

Vegetation Segmentation for Sensor Fusion of Omnidirectional Far-Infrared and Visual Stream

David L. Stone , *Member, IEEE*, Guarav Shah , *Member, IEEE*, Yuichi Motai , *Senior Member, IEEE*, and Alex J. Aved, *Senior Member, IEEE*

Abstract—In the context of vegetation detection, the fusion of omnidirectional (O-D) infrared (IR) and color vision sensors may increase the level of vegetation perception for unmanned robotic platforms. Current approaches are primarily focused on O-D color vision for localization, mapping, and tracking. A literature search found no significant research in our area of interest. The fusion of O-D IR and O-D color vision sensors for the extraction of feature material type has not been adequately addressed. We will look at augmenting indices-based spectral decomposition with IR region-based spectral decomposition to address the number of false detects inherent in indices-based spectral decomposition alone. Our work shows that the fusion of the normalized difference vegetation index (NDVI) from the O-D color camera fused with the IR thresholded signature region associated with the vegetation region minimizes the number of false detects seen with NDVI alone. The contribution of this paper is the demonstration of a new technique, thresholded region fusion technique for the fusion of O-D IR and O-D color. We also look at the Kinect vision sensor fused with the O-D IR camera. Our experimental validation demonstrates a 64% reduction in false detects in our method compared to classical indices-based detection.

Index Terms—Camera, image processing, measurement false positive (FP), omnidirectional (O-D) far-infrared (IR), robot sensing systems, sensor fusion, vegetation detection, vision, visualization.

I. INTRODUCTION

WHILE there has been a great deal of research in the area of intelligent perception, the practical application of omnidirectional (O-D) instrumentation, both O-D infrared (IR) and O-D color vision sensors, to vegetation detection has not received adequate attention. This paper explores the reduction of false positives (FPs) in index-based vegetation recognition by the fusion of these two sensors. The contribution of this paper

Manuscript received August 1, 2018; revised October 24, 2018 and December 6, 2018; accepted December 24, 2018. Date of publication February 5, 2019; date of current version March 4, 2019. This work was supported in part by the Office of Naval Research, in part by the Naval Surface Warfare Center Dahlgren Division, in part by the Marine Corps Warfighting Lab, Quantico, VA, USA, and in part by Air Force Research Laboratory. (*Corresponding author: Yuichi Motai.*)

D. L. Stone is with the Department of Electrical and Computer Engineering, Virginia Commonwealth University, Richmond, VA 23284-3068 USA (e-mail: stonedl3@vcu.edu).

G. Shah is with the Bon Secours Health Systems, Richmond, VA 23226 USA. Y. Motai is with the Virginia Commonwealth University, Richmond, VA 23284-3068 USA (e-mail: ymotai@vcu.edu).

A. J. Aved is with the REID, Air Force Research Laboratory, Rome, NY 13441 USA (e-mail: Alexander.Aved@us.af.mil).

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/JSTARS.2019.2891518

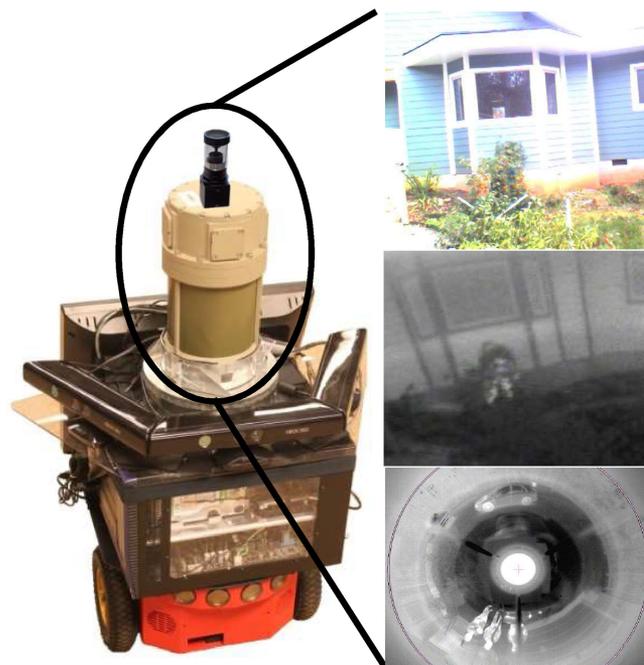


Fig. 1. O-D IR, O-D Color, Kinect cameras, and onboard computer. Showing a sample O-D IR image and two extracted IR and color images of front of the house and a man crouched behind the bush.

is the demonstration of a new technique called thresholded segmentation fusion (TRF) to fuse the visual/IR normalized difference vegetation index (NDVI)-based results with the segmented thermal vegetation region from the O-D IR sensor.

The authors are leveraging the benefits of potentially low-cost O-D IR and O-D color cameras, see Fig. 1. While we were waiting for our O-D color camera, we investigated the fusion of the O-D IR camera with the Kinect color vision sensor. These sensors coupled with intelligent perception algorithms may provide the solution of this problem. With improved omnidirectional systems, the corrections required are reduced, compared to multiple camera stitching, to create 360° fusion of our O-D IR and O-D color cameras, the computation requirements are reduced and sources of error are improved for electro-optical cameras. Additionally, multi-spectral sources can be fused with reduced error and the calculation errors for sensor fusion do not propagate forward in the perception system world model. This makes these O-D sensors ideal for low-cost perception of the robot's environment.

TABLE I
CAMERA SETTING

Sensor	Image Type	Advantage	Disadvantage
<i>O-D IR</i>	far-IR	Wide FoV	Low Resolution
Kinect	Red/Near-IR	High Resolution	Narrow FoV
O-D Color	Visual	Wide FoV	Mid Resolution

This paper is laying out the concept for improved vegetation detection through the fusion of the IR and visual streams. We initially provide experimental results using a 360° far-IR camera coupled with three Kinect cameras. We also apply the techniques to an O-D IR camera and an O-D visual camera. The current sensors are mounted on top of a robot platform and we will develop a more robust dataset and analysis. Fig. 1 is an example of the O-D IR, O-D color, and Kinect cameras mounted on top of the robot, and a sample of a frame extracted from the scene by the IR and color cameras.

The reason for choosing O-D IR and O-D vision is due to their superior performance in the following areas.

- 1) Wide field of view (FoV)—360°.
- 2) Optical flow or structure from motion—translation and rotational flow fields has different characteristic shape in spherical system.
- 3) With camera and mirror axis perpendicular to the floor—vertical lines map to radial lines in the spherical coordinate system.
- 4) O-D far-IR provides robustness in varying light conditions day or night.
- 5) A single O-D image completely defines the visual characteristics of a location.

Table I summarizes the camera setting.

The contribution of this paper is to adopt an O-D far-IR camera and O-D color camera on our mobile robot and developing a new fusion framework, called thresholded region fusion (TRF) for semantic extraction multi-spectral fusion of O-D IR and O-D color vision. Our approach will fuse a modified NDVI (MNDVI), modified for far IR instead of near-IR (NIR) with a region-based thermal semantic structure to improve the number of false detects with the traditional NDVI approach.

The remainder of this paper is organized as follows: Section II describes some prior work in O-D applications, and sensor fusion, Section III introduces the proposed system approach, which consists of the utilization of region-based thresholded histogram segmentation coupled with multi-spectral signatures from both O-D vision and O-D IR cameras to output the vegetation regions. Section IV provides the experimental setting and results for semantic extraction of context-dependent features. Section V concludes the paper.

II. PRIOR RELATED WORKS

Robust and inexpensive intelligent perception is a key enabler to the practical application of unmanned ground vehicles (UGV) operating with teams of humans in areas of police, rescue, and

TABLE II
OMNIDIRECTIONAL (O-D) CAMERA APPLICATIONS SUMMARY

Method	Reference	Model	Approach
Localization	[1]	SLAM	Mixed
Homography	[2, 3]	Visual Geometry	Mixed
Navigation	[4, 5]	Sphere	Optical Flow
Robotic Feature Tracking	[6]	Vertical Line Recognition	Robust Feature Descriptor

military applications. In addition, truly commercial applications will benefit from this research and other intelligent perception work as well. Our review of the literature highlights that the focus of recent research with O-D cameras has been primarily in the area of improvements in mapping, localization and tracking, robot navigation, and obstacle detection.

The authors believe that the fusion of O-D IR and vision has the potential to provide gains in intelligent robotic perception. This section covers related works and is organized as Section II-A *Omni-Direction Camera Setting*, Section II-B *Visual and IR Index-Based Vegetation Detection*, Section II-C *IR Stream Segmentation Using Region-Based Thermal Analysis*, and Section II-D *Sensor Fusion Methodologies*.

A. Omni-Direction Camera Setting

The spherical geometry characteristics of the O-D sensor are leveraged. Table II summarizes the literature review for this section.

There are several approaches to localization and homography (determining the visual geometry of the scene) [1]–[6]. Anderson *et al.* [1] apply O-D vision and odometry to self-localization of a UGV in a non-static environment, and to determine robot’s global pose (x, y, θ) . They also looked at several sensor fusion models and whether fusing data then detecting/classifying (low-level fusion) or the utilization of individual sensor modality to detect/classify and then fusing the results (high-level fusion) gave better overall results. Our approach is to use feature level fusion coupled with segmentation techniques as described in Section III.

Homography (visual geometry) estimation from feature points extracted from a moving omni-vision sensor is another popular approach used in both [2] and [3]. The application of multiple homographies extracted from virtual image planes from O-D visual camera pairs was explored in [2]. Navigation is explored in [4] and [5]. Identification of vertical line geometry in the image is addressed in [6].

B. Visual and IR Index-Based Vegetation Detection

Vegetation detection is required because the identification of vegetation aids the robot in detecting materials that it can pass through such as grass. The work in [7]–[15] addresses different aspects of vegetation detection. However, these approaches are prone to false detects, which we hope to minimize through our approach.

TABLE III
VEGETATION DETECTION METHODOLOGY SUMMARY

Method	Reference	Model	Approach
NIR	[7, 8]	Index Based	Satellite Remote Sensing
Multi Spectral	[9]	Hyperspectral	Mixed
NDVI	[10 - 12]	Various	Improve NDVI
NDVI Variant	[13]	Principal Component Analysis	Statistical Framework
NDVI	[14,15]	Topological Effects	Sensitivity Analysis

TABLE IV
SEGMENTATION AND HISTOGRAM METHOD SUMMARY

Method	Reference	Model	Approach
LIDAR and NDVI Fusion	[16]	Sensor Fusion	Fusion of Lidar and NDVI results
Multi-sensor	[17]	NDVI time series	Spectral Response Functions
Histograms	[18-20]	RGB and Gray Scale Histogram	Region matching
Region Segment	[21-25]	Mixed	Region and Edge, Histograms

Table III summarizes the literature review for this section. The identification of vegetation is especially useful. If the object detected is made out of leafy vegetation or grass, it is more likely passable than a tree trunk or a man-made object. The use of the NIR spectrum in addition to the visual spectrum has been shown to provide useful information in the detection of vegetation in remote satellite imaging [7], [8]. The application of multispectral and hyperspectral techniques is explored in [9]. NDVI [10]–[15] is a very popular index approach to vegetation classification in remote sensing. However, these approaches that are applied in satellite imagery are taking a macro look at large datasets to classify ground cover. In the NDVI approach, a comparison of the levels of red reflectance to the NIR reflectance has been shown to correspond to vegetation.

C. IR Stream Segmentation Using Region-Based Thermal Analysis

Segmentation and fusion techniques have been explored [16]–[20], and both region-based and edge-based approaches have been demonstrated. The work in [16] augments the NDVI with three-dimensional (3-D) LIDAR point cloud to compensate for the vegetation index shortcomings; this is an approach which utilizes expensive sensors.

The work in [17] evaluates the calibration of spectral response functions from multiple sensors to improve the fusion of multi-sensor NDVI time series. Our approach will fuse low-cost optical and IR sensors to minimize the false detects. Localization from O-D vision using RGB color histograms [18]–[20] to match images in the database of regions the robot may visit is another approach. Table IV summarizes the literature review for this section.

Region segmentation techniques have been explored [21]–[25], and both region-based and edge-based approaches have been demonstrated. Taiana *et al.* in [21] apply a multi-mode sensor fusion of Lidar, color vision, and near-IR cameras fused to provide perception of the terrain around a mobile robot.

Localization from O-D vision using RGB color histograms [22], and a similar approach saves computing power by extracting a numerical signature gray scale histogram using Haar discrete wavelet transform [23]. This approach also uses expensive Lidar sensors, which we are trying to avoid. Histogram approaches have also been used for region segmentation [22]–[25] using clustering techniques based on histogram thresholds.

D. Sensor Fusion Methodologies

Sensor fusion techniques have been explored in several multi-sensor approaches [26]–[50]. Table V summarizes the literature review for this section. The following are presented:

- 1) local point statistic;
- 2) conditional local point statistic;
- 3) 2-D–3-D feature fusion;
- 4) laser remission;
- 5) NDVI-laser fusion;
- 6) SVM multi-cam fusion;
- 7) multi-cam NDVI fusion.

The work in [26] provides a survey of sensor fusion methods, which are categorized into the following three levels.

- 1) Raw data level fusion—Sensors with similar characteristics are fused to improve signal to noise [27], [28], [33], [39], [43].
- 2) Feature level fusion—Features are extracted from different types of sensors and then similar features are fused for improved confidence [29], [30], [37], [38], [40]–[42].
- 3) Decision level fusion—The sensors are processed individually and then the relevant information is then fed into a separate decision processor to decide on the most confident outcome [31], [32], [35], [36].

The work in [31] applies decision level fusion to fuse the process sensors of vibration, noise, and force. The work in [32] applies sensor fusion to robot off-road navigation, fusing individual sensor results by probability of blocked/unblocked and feeding this into high-level behavior decision model. A data level fusion of sensor signals [33] is used to develop a thresholded composite failure index that combines both data fusion and degradation modeling to establish the composite index. The authors [33] apply the fusion of multiple machining quality sensors to achieve this data level fusion. The work in [34] characterizes forest canopy using airborne LIDAR and hyperspectral feature-based fusion. The work in [35] applies data fusion to an omnidirectional and PTZ camera for optimal multi-target tracking.

The fusion of environmental features [36] from visual and IR cameras is used to calculate the probability of pedestrian location. The automatic scene segmentation into salient features is accomplished by different methods and their results

TABLE V
SENSOR FUSION METHOD SUMMARY

Method	Reference	Model	Approach
Structured overview	[26]	Survey	Various
Extended Kalman Filter	[26]	Decision fusion	Composite process health index
Raw Data Level Fusion	[27, 28], [33], [39], [43]	Data fusion	Improve signal to noise similar sense
Feature Level Fusion	[29, 30], [37, 38], [40–42]	Feature fusion	Feature Extraction at sensor level
Decision Level Fusion	[31, 32], [35, 36]	High level decision	Individual sensor analysis high level probabilistic confidence fusion
Human-Robot Fusion	[34]	Bayesian Soft Max	Gaussian priors and human guidance
IR-Vision Fusion	[44]	Visual Gradient and infrared intensity	Feature Level Fusion
IR-Vision Fusion	[45]	Non subsampled Contour-let Transform	Fuse sub-bands and low level features
IR-Vision Fusion Label-Me	[46]	Label-Me	MIT CSAIL Labeling Tool
Compressive Data Fusion	[47]	Compressive Fusion	Random Projections
NDVI	[48]	NDVI	Analysis of ground cover for loss of forest cover.
Lidar-Hyperspectral Fusion	[49, 51] [52–53]	Lidar point cloud plus Mixture-tuned matched filter for Hyperspectral	Lidar – canopy structure, Hyperspectral – species spectral signature

fused. The fusion of depth and color features [37] is used to distinguish crops from weeds. Feature fusion of visual and synthetic aperture radar is used to extract roads and buildings [38].

The work in [39] uses the data level fusion of IR and visual camera, preserving the intensity distribution of the IR image and the gradient variation of the visible image. Chen *et al.* [40] survey several feature fusion methods, including substitution techniques-based, independent component analysis-based, principal component analysis-based, segmentation-based, neural network-based, mathematical morphology-based, and multi-scale transform-based fusion schemes. The work in [41] and [42] utilize feature extraction at the sensor level. The work in [43] provides fusion at the raw data level. Jiayi *et al.* [44] apply gradient transfer fusion to extract the gradient information from the visual image fused with minimization of the total variation of the infrared intensity data. Li *et al.* [45] apply a multi-scale transform using a non-subsampled contour-let transform to extract

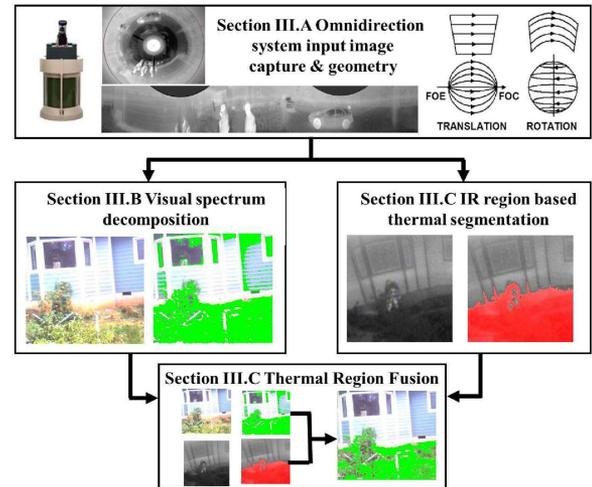


Fig. 2. Technical approach consists of the following subsections: III.A: Omnidirection camera setting. III.B: Visual and IR index-based vegetation detection. III.C: IR stream segmentation using region-based thermal analysis and thresholded region fusion.

high- and low-frequency sub images and fuse with low-level features. Mancas *et al.* in [46] use region growing for segmentation. Russell *et al.* applied the MIT CSAIL LabelMe database and web tool for image annotation to label the images [47].

III. TECHNICAL APPROACH

The fusion of the O-D color visual and O-D IR data was used to enable the robotic perception system to adapt to different lighting and environmental conditions. In this paper, we focus on the fusion of O-D far-IR and visual stream with the system approach shown in Fig. 2.

We compare our results to [48]–[50] in Section IV. Prasad *et al.* in [48] present a compressive data fusion approach for multi-sensor fusion. Zaitunah *et al.* [49] studied land cover changes using NDVI from 2005 to 2015 to assess the impact of changes in watershed on ground cover. Results included shrub 0.67–0.737, secondary forest 0.737–0.804, and primary forest 0.804–0.876. Sankey *et al.* [50] used a UAV to gather Lidar and Hyperspectral data. The Lidar point cloud determined the canopy structure and the hyperspectral identified species-specific spectral signatures. The data were fused to achieve overall 0.76 accuracy. The work in [51]–[53] shows results for FPs ranging from 32.5% to 61.29% with FP peaks as high as 77.4%.

Our proposed system architecture applies texture and frequency analysis (segmentation and signature) to images from low cost, O-D visual, and IR sensors to provide object detection, classification, and tracking. The final approach is the fusion of visible and IR perception systems using O-D sensors. Each block in Fig. 2 represents the subsections as follows.

- 1) Section III-A discusses the O-D camera system and the O-D coordinate system setting.
- 2) Section III-B develops the computational foundation for a new approach to multi-spectral sensor fusion that identifies semantically significant object classes based on

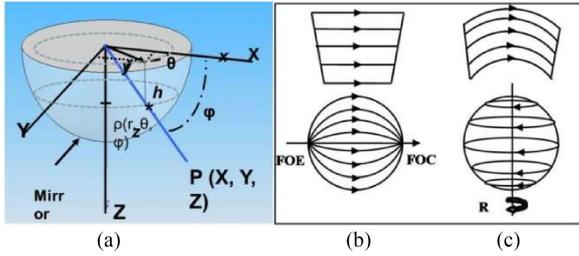


Fig. 3. (a) Geometry of the O-D camera. (b) Translation. (c) Rotation.

spectral and thermal signatures using visual and IR spectrum decomposition, finally.

- 3) Section III-C discusses the method for IR region segmentation using region-based thermal threshold analysis, and discusses the TRF method.

A. Omni-Direction Camera Setting

Using spherical reflecting surfaces, a projected 360° FoVis created and transformed into a single 360 linear plane. With a single omni-vision approach, reconstructions of scenes create a reduced set of unknowns in the equations as compared to the many camera approach. With improved omni systems, the corrections required are reduced to create 360 fusions, the computation requirements are reduced and sources of error are improved for electro-optical cameras. Additionally, multi-spectral sources can be fused with reduced error and the calculations errors for sensor fusion does not propagate forward in the perception system world model.

The transformational relationship between rectangular coordinates and spherical coordinates as shown in Fig. 3 is described in the following equations:

$$r^2 = x^2 + y^2 \quad (1)$$

$$\theta = \tan^{-1}(x/y) \quad (2)$$

$$\varphi = \tan^{-1}(r/z) \quad (3)$$

$$z = (r^2 - h^2)/2h. \quad (4)$$

In O-D optical flow, the difference between flow fields in rectangular space and spherical space can be used to extract semantics from the O-D image. Translation in rectangular space maps to arcs in the spherical space in the direction of flow and that is why O-D vision is useful for determining ego-motion. Note that translation has focus of expansion (FOE) and focus of contraction (FOC) points in the image where rotation does not. Translation flow is from the FOE to the FOC. This motion can be calculated. The rotation motion is preserved as circles on the surface of the mirror in the spherical coordinates. Vertical lines in the rectangular space map to radial lines in the spherical image space.

In Fig. 3, x and y are the image plane coordinates of the pixel, P is the point in space being observed by the O-D camera, and the parameter z traces the surface of the mirror. In the spherical coordinate system, θ and φ define the direction of the ray and

r is the distance to the origin of the mirror. The orientation θ ranges from 0° to 360° around the edge of the mirror, and the pitch φ ranges from 0° , when pointing straight down, to 90° , when pointing at the horizon. The calibration parameter h and the calibration process were presented in [1], where the camera was calibrated.

B. Visual and IR Index-Based Vegetation Detection

The classical vegetation detection looks at red color bands in the visual spectrum and compares this to near-IR spectrum. The NDVI feature works well in chlorophyll-rich vegetation, but does not work well in dry vegetation or desert scenes. It also was a problem when operating in new areas that were not trained. Our approach hopes to use the thermal threshold for vegetation fused with the color vision MNDVI to improve the performance in dry vegetation and false detects.

Our approach modifies the NDVI to look at the red band and far-IR band difference and fuse this with the thermal characteristics of the vegetation. Vegetation is detected based on the ratio of red reflectance and far-IR reflectance. This technique allows us to take advantage of the physical properties of vegetation and how they reflect and absorb light. This is influenced both by the absorption of the chlorophyll and the water content in the vegetation. These techniques apply the band ratio of the far-IR bands and red band in the visual range. The water content of the vegetation absorbs light above 1400 nm, and the chlorophyll absorbs the red and blue bands leaving the green band. The result is that the vegetation shows a high reflectance in the green and IR bands. Several vegetation indices have been developed, which are as follows.

The NDVI is given as (5), which is the ratio of the difference and sum of the NIR reflectance and the RED reflectance

$$\text{NDVI} = (I_{\text{NIR}} - I_{\text{RED}}) / (I_{\text{NIR}} + I_{\text{RED}}). \quad (5)$$

Several other indices are also used such as the difference vegetation index (DVI) as $\text{DVI} = I_{\text{NIR}} - I_{\text{RED}}$, which is the difference between NIR reflectance and the RED reflectance. The perpendicular vegetation index (PVI) is given as $\text{PVI} = \sin(\alpha) I_{\text{NIR}} - \cos(\alpha) I_{\text{RED}}$, where PVI is the difference between the NIR reflectance times $\sin(\alpha)$ and RED reflectance times $\cos(\alpha)$ and represents the perpendicular distance from the NIR reflectance point to the soil line. The angle alpha (α) is the angle between the soil line and the near infrared reflectance axis.

We use the NDVI, where $\text{NDVI} = (I_{\text{NIR}} - I_{\text{RED}}) / (I_{\text{NIR}} + I_{\text{RED}})$ modified for far-IR (6)

$$\text{MNDVI} = (I_{\text{IR}} - I_{\text{RED}}) / (I_{\text{IR}} + I_{\text{RED}}). \quad (6)$$

The MNDVI was fused with the region-based IR signature explained in Section III-C to enhance the semantic segmentation of the IR/Vision stream. We are using potentially low-cost IR and visual sensors to achieve multiband spectral signatures. Fig. 4 shows the location of a point of interest in the O-D IR sensor, unwrapped image, and extracted IR and visual images.

We looked at multispectral bands and region-based IR spectral analysis to better characterize the spectral signature of different materials. This paper looks at the effectiveness of the

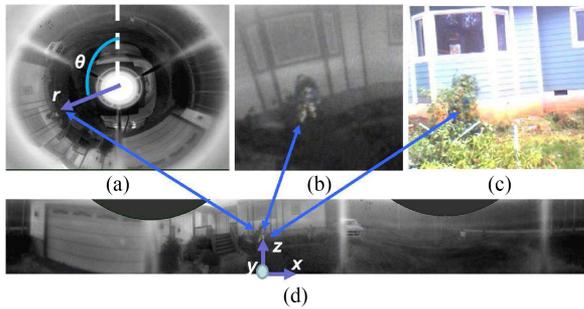


Fig. 4. O-D geometry (a) shows 360° O-D image with the r , θ geometry overlaid. (b) Cropped IR region of interest. (c) Visual cropped region of interest. (d) Unwrapped rectangular image, and the X , Y , Z geometry axis.

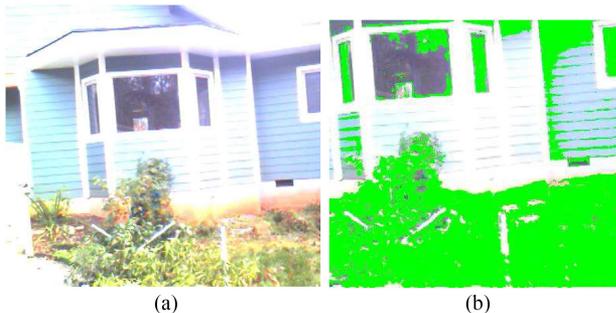


Fig. 5. MNDVI approach. (a) Original image. (b) Processed image using MNDVI vegetation index approach. This result has a relatively high number of false detects.

MNDVI and region-based IR method as compared to the classical NDVI approach to more fully characterize the signatures of vegetation and other materials. The indices value is then used as a threshold to detect vegetation. There are issues with false detects in the areas where the bands overlap when using just the MNDVI. We used the MNDVI approach on our data in Fig. 5 below to show the effect of this approach and the number of false detects.

We as humans are accustomed to using color to distinguish materials. The color and reflectivity are important indicators of the material composition of an object.

The spectral signature of the material or material reflectance spectrum $\rho(\lambda)$ is related to the reflected light or spectral radiance $L_s(\lambda)$ and the impending scene radiance $L_i(\lambda)$ by the following:

$$L_s(\lambda) = \frac{\rho(\lambda)}{L_i(\lambda)}. \quad (7)$$

This can be rearranged into the reflectance spectrum $\rho(\lambda)$ in terms of the ratio of spectral radiance $L_s(\lambda)$ to the scene radiance $L_i(\lambda)$ as given in the following equation:

$$L_s(\lambda) = \rho(\lambda) * L_i(\lambda). \quad (8)$$

This wavelength λ is filtered into k spectral bands. Fig. 6 shows a representative cumulative density function (CDF) for the three primary color bands (red, blue, green) in the visible spectrum. The 360-color camera is the Hummingbird360 with a Point Gray Flea3 with 12 bits and 8.8-megapixel camera.

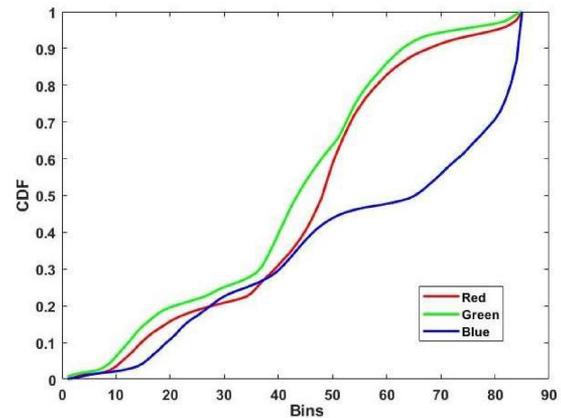


Fig. 6. Representative cumulative density function (CDF) for the three color bands in the image is a summation of the probabilities of each intensity at the different wavelengths.

The CDF is defined as the accumulation of pixels in the histogram of each color band.

The approach we are developing is to fuse the red band visible and IR region spectrum and utilize these spectral signatures to extract scene semantics from the O-D IR and visual images. We demonstrate that each material has a characteristic visual and thermal spectral signature from which the material is to be classified. We then compare the MNDVI red band approach to the combined MNDVI red band and IR region-based thermal signature approach, and finally compare the fused results to the baseline MNDVI indexed-based approach.

C. IR Stream Segmentation Using Region-Based Thermal Analysis

To address the issue of false detects in the index-based approach, semantic extraction based on thermal regions will be investigated to accentuate vegetation detection and reduce the number of FPs.

Fig. 7 is the IR region segmentation using region-based thermal segmentation analysis: (a) is the original image, (b) shows that all of the vegetation is at the same temperature, and (c) shows that the hard concrete road and sidewalks are at a different temperature. The sky is dark and has not absorbed any temperature.

Fig. 7 shows the region-based segmentation of the IR image. The areas of similar gray level (same temperature) will be segmented and matched to the spectral signature for that region and then classified based on the spectral content and thermal characteristics. For instance, the sky is dark with no thermal content, the vegetation is dark gray due to its water content causing it to pick up less heat than the sidewalks or pavement, and the pavement and sidewalks are white due to picking up heat during the day. These thermal-based material regions can then be correlated with the material spectrum to aid in identifying the material.

Fig. 7(a) is the original unwrapped thermal image, Fig. 7(b) shows the vegetation region segmented and has the formula for the region overlaid. The sky is the background and has

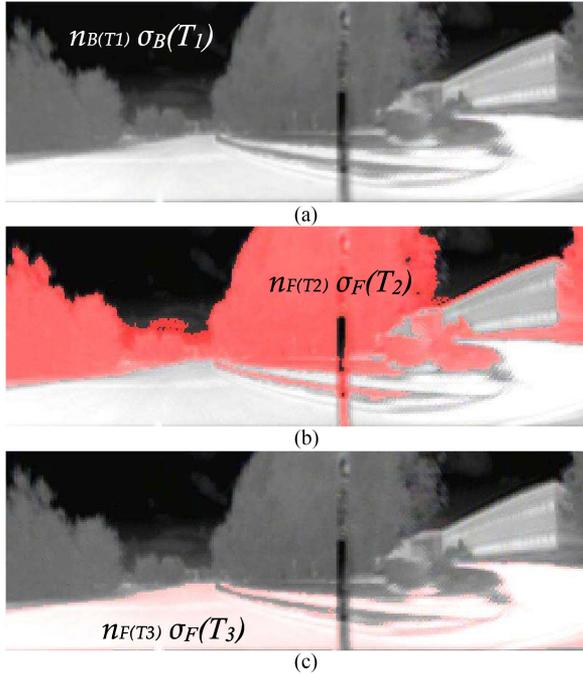


Fig. 7. IR segmentation using region-based spectral analysis. (a) Original image. (b) All of the vegetation is at the same temperature. (c) Hard concrete road and sidewalks are at a different temperature. The sky is dark and has not absorbed any temperature.

the background CDF times the background variance squared at threshold T_1 , and the vegetation has the foreground CDF and foreground variance at threshold T_2 . Fig. 7(c) shows the pavement and sidewalk regions segmented at foreground CDF and variance at threshold T_3 . After segmenting the various material regions and extracting the spectral signature for the regions, the two results will be fused by the union of the two sets.

The thermal image has similar materials clustered in the same region of the IR thermal space. At different ambient temperatures, these materials will shift in color but will remain clustered. The thermal signature of certain terrain types and vegetation will have different emissivities but the behavior is characteristic of the material. By extracting this region-based signature from the thermal behavior and fusing it with color multispectral signature, different types of vegetation and terrain can be classified. During the day, vegetation is cooler than the soil and is darker in the IR image.

Representing the gray scale of the IR image as histograms, we partitioned the image (segmentation) into homogeneous regions representing areas of similar temperature. During the day, the ground and other solid objects absorb heat faster than the vegetation so the vegetation areas will be lighter than the ground. At night, this reverses and the foliage will look darker. Considering this grayscale image, we let the number of times gray level i occur as n_i . The probability P that a pixel at (r, θ, φ) will be at level i is given as follows:

$$P_{r,\theta,\varphi}(i) = \frac{n_i}{n}, 0 \leq i < l \quad (9)$$

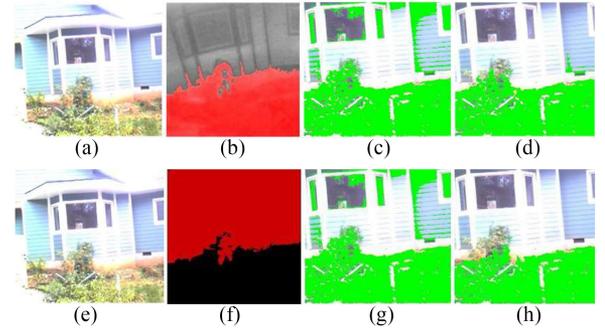


Fig. 8. Row 1: Fusion of MNDVI and thermal vegetation region. (a) Original image. (b) Vegetation thermal region. (c) MNDVI result. (d) Fusion of MNDVI and vegetation thermal region. (b) and (c) still have some false detects. Row 2: Fusion of MNDVI and non-vegetation region. (e) Original image. (f) Non-vegetation thermal region. (g) MNVDI result. (h) Fusion of MNDVI and non-vegetation thermal region. Gives fewer false detects, but greater number of missed areas in the vegetation region.

where l is the total number of gray levels in the image, and n is the total number of pixels. So the probability of intensity i at a given (r, θ, φ) location is the ratio of the number of pixels at intensity i or n_i to the total number of pixels' $n P_{r,\theta,\varphi}(i)$ is the image histogram for pixel i , the value is normalized to $[0, 1]$. We define the CDF as follows:

$$\text{CDF}_{r,\theta,\varphi}(i) = \sum_{j=0}^I P_{r,\theta,\varphi}(j). \quad (10)$$

The CDF is the summation of the probabilities for the intensity at each pixel. Using (11), for optimal thresholding, to segment regions we achieve a mapping of regions of like thermal intensity. Different materials absorb heat differently, and thus we can use this in conjunction with the spectral signature to determine the regions of similar material

$$\sigma_{\text{within}}(T) = n_B(T) \sigma_B^2(T) + n_F(T) \sigma_F^2(T) \quad (11)$$

where η_B is the background histogram CDF

$$n_B(T) = \sum_{j=0}^{T-1} P_{r,\theta,\varphi}(j)$$

where η_F is the foreground histogram CDF

$$n_F(T) = \sum_{j=T}^{N-1} P_{r,\theta,\varphi}(j).$$

σ_B is the variance of the pixels in the background ($<T$)

$$\sigma_B(T) = \sqrt{(h_P - h_{Avg})^2}$$

σ_F is the variance of the pixels in the foreground ($>T$)

$$\sigma_F(T) = \sqrt{(h_P - h_{Avg})^2}$$

where h_P is the histogram value at the pixel of interest and h_{Avg} is the average value of all the pixels above the threshold for foreground or below the threshold for background. The optimum threshold for Figs. 8–10 is determined by a combination of the triangle threshold method and the optimization for T of (11) plus

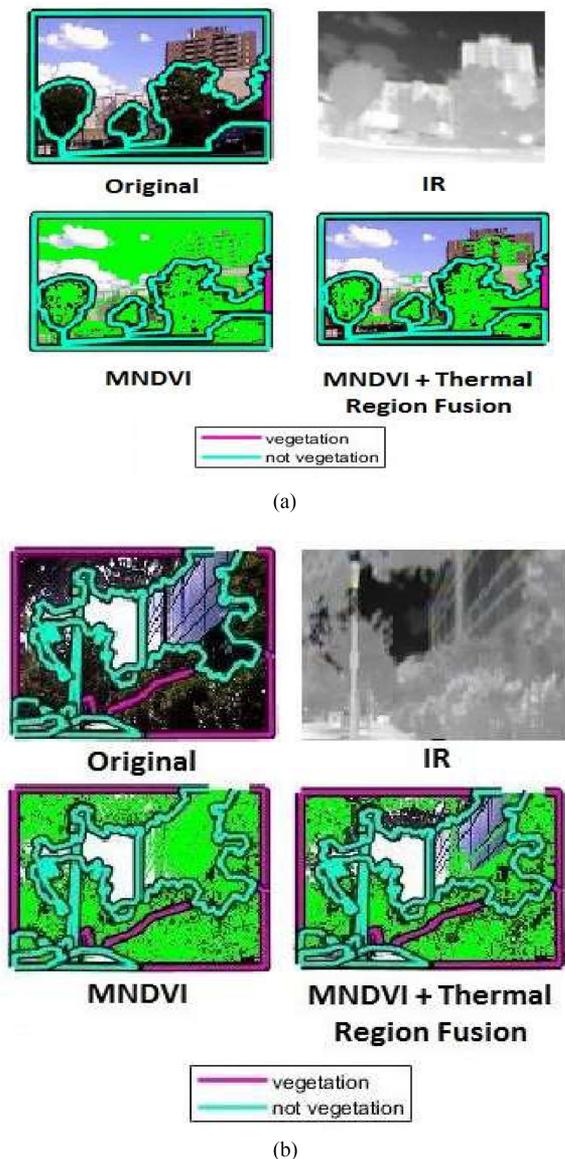


Fig. 9. Two original images (a) and (b) and for each image the following are shown: The original image with ground truth labeling, the IR image, the MNDVI solution, and the thermal region fusion. The vegetation detection and false positive reduction for the IR and Kinect visual images are shown.

the statistical extent of the thermal IR region. Equation (11) is used as input to the fitness function for a genetic algorithm, which optimizes the regions of interest in the given image. For a starting T the standard deviation and variance of the foreground and background histogram are calculated. The CDF above and below that threshold is also calculated. These are then combined in (11) to find the standard deviation of the region of interest. This is tested against the thermal band for vegetation and iterated using the genetic algorithm until the optimum thresholds for the vegetation region are found.

Fig. 8(b) shows the fusion of the region representing the vegetation thermal signature with the MNDVI result in 8(c). It still has some false detects but they are significantly reduced. Fig. 8, second row, is the fusion of the MNDVI result with the

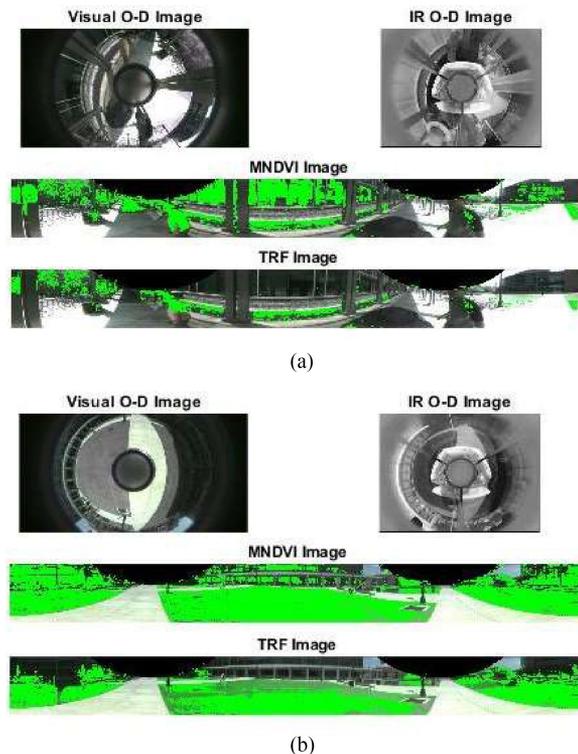


Fig. 10. Two sets of data each showing the original omnidirectional visual image, omnidirectional IR image, the unwrapped O-D visual and O-D IR images with the MNDVI vegetation detection with the thermal region fusion vegetation detection results overlaid. Note that in (a) TRF fused image shows a good capture of the vegetation and rejection of false positives. While in (b) it has good FP rejection, but less accurate vegetation detection.

thermal region representing the non-vegetation region. In this second application of the algorithm, this time the non-vegetation thermal region 8(f) is fused with the MNDVI result 8(g).

Fig. 9 is a comparison for two original images (a) and (b), and for each image the following are shown: the original image with ground truth labeling, the IR image, the MNDVI image, and the thermal region fusion (TRF). The images show the false detects in the MNDVI image and the reduction in those false detects in the TRF image. We used the MIT LabelMe website [47] to label the vegetation and non-vegetation regions shown in Fig. 9.

Utilizing (11) along with the triangle threshold method to find the threshold for the vegetation region, (11) is minimized to find the extent of the region. We also applied a genetic algorithm to find the region of interest. We first apply the modified vegetation index using far-IR instead of near-IR. We next apply thresholding and region growing techniques to identify the regions of vegetation and non-vegetation. The results of the MNDVI and thermal region extraction are then fused to minimize false detects and the results are presented.

The contribution of this work is a computational foundation for a new approach to multi-spectral sensor fusion that identifies semantically significant object classes using sensor fusion of the O-D vision (electro-optical) and O-D IR sensors, applying the two streams into a thresholded segmentation and fusion genetic algorithm for the fusion of the IR and visual Spectral

TABLE VI
PROPOSED TRF ALGORITHM: SENSOR FUSION OF MNDVI
VEGETATION INDEX AND REGION-BASED IR

STEPS	STEP DETAILS
STEP 1	Apply the MNDVI vegetation index to extract region vegetation $MNDVI = (I_{FIR} - I_{RED}) / (I_{FIR} + I_{RED})$
STEP 2	Apply thresholding to the IR image to identify region similar temperature $\sigma_{within}(T) = n_{B(T)}\sigma_B^2(T) + n_F(T)\sigma_F^2(T)$
STEP 3	Apply region growing using a genetic algorithm to find of the areas with the same temperature
STEP 4	Fuse the MNDVI vegetation highlight with the IR re based thermally detected vegetation area to minimize : detects in the MNDVI signature
STEP 5	Compare the vegetation index based results, with the combined MNVDI and Thermal Region Threshold Segmentation Fusion results.

content to extract the regions of vegetation and other materials. Our algorithm applies both a modified segmentation feature extractors into the input fusion/classification. We will evaluate the TRF architecture against the baseline index-based vegetation detection.

Fig. 10 shows two different images of different scenes with the original O-D visual and O-D IR image shown for each case. The unwrapped O-D images are then shown with the MNDVI and TRF detected vegetation regions overlaid. In each case, the MNDVI image shows significant false detects. In Fig. 10(a), the TRF method shows good vegetation detection and also has good FP rejection. However, in Fig. 10(b), while the FP rejection is still good, the vegetation detection is less accurate than the MNDVI, but is still in the range of other methods (see Table IX below). It can be seen that there is a registration problem between the sidewalk in the visual image and the TRF overlay. The authors believe this is due to the difference in viewing angle from the two cameras. The effects of this mismatch can be seen in Fig. 10(b).

The overall process is shown in Table VI. The two algorithms will be applied and the results fused using the TRF approach.

- 1) MNDVI index based modified for far-IR vegetation detection.
- 2) Region-based thresholded thermal segmentation augmented with a genetic algorithm to optimize the selection of the region of interest.

The experimental results are presented below in Section IV.

IV. EXPERIMENTAL RESULTS

The experimental results section follows the organization of Sections II and III. Section IV-A describes the *Omni-Direction Camera and Data Setting*; Section IV-B describes the results *Visual and IR Index-Based Vegetation Detection*; Section IV-C, *IR Stream Segmentation Using Region-Based Thermal Analysis*, describes the results of the combination of the MNDVI and thermal region-based stream segmentation approaches, using region-based IR thresholding with thermal analysis and region

TABLE VII
CAMERA DATA SETTINGS

Dataset	Image Type	Size	Sensor
1	far-IR O-D	640 x 480	O-D IR
2	Visual Kinect	Various	Kinect Visual
3	O-D IR	640 x 480 Unwrapped to 181 x 1760	O-D IR
4	O-D Visual	640 x 480 Unwrapped to 181 x 1760	O-D Visual

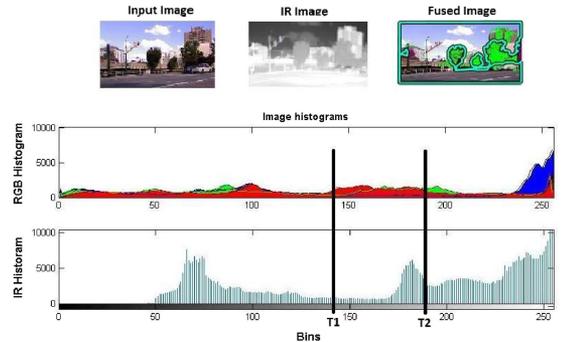


Fig. 11. Using datasets 1 and 2. A representative input Kinect visual and O-D IR images. The first plot is the histogram distribution of the blue, green, and red bands of the input image shown above. The second plot is the gray scale histogram of the IR image. The thresholds are overlaid on the plot.

growing threshold segmentation fusion of O-D IR and visual stream.

A. Omni-Direction Camera and Data Setting

Data were captured from the O-D far-IR camera and Kinect camera systems; we later captured additional data from our O-D IR and O-D visual camera. The datasets are presented in Table VII. We then processed the data with the two approaches, the MNDVI and the TRF fused, and compared their results.

A representative sample was used from dataset 1 (IR) and dataset 2 (Kinect) to produce Fig. 11. O-D IR and O-D visual samples from datasets 3 and 4 were used to produce Fig. 12. In Figs. 11 and 12, the visual image, the IR image, and the thermal-fused image are shown along with the RGB and IR histograms. Both figures have the vegetation region threshold values superimposed. Figs. 11 and 12, second plot, shows the spectral bins in the thermal signature.

Fig. 11, using datasets 1 and 2, shows the relationship between the red, blue, and green bands and IR grayscale bins for a representative sample of a segment of the unwrapped O-D IR, and Kinect images. It also shows the fused image, and the histograms have the thresholds overlaid.

Fig. 12, using datasets 3 and 4 (O-D IR and O-D visual), shows the same relationship for a representative unwrapped O-D color and O-D IR sensors along with the fused image showing the detected vegetation region. Again, in Fig. 12, the RGB and grayscale histograms with thresholds overlaid are shown.

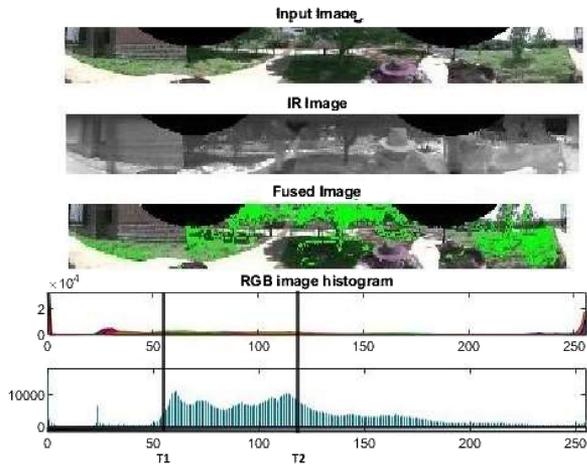


Fig. 12. Using datasets 3 and 4. A representative unwrapped input O-D visual and O-D IR images. The first plot is the histogram distribution of the blue, green, and red bands of the input image shown above. The second plot is the gray scale histogram of the IR image with the thresholds annotated.

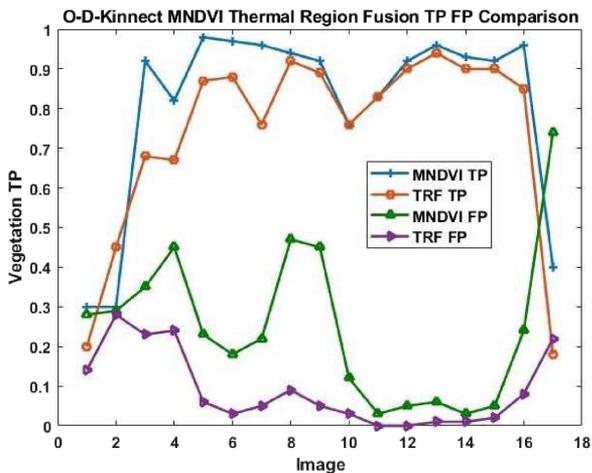


Fig. 13. O-D IR with Kinect cameras (datasets 1 and 2) comparison of MNDVI and TRF true positive compared to false positive rates.

B. Visual and IR Index-Based Vegetation Detection

In this section, we present the results from the index-based MNDVI approach. One of the key issues with the current implementations of vegetation index-based processes is the FPs that it produces. It particularly struggles with synthetic materials that have high red absorption. The vegetation index approach has known failures in areas such as manmade materials and paints that have high red absorption and behave similar to vegetation. Since the index-based approach compares the difference of the red band with the IR it can be seen how the MNDVI approach will be confused by this information. Applying MNDVI index-based vegetation detection alone has a high incidence of FPs.

Figs. 13 and 14 compare MNDVI true positives (TPs) with TRF TP and MNDVI FP with TRF FP for the two different camera setups. We can see the relationship of TPs versus FPs in Fig. 13 for the O-D IR and Kinect cameras using datasets 1 and 2. We can see the relationship of TPs versus FPs in Fig. 14 for the O-D IR and O-D visual cameras using datasets 3 and 4.

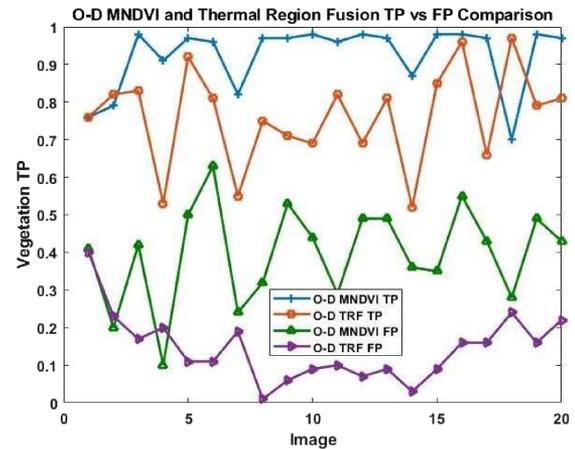


Fig. 14. O-D IR and O-D visual cameras (datasets 3 and 4) comparison of MNDVI and TRF true positive to false positive rates.

TABLE VIII
THERMAL SEGMENTED REGION FUSION RESULTS

Attribute	Comparison of MNDVI to TRF				
	Data Set	True % Positive	False % Positive	Ratio FP/TP	Peak FP/TP
<i>MNDVI</i>	1-2	81.12	25.00	0.3082	0.7400
<i>MNDVI</i>	3-4	92.35	39.72	0.4301	0.6300
<i>MNDVI</i>	Avg.	86.74	32.5	0.3731	0.6850
<i>TRF</i>	1-2	74.00	9.00	0.1216	0.2800
<i>TRF</i>	3-4	76.31	14.49	0.1899	0.3990
<i>TRF</i>	Avg.	75.16	11.75	0.1563	0.3940

It can be seen in Table VIII that the MNDVI approach has an average accuracy of 87.4% of the TPs compared to the FPs at 32.5% with some as high as 40%–70%. The visual MNDVI vegetation detection is shown in Row 1 of Table VIII, and the fusion of the index-based vegetation detection and the IR thermal-based region fusion TRF results are shown in Row 2 of Table VIII. As a result, we need a more robust method to detect the vegetation.

The MNDVI approach has an average TP rate of 86.75% averaged across all four datasets. However, the FP rate (FPR) is high with an average of 32.5% and a peak of 68.5%. The TRF approach has an average overall accuracy of 75.16% for the TPs compared to the average FPs at 11.5% with a peak as high as 40%. The TRF method has a better FP rejection rate, but has less accurate recognition. The TRF TP rate is mid-range of the NDVI spread in Table IX and therefore this is not considered a significant detractor. The authors believe that this degradation in performance is a result of the camera not being co-linear with the same viewing axis. This could be improved with a better choice and design of the camera. This could be further improved by the application of an in-depth learning approach to the vegetation detection problem.

TABLE IX
OTHER METHOD COMPARISON RESULTS

Comparison of Other Methods to our MNDVI to TRF				
Attribute	True Positive	% False Positive	Ratio FP/TP	Peak FP/TP
Reflectance NDVI [51]	95.2	32.5	0.3412	0.8124
MODIS-NDVI [52]	72.6	35.5	0.4889	0.5510
NDVI[53]	47.7	61.29	1.2849	1.7442
MNDVI datasets 1 - 4	86.74	32.5	0.3747	0.8623
TRF datasets 1 - 4	75.16	11.5	0.1533	0.3940

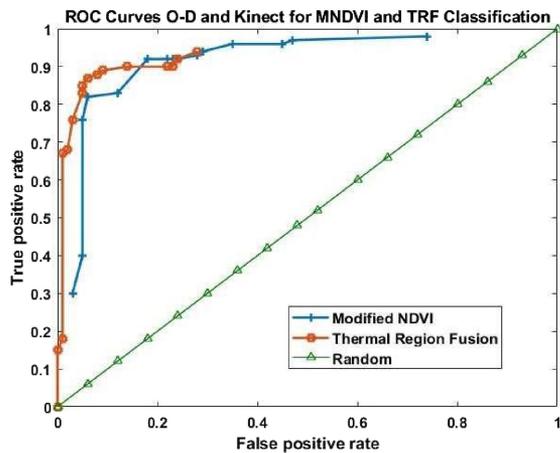


Fig. 15. ROC curves for the MNDVI and TRF approaches using the OD IR and Kinect cameras (datasets 1 and 2). It can be seen that the TRF has fewer false detects but slightly worse recognition of the vegetation areas.

C. IR Stream Segmentation Using Region-Based Thermal Analysis

The IR TRF results are shown in Figs. 13 and 14. Table VIII shows the comparison of the MNDVI and TRF approach for the overall experiment averaging the O-D Kinect and O-D IR and O-D visual camera setup. Table VIII summarizes the results for the thermal region-based analysis segmentation and fusion. The segmentation finds the regions of similar temperature, which can then be fused with the index-based results in Section IV-D.

The ROC curves for the MNDVI and the TRF fused visual and far-IR streams to extract salient vegetation features are presented in Figs. 15 and 16.

The TRF approach was better at FPs but had a lower true recognition rate, identifying the vegetation regions with fewer FPs than the baseline MNDVI approach. The positive impact of using TRF is that it effectively captures the vegetation pattern and fuses the two input streams from the O-D IR and vision cameras while rejecting FPs.

The better overall performance is demonstrated by less FPs and reasonable computation time. The receiver operating curve (ROC) plots the TP rate (TPR) or (sensitivity) against FPR

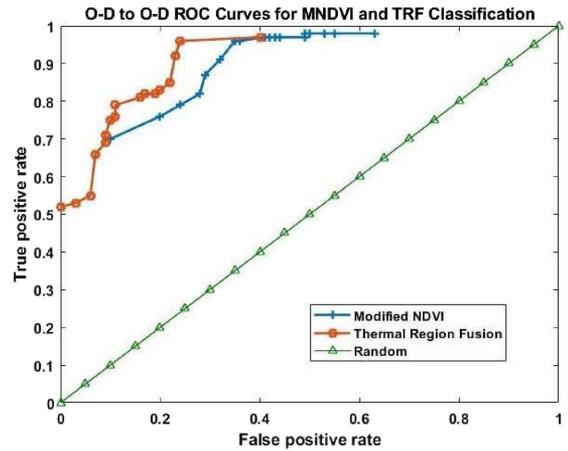


Fig. 16. ROC curves for the MNDVI and TRF approaches using the O-D IR and O-D visual cameras (datasets 3 and 4). It can be seen that the TRF has fewer false detects but slightly worse recognition of the vegetation areas.

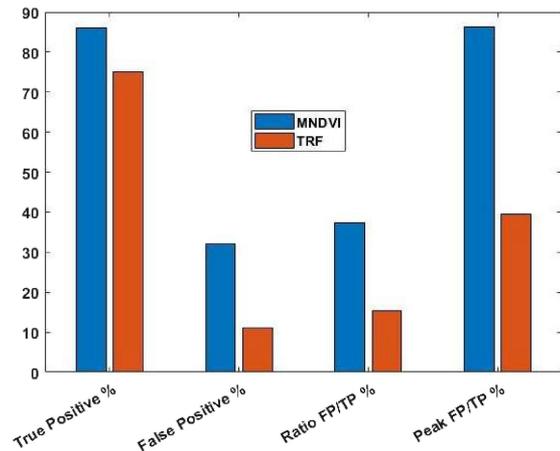


Fig. 17. MNDVI and TRF fusion results (average of datasets) showing % true positive (TP), percent false positive (FP), and ratio FP/TP and peak FP/TP.

or (1-specificity). The ROC is commonly used to visualize the performance of a binary classifier. Fig. 15 shows the ROC curves for the MNDVI compared to the segmented TRF for the O-D IR and Kinect camera setup (datasets 1 and 2), and Fig. 16 captures the ROC curves for the O-D IR and O-D visual camera setup (datasets 3 and 4). It can be seen that the TRF thermal region segmentation also does not capture as much of the vegetation region as the MNDVI, but has fewer false detects. The ratio of FP to TP for MNDVI is 0.37, where TRF is 0.15. This again highlights the need for a new fusion method. It can be seen that the relationship between TP and FP is best for the TRF fusion approach.

Table IX lists other methods and how we relate in this context. Our accuracies are consistent with the average of the other methods. However, our FPRs are reduced. The first three references used provided no data from which we could estimate the FPR, so the authors found three additional references with data. Moreover, we have referenced their results in Table IX.

Fig. 17 summarizes the results in a bar chart showing the average percent across the datasets for TP, percent FP, and the ratio

of FP/TP along with the peak FP/TP. Overall, the TRF approach performs the best. However, the results still need improvement, leading us to the need for an in-depth learning approach as our future research.

V. CONCLUSION

We presented a TRF Fusion approach for O-D IR and vision stream and compared the results to our MNDVI approach and a fusion of the MNDVI and IR thermal region-based approaches to detect and classify vegetation. Table VIII summarizes the baseline sensor fusion results with MNDVI having an 86.73% detect rate, but it had a 32.5% FPR. The IR thermal region fused with the visual index based lowered the false detect rate to 11.5%, but was not as good at detecting vegetation dropping to 75.16%. The method of determining the thermal region was a threshold region growing and segmentation. We demonstrated a 64% improvement in FP but resulted in a 14.5% reduction in TP with our TRF fusion approach. The reduction in TP results is still mid-range of the NDVI results presented in [49] and Table IX. The authors hypothesize that the reduction in TP recognition was due to the two cameras not having the same vertical viewing axis. Also, the IR and visual cameras did not have the same vertical viewing angle and could not be fully registered. While our TRF approach made a significant reduction in FPs, it lost ground in TP recognition. However, the authors feel that these proof of concept results merit further work in this area. The approach had reasonable performance in varying lighting conditions, but was not robust to the different viewing angle of the two cameras. This demonstrates a need to use a learning approach that will address the issue of camera quality.

The follow on work will fuse the O-D IR and O-D visual streams utilizing the application of deep learning through convolution neural networks and context-based reasoning algorithms to discriminate between objects in a spatial scene (grass-wall, rock-bush, puddle-hole, and door-window). This will then be applied to a robot platform for object detection classification and tracking.

REFERENCES

- [1] H. Anderson, A. Treptow, and T. Duckett, "Self-localization in non-stationary environments using omni-directional vision," *Robot. Auton. Syst.*, vol. 55, pp. 541–551, 2007.
- [2] G. Lopez-Nicolas, J. Guerrero, and C. Sagues, "Multiple homographies with omnidirectional vision for robot homing," *Robot. Auton. Syst.*, vol. 58, no. 6, pp. 773–783, 2010.
- [3] W. Lui and R. Jarvis, "Eye-full tower: A GPU-based variable multi-baseline omnidirectional stereovision system with automatic baseline selection for outdoor mobile robot navigation," *Robot. Auton. Syst.*, vol. 58, no. 6, pp. 747–761, 2010.
- [4] J. Lim and N. Barnes, "Estimation of the epipole using optical flow at antipodal points," *Comput. Vision Image Understanding*, vol. 114, no. 2, pp. 245–253, 2010.
- [5] D. Scaramuzza, F. Fraundorfer, and M. Pollefeys, "Closing the loop in appearance-guided omnidirectional visual odometry by using vocabulary trees," *Robot. Auton. Syst.*, vol. 58, no. 6, pp. 820–827, 2010.
- [6] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A robust descriptor for tracking vertical lines in omnidirectional images and its use in mobile robotics," *Int. J. Robot. Res.*, vol. 28, no. 2, pp. 149–171, 2009.
- [7] S. Aksoy, "Spatial techniques for image classification," in *Signal and Image Processing for Remote Sensing*. New York, NY, USA: Taylor and Francis, 2016, pp. 491–513.
- [8] O. Helwich and C. Wiedemann, "Object extraction from high-resolution multi-sensor image data," *Image Signal Process. Remote Sens. Int. Soc. Opt. Eng.*, vol. 3871, pp. 284–295, 1999.
- [9] D. Landgrebe, "Information extraction principles and methods for multi-spectral and hyperspectral image data," in *Information Processing for Remote Sensing*. Singapore: World Scientific, 1998.
- [10] Y. J. Kaufman and D. Tanré, "Atmospherically resistant vegetation index (ARVI) for EOS-MODIS," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 2, pp. 261–270, Mar. 1992.
- [11] H. Q. Liu and A. R. Huete, "A feedback modification of the NDVI to minimize canopy background and atmospheric noise," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 2, pp. 457–465, Mar. 1995.
- [12] M. Verstraete and B. Pinty, "Designing optimal spectral indices for remote sensing applications," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 5, pp. 1254–1265, Sep. 1996.
- [13] C. Ünsalan and K. Boyer, "Linearized vegetation indices based on a formal statistical framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1575–585, Jul. 2004.
- [14] M. Bunkei, Y. Wei, Y. C. Jin, O. Yuichi, and Q. Guoyu, "Sensitivity of the enhanced vegetation index (EVI) and normalized difference vegetation index (NDVI) to topographic effects: A case study in high-density cypress forest," *Sensors*, vol. 7, no. 11, pp. 2636–2651, 2007.
- [15] M. Bradley, S. M. Thayer, A. Stentz, and P. Rander, "Vegetation detection for mobile robot navigation," *Robot. Inst.*, Carnegie Mellon Univ., Pittsburgh, PA, USA, CMU-RI-TR-04-12, 2004.
- [16] D. Bradley, R. Unnikrishnan, and J. Bagnell, "Vegetation detection for driving in complex environments," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2007, pp. 503–508.
- [17] P. D'Odorico, A. Gonsamo, A. Damm, and M. Schaepman, "Experimental evaluation of sentinel-2 spectral response functions for NDVI time-series continuity," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1336–1348, Mar. 2013.
- [18] K. Bhoyar and O. Kakde, "Color image segmentation based on JND color histogram," *Int. J. Image Process.*, vol. 3, no. 6, pp. 283–292, 2010.
- [19] A. Briggs, C. Detweiler, P. Mullen, and D. Sharstein, "Scale space features in 1D omnidirectional images," in *Proc. Workshop Omnidirectional Vis., Camera Netw. Non-classical Cameras*, 2004, pp. 115–126.
- [20] A. Pretto, E. Minegatti, J. Yoshiaki, U. Ryuichi, and A. Tamio, "Image similarity based on discrete wavelet transform for robots with low-computational resources," *Robot. Auton. Syst.*, vol. 58, no. 7, pp. 879–888, 2010.
- [21] M. Taiana, J. Santos, J. Gaspar, J. Nascimento, A. Bernardino, and P. Lima, "Tracking objects with generic calibrated sensors: An algorithm based on color and 3D shape features," *Robot. Auton. Syst.*, vol. 58, no. 6, pp. 784–795, 2010.
- [22] A. Bagnell, D. Bradley, D. Silver, and B. Sofman, "Learning rough-terrain autonomous navigation," *IEEE Robot. Automat. Mag.*, vol. 17, no. 2, pp. 74–84, May 2011.
- [23] A. Rankin *et al.*, "Unmanned ground vehicle perception using thermal infrared cameras," *Proc. SPIE*, vol. 8045, 2011, Art. no. 804503.
- [24] D. Wolf and G. Sukhatme, "Activity-based semantic mapping of an urban environment," in *Experimental Robotics*. New York, NY, USA: Springer, 2008, pp. 321–329.
- [25] A. Mishra and Y. Alimonos, "Visual segmentation of simple objects for robots," in *Robotics: Science and Systems*, vol. 7. Cambridge, MA, USA: MIT Press, pp. 30–38, 2012.
- [26] D. Nguyen, L. Kuhnert, and K. Kuhnert, "Structure overview of vegetation detection. A novel approach for efficient vegetation detection using an active lighting system," *Robot. Auton. Syst.*, vol. 60, pp. 498–508, 2012.
- [27] T. Hoang, P. Duong, N. Van, D. Viet, and T. Vinh, "Multi-sensor perceptual system for mobile robot and sensor fusion-based localization," in *Proc. Int. Conf. Control, Automat., Inf. Sci.*, 2012, pp. 259–264.
- [28] C. Cimander, M. Carlsson, and C. F. Mandenius, "Sensor fusion for on-line monitoring of yoghurt fermentation," *J. Biotechnol.*, vol. 99, pp. 237–248, 2002.
- [29] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern Recognit. Lett.*, vol. 24, pp. 2115–2125, 2003.
- [30] M. A. K. Jaradat and R. Langari, "A hybrid intelligent system for fault detection and sensor fusion," *Appl. Soft. Comput.* vol. 9, pp. 415–422, 2009.
- [31] C. Aliustaoglu, H. M. Ertunc, and H. Ocak, "Tool wear condition monitoring using a sensor fusion model based on fuzzy inference system," *Mech. Syst. Signal Process.*, vol. 23, pp. 539–546, 2009.

- [32] S. Tangjitsitcharoen and C. Rungruang, "In-process monitoring and estimation of tool wear on CNC turning by applying multi-sensor with back propagation technique," *Adv. Mater. Res.*, vol. 291, pp. 3036–3043, 2011.
- [33] Q. Cheng, P. K. Varshney, J. H. Michels, and C. M. Belcastro, "Distributed fault detection with correlated decision fusion," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 45, no. 4, pp. 1448–1465, Oct. 2009.
- [34] K. Liu and S. Huang, "Integration of data fusion methodology and degradation modeling process to improve prognostics," *IEEE Trans. Automat. Sci. Eng.*, vol. 13, no. 1, pp. 344–354, Jan. 2016.
- [35] N. Ahmed, E. Sample, and M. Campbell, "Bayesian multi-categorical soft data fusion for human–robot collaboration," *IEEE Trans. Robot.*, vol. 29, no. 1, pp. 189–206, Feb. 2013.
- [36] O. Beyca, P. K. Rao, Z. Kong, S. Bukkapatnam, and R. Komanduri, "Heterogeneous sensor data fusion approach for real-time monitoring in ultra-precision machining (UPM) process using non-parametric bayesian clustering and evidence theory," *IEEE Trans. Automat. Sci. Eng.*, vol. 13, no. 2, pp. 1033–1044, Apr. 2016.
- [37] B. Schäfer, C. Armbrust, T. Föhst, and K. Berns, "The application of design schemata in off-road robotics," *IEEE Intell. Transp. Syst. Mag.*, vol. 5, no. 1, pp. 4–27, Spring 2013.
- [38] Z. Kang, J. Yang, and R. Zhong, "A bayesian-network-based classification method integrating airborne LiDAR data with optical images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1651–1661, Apr. 2017.
- [39] D. Dutta *et al.*, "Characterizing vegetation canopy structure using airborne remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 1160–1178, Feb. 2017.
- [40] C.-H. Chen, "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.
- [41] N. Robertson and J. Letham, "Contextual person detection in multi-modal outdoor surveillance," in *Proc. 20th Eur. Signal Process. Conf.*, 2012, pp. 1930–1934.
- [42] D. Andújar, J. Dorado, C. Fernández-Quintanilla, and A. Ribeiro, "An approach to the use of depth cameras for weed volume estimation," *Sensors*, vol. 16, no. 972, pp. 1–11, 2016.
- [43] V. Poulain, J. Inglada, M. Spigai, J.-Y. Tourneret, and P. Marthon, "High-resolution optical and SAR image fusion for building database updating," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 2900–2910, Aug. 2011.
- [44] M. Jiayi, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Inf. Fusion*, vol. 31, pp. 100–109, 2016.
- [45] H. Li, H. Qiu, Z. Yu, and Y. Zhang, "Infrared and visible image fusion scheme based on NSCT and low-level visual features," *Infrared Phys. Technol.*, vol. 76, pp. 174–184, 2016.
- [46] M. Mancas, B. Gosselin, and M. Benoit, "Segmentation using a region growing thresholding," *Proc. SPIE*, vol. 5672, pp. 388–398, 2006.
- [47] B. Russell, A. Torralba, K. Murphy, and W. T. Freeman, "LabelMe: A database and web-based tool for image annotation," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 157–173, 2007.
- [48] S. Prasad, H. Wu, and J. E. Fowler, "Compressive data fusion for multi-sensor image analysis," in *Proc. Int. Conf. Image Process.*, Paris, France, Oct. 2014, pp. 5032–5036.
- [49] A. Zaitunah *et al.*, "Normalized difference vegetation index (NDVI) analysis for land cover types using Landsat 8 oli in besitang watershed, Indonesia," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 126, pp. 1–10, 2018, Art. no. 012112.
- [50] T. Sankey, J. McVay, T. Swetnam, M. McClaran, P. Heilman, and M. Nichols, "UAV hyperspectral and LIDAR data and their fusion for arid and semi-arid land vegetation monitoring," *Remote Sens. Ecol. Conserv.*, vol. 4, no. 1, pp. 20–33, 2018.
- [51] L. Leite, L. Carvalho, and F. Silva, "Change detection in forests and savannas using statistical analysis based on geographical objects," *Boletim Ciências Geodesicas*, vol. 23, no. 2, pp. 284–295, 2017.
- [52] L. Reymondina *et al.*, "A methodology for near real-time monitoring of habitat change at continental scales using MODIS-NDVI and TRMM," *Terra-i-Method, Report*, pp. 1–33, Dec. 2012.
- [53] R. Carlà, L. Santurri, L. Bonora, and C. Conese, "Multi-temporal burnt area detection methods based on a couple of images acquired after the fire event," *Proc. SPIE*, vol. 7472, 2011, Art. no. 74720C, doi: [10.1117/12.832908](https://doi.org/10.1117/12.832908).



David L. Stone (M'10) received the B.S. Eng. degree in electrical and electronic engineering from Purdue University, West Lafayette, IN, USA, in 1973, the M.S. Eng. degree in mechanical engineering and ocean engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1982.

He currently works full time with the Naval Surface Warfare Center Dahlgren, Dahlgren, VA, USA. His research interests include deep neural networks and intelligent robotic perception.



Guarav Shah (M'13) received the B.Com. degree in economics and accounting from The Maharaja Sayajirao University of Baroda, Vadodara, India, in 2005, and the B.S. degree in computer engineering from Virginia Commonwealth University, Richmond, VA, USA, in 2014.

He is currently a Web Integration Developer with Bon Secours Health Systems, Richmond. His research interests include web, mobile, and data integration in health care and backend and database programming.



Yuichi Motai (S'00–M'03–SM'12) received the B.E. degree in instrumentation engineering from Keio University, Tokyo, Japan, in 1991, the M.E. 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 research interests include the broad area of sensory intelligence, particularly

in medical imaging, pattern recognition, computer vision, and sensory-based robotics.



Alex J. Aved (M'13–SM'15) received the B.A. degree in computer science and mathematics from Anderson University, Anderson, Indiana, in 1999, the M.S. degree in computer science from Ball State University, Muncie, Indiana, in 2001, and the Ph.D. degree in computer science (*focus area: real-time multimedia databases*) from the University of Central Florida, Orlando, FL, USA, in 2013.

He is currently a Technical Advisor with the Air Force Research Laboratory Information Directorate, Rome, NY, USA. His research interests include multimedia databases, stream processing (via CPU, GPU or coprocessor), and dynamically executing models with feedback loops incorporating measurement and error data to improve the accuracy of the model.