracking with a mixed omnidirectional stereo camera system," *Mechatronics*, vol. 21, no. 2, pp. 390–398, Mar. 2011.
[24] T. Svoboda and T. Pajdla, "Epipolar geometry for central catadioptric cameras," *Int. J. Comput. Vis.*, vol. 49, no. 1, pp. 23–37, 2002.
[25] "Computer vision system toolbox documentation," 2017. [Online]. Available:https://www.mathworks.com/help/vision/

Emrah Benli (S'15–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 was 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 Assistant Professor of Electrical and Electronics Engineering with Gümüşhane University, Gümüşhane, Turkey. His research interests include intelligent systems, computer vision, artificial intelligence, multimodal sensory, robotic system design and control, and human–robot interaction.

Richard Lee Spidalieri received the B.Sc. degree (cum laude) in electrical engineering from Virginia Commonwealth University, Richmond, VA, USA, in 2018.

He is currently a Project Engineer in the power industry in Chester, VA, USA. His primary responsibilities include testing, maintaining, and commissioning industrial power systems and their protection schemes.

Yuichi Motai (S'00–M'03–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 with Virginia Commonwealth University, Richmond, VA, USA. His current research focuses on the broad area of sensory intelligence, particularly in medical imaging, pattern recognition, computer vision, and sensory-based robotics.