Multi-Resolution Multi-Modal Sensor Fusion For Remote Sensing Data With Label Uncertainty

Xiaoxiao Du[a], Alina Zare[b,*]

[a]*Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211*
[b]*Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611*

Abstract

In remote sensing, each sensor can provide complementary or reinforcing information. It is valuable to fuse outputs from multiple sensors to boost overall performance. Previous supervised fusion methods often require accurate labels for each pixel in the training data. However, in many remote sensing applications, pixel-level labels are difficult or infeasible to obtain. In addition, outputs from multiple sensors may have different levels of resolution or modalities (such as rasterized hyperspectral imagery versus LiDAR 3D point clouds). This paper presents a Multiple Instance Multi-Resolution Fusion (MIMRF) framework that can fuse multi-resolution and multi-modal sensor outputs while learning from ambiguously and imprecisely labeled training data. Experiments were conducted on the MUUFL Gulfport hyperspectral and LiDAR data set and a remotely-sensed soybean and weed data set. Results show improved, consistent performance on scene understanding and agricultural applications when compared to traditional fusion methods.

Keywords: sensor fusion, multi-resolution, multi-modal fusion, multiple instance learning, scene understanding, label uncertainty, hyperspectral, LiDAR, Choquet integral.

[*]Corresponding author

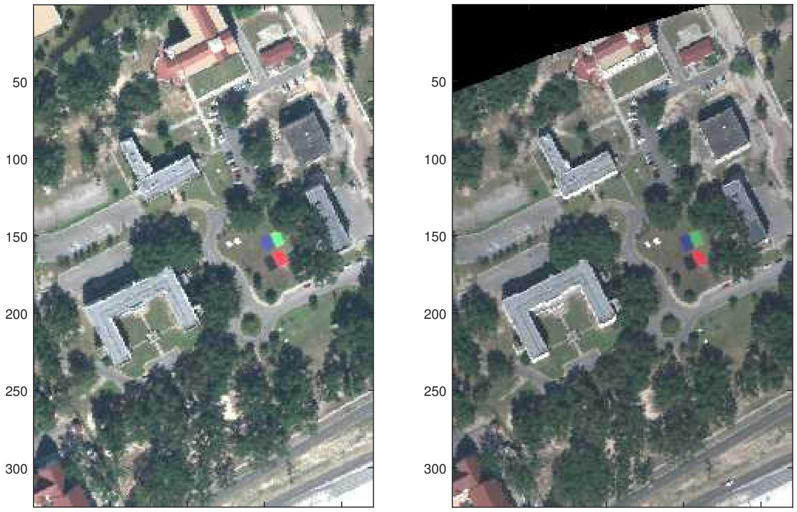
Email addresses: xdy74@mail.missouri.edu (Xiaoxiao Du), azare@ece.ufl.edu (Alina Zare)

1. Introduction and Motivation

In multi-sensor fusion, each sensor may provide complementary and reinforcing information that can be helpful in target detection, classification, and scene understanding [1]. It is useful to integrate and combine such outputs from multiple sensors to provide more accurate information over a scene. In this study, hyperspectral (HSI) and LiDAR data collected by us over the University of Southern Mississippi-Gulfpark campus were used [2, 3]. Figure 1 shows an RGB image over the scene. Figure 2 shows a scatter plot of the LiDAR point cloud data. As can be seen, there are a wide variety of materials in the scene, such as buildings, road, etc. The task is to use the HSI and LiDAR data we collected to classify materials in the scene.

There are several advantages to fuse both HSI and LiDAR data rather than using individual HSI and LiDAR information for this task. If a road and a building rooftop are built with the same material (say, asphalt), hyperspectral information alone may not be sufficient to tell the rooftop and the road apart. However, the LiDAR data provides elevation information and can easily distinguish the two. On the other hand, a highway and a biking trail can be at the same elevation and using LiDAR data alone may not be sufficient to distinguish the two types of roads, but hyperspectral sensor can be very helpful in identifying the distinctive spectral characteristics between a highway, which is likely to be primarily asphalt, and a biking trail, which is likely to be covered in dirt. It would be, thus, valuable to fuse information from both HSI and LiDAR sensors to obtain a better classification result and a more comprehensive understanding of a scene [4, 5, 6, 7].

There are four challenges associated with this fusion task. First, the HSI and LiDAR data we have are of different modalities. The HSI imagery is on a pixel grid and the raw LiDAR data is a point cloud. Many previous HSI and LiDAR fusion methods have been proposed in the literature [8, 9, 10, 11]. However, nearly all of these fusion methods for HSI and LiDAR rely on having a rasterized LiDAR image with accurate image registration. In these methods, the LiDAR point cloud data was mapped in the same grid as the hyperspectral imagery and the fusion methods work with both HSI and LiDAR data in image form. However, the raw dense LiDAR point cloud can offer



(a)

(b)

Figure 1: RGB images of the MUUFL Gulfport data. (a) Campus 1; (b) Campus 2. The black region in the top left corner of (b) is an invalid region from the data collection and was excluded from any training or testing process in the experiment. The "campus 1" and "campus 2" refer to data from two flights (more details please see Section 4.1).

higher geographic accuracy [12] as the raw data does not depend on grid size. Besides, image registration and rasterization may introduce additional inaccuracy [13, 14]. It would be valuable to develop a multi-modal fusion algorithm that can handle directly the HSI imagery and raw LiDAR point cloud.

Second, the HSI and LiDAR data have different resolutions. The hyperspectral imagery we collected have a ground sample distance of $1m$. That is to say, each pixel in the hyperspectral imagery covers $1m^2$ area. The LiDAR data, on the other hand, have a higher resolution with 0.60m cross track and 0.78m along track spot spacing. There is more than one LiDAR data point inside a $1m^2$ area. This means, for each pixel in the hyperspectral imagery, there are more than one LiDAR data point to which it corresponds. It would be, thus, important to develop a fusion method that can handle multi-resolution outputs from multi-modal sensors.

Third, there are inaccuracies of LiDAR measurements in the given data set after

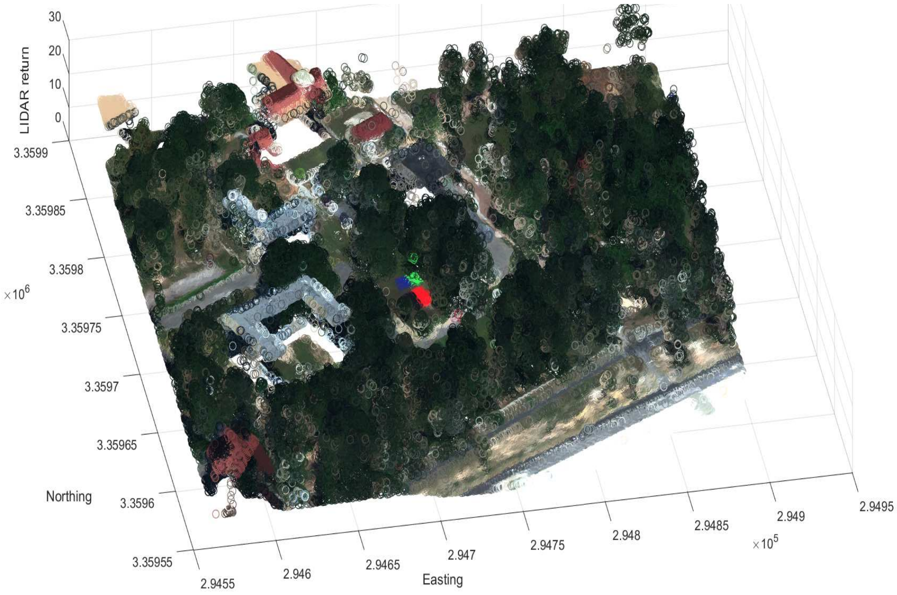


Figure 2: An example of 3D scatterplot of LiDAR data over the University of Southern Mississippi - Gulfpark campus. The LiDAR points were colored by the RGB imagery provided by HSI sensors over the scene. “X” and “Y”-axes are Easting and Northing locations and “Z”-axis corresponds to the elevation information.

rasterization. For example, if we look at building edges on the grey roofs, some of the building points have similar elevation as the neighborhood grass pixels on the ground, which is obviously wrong. Some points on the sidewalks and dirt roads, on the other hand, have high elevation values similar to nearby trees and buildings, which is also wrong. Such incorrect information can be caused by a variety of factors such as poor alignment and rasterization to HSI grid, sensor inaccuracy, or missed edge points from the laser pulse discontinuousness in LiDAR [15]. Directly using such inaccurate measurements for fusion can cause further inaccuracy or error in classification, detection, or prediction. Therefore, we must develop a fusion algorithm that is able to handle such inaccurate/imprecise measurements.

Finally, to train a supervised fusion algorithm, training labels are needed. However, in this data set, accurate pixel-level labels for both HSI and LiDAR data are not available. There are more than 70,000 pixels in one HSI image and more than 170,000 data points in one LiDAR point cloud. It would be expensive and difficult for humans to manually label each pixel/data point. Even if we have experts manually label each pixel or data point in the scene, there are certain transition areas that are impossible to label. For example, at the edge of a building, it is impossible for humans to visually identify which pixel belongs to the building and which pixel belongs to the ground surface nearby. Besides, as mentioned above, a pixel in the HSI image corresponds to $1m^2$. It is highly likely that a pixel at the edge of the building is mixed with both building and ground surface materials. In this case, it is impossible to provide accurate labels for each pixel/data point in the training data. However, in this example, one can easily circle a region in the scene that contains a building. That is to say, it is possible to identify a region or a set of pixels that contains the target material in the scene. We call such a region or a set of pixels a "bag", based on the Multiple Instance Learning (MIL) framework [16]. Previous supervised fusion methods often require accurate pixel-level labels and cannot handle "bag-level" labels [17, 18]. We aim to develop a trained fusion algorithm that can handle such uncertainty in training labels.

This paper proposes a Multiple Instance Multi-Resolution Fusion (MIMRF) framework to address the above four challenges. The proposed MIMRF can fuse multi-resolution and multi-modal sensor outputs from multiple sensors while effectively learning from bag-level training labels.

2. Related Work

The proposed MIMRF algorithm handles label uncertainty in training data by formulating the supervised fusion problem under the Multiple Instance Learning (MIL) framework. The proposed MIMRF algorithm is based on the Choquet integral, an effective nonlinear fusion tool widely used in the literature. In this section, related work on multi-sensor fusion and the MIL framework is described. The basis of the proposed MIMRF fusion, the Choquet integral, is also described in this section.

2.1. Multi-Resolution and Multi-Modal Sensor Fusion

Existing optical sensors produce data at varying spatial, temporal and spectral resolutions [19]. Those sensor outputs may also have different modalities, such as imagery and point clouds. Multi-resolution and multi-modal fusion methods have been studied and developed in the literature in order to better fuse such information from different sensors. Multi-sensor fusion have wide applications in remote sensing such as the extraction of urban road networks [20], building detection [21], precision agriculture [22], and anomaly detection in archaeology [23].

Nearly all previous multi-resolution fusion methods focus on fusing image data only. Pan-sharpening methods, for example, use a panchromatic image with higer spatial resolution to fuse with multi- or hyper-spectral images in order to obtain images with higher spectral and spatial resolution. However, methods like pan-sharpening only handles imagery and mostly focus on fusing only two images (a panchromatic and a multispectral or hyperspectral image) [24, 25, 26, 27, 28, 29, 30, 31].

More specifically, regarding HSI and LiDAR fusion, previous fusion methods can only work with rasterized LiDAR images [8, 9, 10, 11]. In these methods, the LiDAR point cloud data was mapped in the same grid as the hyperspectral imagery and the fusion methods work with both HSI and LiDAR data in image form. However, as discussed in Introduction, the raw dense LiDAR point cloud can offer higher geographic accuracy and alleviate the necessity to geometrically align the HSI and LiDAR data [32, 33]. We are not aware of any previous methods that can directly handle HSI imagery and raw LiDAR point clouds with different resolutions without rasterization or alignment.

2.2. Multiple Instance Learning

As discussed in Introduction, the proposed MIMRF needs to be able to handle bag-level training labels rather than pixel-level labels. Here, this label uncertainty problem is formulated under the Multiple Instance Learning (MIL) framework.

The MIL framework was first proposed in [16] to deal with uncertainties in training labels. In the MIL framework, training labels are associated with sets of data points ("bags") instead of each data point ("instance"). In two-class classification, the standard MIL assumes that a bag is labeled positive if at least one instance in the bag is positive and a bag is labeled negative if all the instances in the bag are negative. The MIL has wide applications in natural scene classification [34, 35], human action recognition in videos [36], object detection and tracking [37, 38, 39], context identification, and context-dependent learning in remote sensing data [40, 41].

The mi-SVM algorithm is a widely cited MIL method for classification [42, 43, 44] and will later be used as a comparison method in the experiments. The mi-SVM algorithm was proposed by Andrews et al. [45] as an MIL extension to support vector machine (SVM) learning approaches. The mi-SVM algorithm can work with bag-level labels and learns a linear discriminate function to separate the positive from the negative classes.

2.3. Fuzzy Measure and Choquet Integral

The proposed MIMRF algorithm uses the Choquet integral (CI) [46] to perform fusion. The CI is a powerful non-linear fusion and information aggregation framework and has wide applications in the literature [47, 48, 49, 50, 51]. In the field of remote sensing, the CI has been applied to multi-sensor fusion in landmine detection [1, 52], classifier fusion [53], and target detection using Landsat and hyperspectral imagery [54, 53] .

Compared with commonly used aggregation operators such as weighted arithmetic means [55], the CI is able to model complex, non-linear relationship amongst the combinations of the sources for fusion. The Choquet integral depends on "fuzzy measures", whose values determine the fusion outcome. Depending on the set of real-valued fuzzy measure $\mathbf{g}$ it learns, the CI can flexibly represent a wide variety of aggregation operators [56, 57, 49, 1, 50, 53].

Suppose we are fusing m sensor outputs using the discrete Choquet integral. These sensor outputs are called "sources" for fusion. Denote $S = \{s_1, s_2, \ldots, s_m\}$ for m sources to be fused. The power set of all (crisp) subsets of S is denoted 2^S. A monotonic and normalized fuzzy measure, $\mathbf{g}$, is a real valued function that maps $2^S \to [0, 1]$ [46, 58, 59, 53]. The fuzzy measure used in this paper satisfies monotonicity and normalization properties, i.e. $g(A) \leq g(B)$ if $A \subseteq B$ and $A, B \subseteq C$; $g(\emptyset) = 0$; and $g(S) = 1$ [60]. Therefore, the power set has $2^m - 1$ non-empty crisp subsets, and each element in the fuzzy measure corresponds to one of the subset. In this paper, the fuzzy measure $\mathbf{g}$ can be written as a vector of length $(2^m - 1)$. Denote $h(s_k; \mathbf{x}_n)$ as the k^{th}

sensor output for the n^{th} data point, the discrete Choquet integral on instance $\mathbf{x}_n$ given sensor outputs S is then computed as [61, 60, 53]:

$$C_{\mathbf{g}}(\mathbf{x}_n) = \sum_{k=1}^{m} \left[h(s_k;\mathbf{x}_n) - h(s_{k+1};\mathbf{x}_n)\right] g(A_k), \tag{1}$$

where S is sorted so that $h(s_1;\mathbf{x}_n) \geq h(s_2;\mathbf{x}_n) \geq \cdots \geq h(s_m;\mathbf{x}_n)$ and $h(s_{m+1};\mathbf{x}_n) \equiv 0$. The term $g(A_k)$ is the fuzzy measure element that corresponds to the subset $A_k = \{s_1, \ldots, s_k\}$.

In the supervised fusion problem, we need to learn the unknown fuzzy measure, $\mathbf{g}$, given training data and known training labels. The learned fuzzy measure $\mathbf{g}$ can then be used to determine the fusion results for test data. In the literature, the fuzzy measure values can be learned using either quadratic programming or sampling techniques. The CI-QP (quadratic programming) approach [60] learns a fuzzy measure for Choquet integral by optimizing a least squares error objective using quadratic programming [62]. However, the CI-QP method requires pixel-level training labels and cannot work for MIL problems. We previously proposed a Multiple Instance Choquet Integral (MICI) classifier fusion framework that extends the standard CI fusion under the MIL framework [53, 63, 64]. However, the previous MICI models still assumes that each sensor sources must have the same number of data points and do not support multi-resolution fusion. It would be interesting, therefore, to develop a novel trained classifier fusion algorithm that can both handle multi-resolution data and learn from uncertain and imprecise training labels.

3. The Multiple Instance Multi-Resolution Fusion (MIMRF) Algorithm

In this section, the proposed Multiple Instance Multi-Resolution Fusion (MIMRF) algorithm is described. The proposed MIMRF algorithm learns a monotonic, normalized fuzzy measure from training data and bag-level training labels. The learned fuzzy measure is used with the Choquet integral to perform multi-resolution fusion.

3.1. Objective Function

We already discussed in Introduction that in our HSI/LiDAR fusion problem, due to the difference in modality and resolution, there are more than one LiDAR data point that correspond to each pixel in the HSI imagery. There may also be inaccuracies in the LiDAR data. Our proposed MIMRF algorithm aims to handle such challenges. We motivate our objective function with a simple example, as follows.

Suppose there are three LiDAR data points that correspond to one pixel in the HSI imagery, as shown in pink in Figure 3. Denote the value of the i^{th} pixel in the HSI imagery as H_i. Denote the values of the LiDAR points as L_{i1}, L_{i2}, L_{i3}. All possible combination of the sensor outputs for pixel i can be written as:

$$\mathbf{S}_i = \begin{bmatrix} H_i & L_{i1} \\ H_i & L_{i2} \\ H_i & L_{i3} \end{bmatrix}. \tag{2}$$

Here, $\mathbf{S}_i$ is called a "collection" of all possible combinations of all outputs from all sensors for the i^{th} data point. By using the collection of all possible combinations in this way, we can handle multi-resolution sensor outputs with unmatched number of data points.

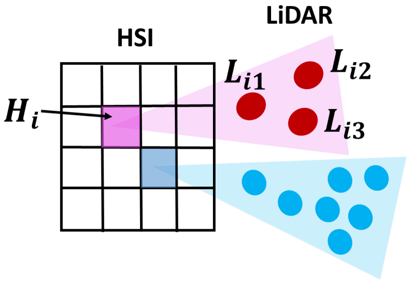


Figure 3: Illustration for HSI and LiDAR fusion. All LiDAR data points in the pink shade fell in the same area covered by the pink pixel in the HSI image, and all LiDAR data points in the blue shade corresponds to the blue pixel in the HSI image. Note that there can be more than one LiDAR points in the area covered by one pixel in the HSI imagery.

We assume that at least one of the LiDAR point is accurate (has the correct height information), but we do not know which one. Other LiDAR points may be erroneous or inaccurate (for example, on the edge of a building). Our proposed algorithm aims to automatically select the "correct" point with accurate information. To achieve this goal, we write the CI fusion for the collection of the sensor outputs of the i^{th} negative data point as:

$$C_{\mathbf{g}}(\mathbf{S}_i^-) = \min_{\forall \mathbf{x}_k^- \in \mathbf{S}_i^-} C_{\mathbf{g}}(\mathbf{x}_k^-); \tag{3}$$

and we write the CI fusion for the collection of the sensor outputs values of the j^{th} positive data point as:

$$C_{\mathbf{g}}(\mathbf{S}_j^+) = \max_{\forall \mathbf{x}_l^+ \in \mathbf{S}_j^+} C_{\mathbf{g}}(\mathbf{x}_l^+), \tag{4}$$

where $\mathbf{S}_i^-$ is the collection of sensor outputs for the i^{th} negative data point and $\mathbf{S}_j^+$ is the collection of sensor outputs for the j^{th} positive data point; $C_{\mathbf{g}}(\mathbf{S}_i^-)$ is the Choquet integral output for $\mathbf{S}_i^-$ and $C_{\mathbf{g}}(\mathbf{S}_j^+)$ is the Choquet integral output for $\mathbf{S}_j^+$. In this way, the min and max operators automatically select one data point (which is assumed to be the data point with correct information) from each negative and positive bag to be used for fusion, respectively.

Moreover, our objective function must be able to handle bag-level training labels. Suppose the desired label for a "negative" (non-target) data point is "0" and the desired label for a "positive" (target) data point is "1", for two-class classification problems. Recall that the MIL assumes a bag is labeled positive if at least one instance in the bag is positive and a bag is labeled negative if all the instances in the bag are negative. That is to say, we want to encourage all points in negative bags to have label "0" and at least one point in positive bags to have label "1". We can write the objective function for negative bags as:

$$\begin{aligned} J^- &= \sum_{a=1}^{B^-} \max_{\forall \mathbf{S}_{ai}^- \in \mathscr{B}_a^-} \left(C_{\mathbf{g}}(\mathbf{S}_{ai}^-) - 0 \right)^2 \\ &= \sum_{a=1}^{B^-} \max_{\forall \mathbf{S}_{ai}^- \in \mathscr{B}_a^-} \left(\min_{\forall \mathbf{x}_k^- \in \mathbf{S}_{ai}^-} C_{\mathbf{g}}(\mathbf{x}_k^-) - 0 \right)^2 . \end{aligned} \tag{5}$$

Similarly, the objective function for positive bags can be written as:

$$\begin{aligned} J^+ &= \sum_{b=1}^{B^+} \min_{\forall \mathbf{S}_{bj}^+ \in \mathscr{B}_b^+} \left(C_{\mathbf{g}}(\mathbf{S}_{bj}^+) - 1 \right)^2 \\ &= \sum_{b=1}^{B^+} \min_{\forall \mathbf{S}_{bj}^+ \in \mathscr{B}_b^+} \left(\max_{\forall \mathbf{x}_l^+ \in \mathbf{S}_{bj}^+} C_{\mathbf{g}}(\mathbf{x}_l^+) - 1 \right)^2 , \end{aligned} \tag{6}$$

where B^+ is the total number of positive bags, B^- is the total number of negative bags, $\mathbf{S}_{ai}^-$ is the collection of i^{th} instance set in the a^{th} negative bag and similar for $\mathbf{S}_{bj}^+$. $C_{\mathbf{g}}$

is the Choquet integral given fuzzy measure $\mathbf{g}$, $\mathscr{B}_a^-$ is the a^{th} negative bag, and $\mathscr{B}_b^+$ is the b^{th} positive bag. The term $\mathbf{S}_{ai}^-$ is the collection of input sources for the i^{th} pixel in the a^{th} negative bag and $\mathbf{S}_{bj}^+$ is the collection of input sources for the j^{th} pixel in the b^{th} positive bag. By minimizing J^- in Eq. (5), we encourage the fusion output of all the points in the the negative bag to the desired negative label 0. By minimizing J^+ in Eq. (6), we encourage the fusion output of at least one of the points in the positive bag to the desired positive label 1. This satisfies the MIL assumption and successfully handles label uncertainty.

Thus, the objective function for the proposed Multiple Instance Multi-Resolution Fusion (MIMRF) algorithm is proposed as follows:

$$\min_{\mathbf{g}} J = \min_{\mathbf{g}} \left(J^- + J^+ \right). \tag{7}$$

3.2. Optimization

An evolutionary algorithm is used to optimize the objective function in Eq. (7) and learn the fuzzy measure $\mathbf{g}$. The evolutionary algorithm used in this paper is similar to the method used in [53]. First, a population of fuzzy measures were generated. Each element in the fuzzy measure was initialized randomly to a set of values between $[0, 1]$. In each iteration, the valid interval of each fuzzy measure element in the population was computed. The "valid interval" of a fuzzy measure element is defined as how much "wiggle room" the measure element value can change while still satisfying the monotonicity property [53].

New measures are sampled based on either small-scale or large-scale mutations. In small-scale mutation, only one measure element is sampled according to their valid interval. In large-scale mutation, all the measure elements are sampled. A Truncated Gaussian distribution [65] is used for sampling new measure values. The rate of small-scale mutation versus large-scale mutation is a user-set parameter. Then, the old and new measure element valuess are pooled together and a new "child" population with top fitness values is selected. The process is iterated until a stopping criteria is met, such as the change in fitness function (Eq. 7) is smaller than a threshold or a maximum number of iteration is reached. The fuzzy measure with the best fitness value is selected as the optimal measure.

Algorithm 1 MIMRF Optimization [53]

```
    TRAINING
Require: Training Bags $\mathscr{B}$, Bag-level Labels, Parameters
    Initialize a population of fuzzy measures, $\mathscr{W}$
    Compute objective values $\mathbf{J}^0_{\mathscr{W}}$
    $J^* = min(\mathbf{J}^0_{\mathscr{W}})$, $\mathbf{g}^* = \arg\min_{\mathscr{W}} J^*$
    for $t := 1 \rightarrow$ Number of Iterations do
        for $p := 1 \rightarrow$ Size of Population do
            Evaluate valid intervals of all fuzzy measures
            if Small-scale mutation then
                Update one element in $\mathscr{W}_p$
            else
                Update $\mathscr{W}_p$ by large-scale mutation
            end if
        end for
        Compute $J(\mathscr{W})$ using (7)
        Select and keep the fuzzy measures with low objective function values
        if $min(\mathbf{J}^t_{\mathscr{W}}) < J^*$ then
            $J^* = min(\mathbf{J}^t_{\mathscr{W}})$, $\mathbf{g}^* = \arg\min_{\mathscr{W}} J^*$
        end if
    end for
        return Optimal fuzzy measure $\mathbf{g}^*$

    TESTING
Require: Testing Data, $\mathbf{g}^*$
    $TestLabels \leftarrow$ CI fusion output computed based on Equation (1) using the learned
    optimal fuzzy measure $\mathbf{g}^*$ obtained from Training
        return $TestLabels$
```

In testing, the learned fuzzy measure is used to compute the CI fusion output for test data. More details about the evolutionary algorithm can be seen in [53] and the pseudocode can be seen in Algorithm 1.

4. Experiments

This section presents experimental results of the proposed MIMRF algorithm on two real remote sensing data sets. First, the proposed MIMRF algorithm is used for hyperspectral and LiDAR fusion for scene understanding in the MUUFL Gulfport data set. Then, agricultural applications of the MIMRF algorithm is shown on a multi-resolution soybean data set for weed detection.

4.1. MUUFL Gulfport HSI and LiDAR Fusion Data Set

This section desribes the MUUFL Gulfport hyperspectral and LiDAR fusion data set and presents experimental results. The proposed MIMRF was used to perform multi-resolution fusion on the hyperspectral imagery and raw LiDAR point cloud data.

4.1.1. Data Set Description

The MUUFL Gulfport hyperspectral imagery and LiDAR data set [2, 3, 66] was collected over the University of Southern Mississippi - Gulfpark campus in November 2011. The data set contains hyperspectral and lidar data from two flights. The RGB images from the HSI imagery of the two flights (named “campus 1” and “campus 2” data) are shown in Figure 1a and Figure 1b. The first 220 columns of the hyperspectral data were kept due to highly bright beach sand materials in the lower right corner in the original images. The first and last four bands of the data were removed due to noise. The size of the HSI image is $325 \times 220 \times 64$ for both flights in this experiment.

The LiDAR raw point cloud data was collected by the Gemini LiDAR sensor at 3500 ft over the scene [2]. Figure 4 shows the flight lines of the LiDAR data. The red pin shows the location of the campus. LiDAR flights 001 and 002 cover the campus area and the LiDAR data from these two lines were used for fusion. The scatter plot of the raw LiDAR point cloud data over the entire Gulfpark campus is shown in Figure 2. Rasterized LiDAR imagery (pre-processed by 3001 Inc. and Optech Inc.) was used to generate results for comparison methods. The rasterized LiDAR imagery for campus 1 and campus 2 are shown in Figure 5a and Figure 5b.

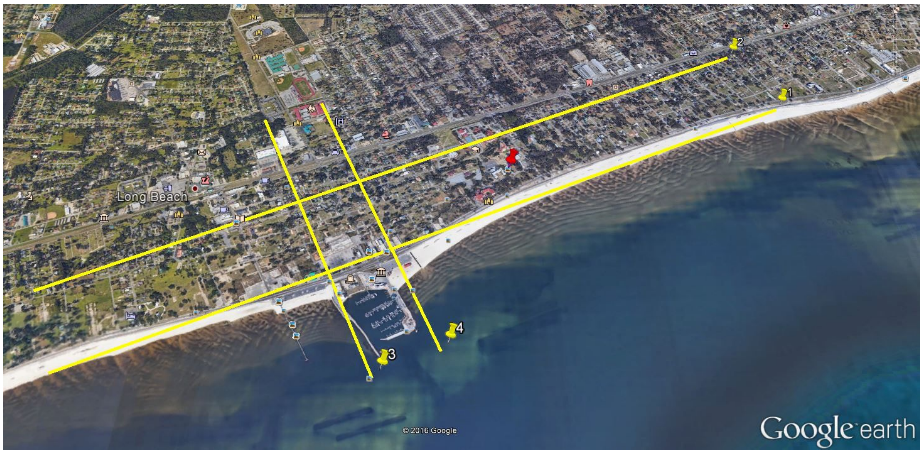


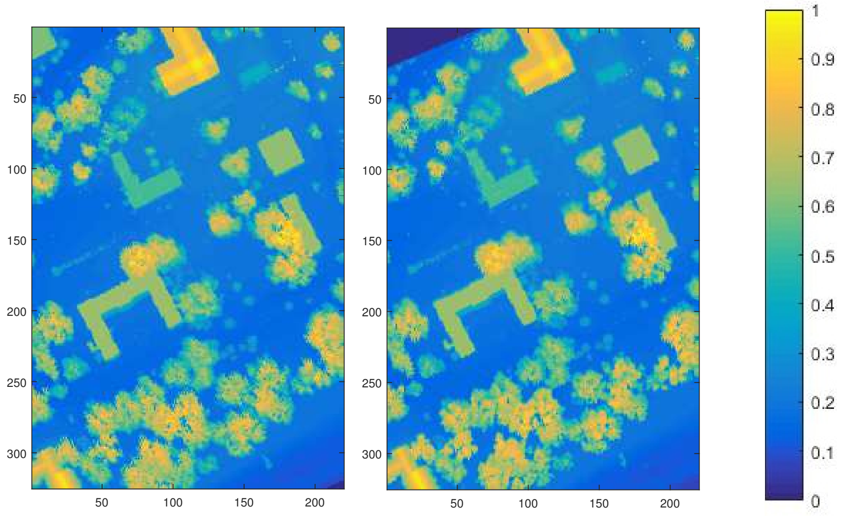
Figure 4: Four LiDAR lines in MUUFL Gulfport data, shown in Google Earth.

4.1.2. Automated Training Label Generation

As discussed in Introduction, hand-labeling is tedious, expensive, and prone to error especially for such large data. For transition regions such as edges of a building, it is impossible for humans to visually identify and correctly hand-label each pixel. In this section, an automated label generation method is proposed to automatically generate bag-level training labels for this data set.

The “bags” for this data set were constructed using the simple linear iterative clustering (SLIC) algorithm [67, 68]. The SLIC algorithm is a widely used, unsupervised superpixel segmentation algorithm that can produce spatially coherent regions [68]. Figure 7b shows the SLIC segmentation result on the MUUFL Gulfport hyperspectral campus 1 data. The lines mark the boundaries for each superpixel. Each superpixel is treated as a “bag” in training and all pixels in each superpixel are all the instances in the bag.

Training labels for each bag were automatically generated based on the crowd-sourced open map data from Open Street Map (OSM) [69]. Information from Google Earth [70], Google Maps [71] and geo-tagged photographs from a digital camera taken



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

Figure 5: Raster image of the first return MUUFL Gulfport LiDAR data. The color represents the lidar height information. The rasterized image is in the same size as the hyperspectral imagery. (a) Campus 1 data; (b) Campus 2 data. (c) Color mapping.

at the scene were also used as auxiliary data to assist the labeling process. Figure 6 shows the map extracted from Open Street Map (OSM) based on the tags available, such as “highway”, “footway”, “building”, “parking”, etc. The blue lines corresponds to asphalt, which includes road, highway and parking lot. The magenta lines corresponds to sidewalk/footway. The green lines marks buildings. The black lines corresponds to “other” tags. Key points (such as corners of the buildings) were selected manually from both OSM map and HSI RGB imagery and affine transformation [72] was used to map between the OSM data and the HSI data coordinates.

Buildings are the easiest to identify in a scene due to their spatial coherence. However, buildings are also the most challenging due to their edges. There can be transition area from the top of a building to the ground surface in a few pixels or less. The edge areas are also the most difficult or impossible to hand-label. Therefore, in the following experiment, we first start with the task of building detection and then further investigate the performance of the proposed algorithm on difficult regions such as building edges.

For building detection tasks, the pixel locations of buildings were automatically extracted from the OSM map coordinates (as marked by the green lines in Figure 6). An affine transformation was used to transform lat/lon coordinates in the OSM map to pixel coordinates in the HSI imagery. Then, all the superpixels that contains at least one building pixel are labeled positive and all the superpixels that do not contain building pixels are labeled negative. Figure 7a shows the ground truth map for buildings (with a grey roof) and Figure 7b shows the bag-generation results with bag-level labels. The red marks the positive bags that contain building pixels and the blue marks the negative bags that do not contain building pixels.

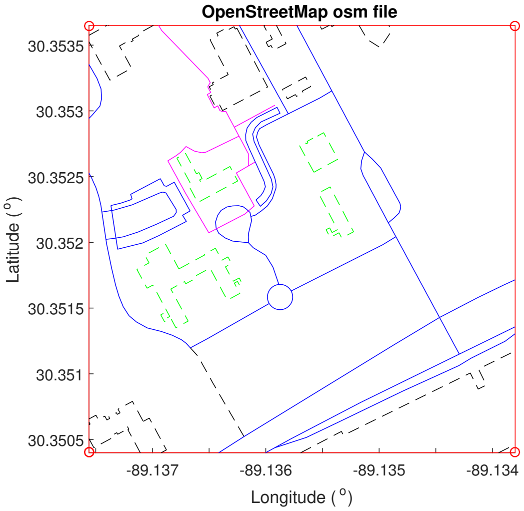
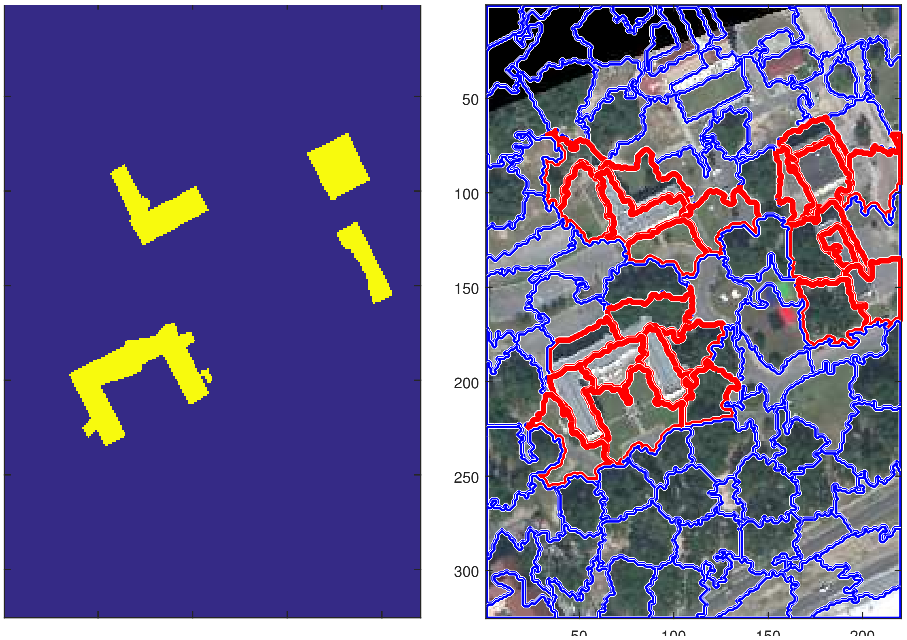


Figure 6: Open Street Map imagery over MUUFL Gulfport campus 1. The blue lines corresponds to road and highway. The magenta lines corresponds to sidewalk/footway. The green lines marks buildings. Here, the “building” tag is specific to the buildings with a grey (asphalt) roof. The black lines corresponds to “other” tags.

4.1.3. Generation of Fusion Sources for Building Detection

In this experiment, three multi-resolution sources were used for fusion, one from HSI and two from LiDAR. First, building points were manually extracted from the hyperspectral imagery and the mean spectral signature of these point were computed. The adaptive coherence estimator (ACE) detector [73, 74, 75] was applied to the HSI imagery using spectral signature of the building points. The ACE detection map for buildings is shown in Figure 9a. As can be seen, the ACE confidence map highlights most buildings, but also highlights ashpalt roads which has similar spectral signature. The ACE detector also failed to detect the top right building due to the darkness of the roof.

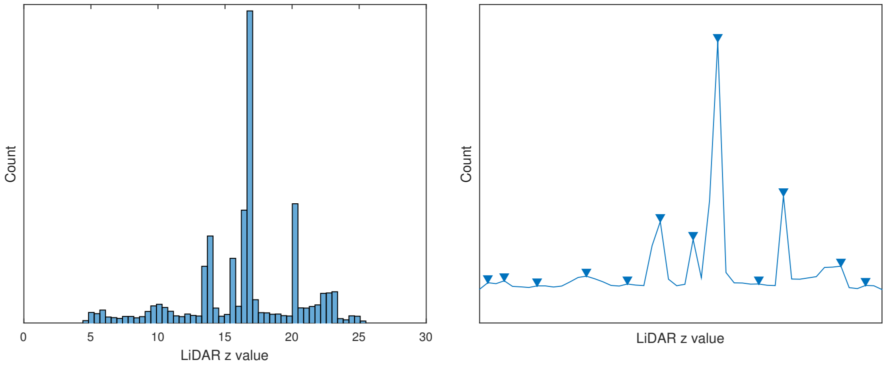


(a)

(b)

Figure 7: The Ground Truth map and the SLIC segmentation map of the MUUFL Gulfport HSI data for building detection. (a) The Ground Truth map of the buildings in MUUFL Gulfport HSI data. The yellow highlights the ground truth building locations [3]. (b)The SLIC segmentation result on MUUFL Gulfport HSI data. Red marks positive training bags and blue marks negative bags for building detection experiment.

Confidence maps from the two LiDAR flights were generated to be fused with the ACE detection map. The LiDAR height information of extracted training building points were plotted in a histogram, as shown in Figure 8a. It is assumed in this experiment, specific to this data set, that buildings heights are similar in training and testing data. The peaks of the histogram was found by using the MATLAB *findpeaks()*



(a)

(b)

Figure 8: The histogram and peaks of the LiDAR values of building points. (a) The histogram of the LiDAR values of building points. (b) The peaks found based on the histogram in (a).

function, as shown in Figure 8b. The Euclidean distance of all the LiDAR points in the scene were computed against the peak height values of the building points. We desire high confidence values on points that have similar height to the training building data. Thus, a Gaussian function was applied to compute confidence values based on the distance, as follows:

$$Conf_{lidar} = \exp\left\{-\frac{d_{lidar}}{2}\right\}, \tag{8}$$

where $Conf_{lidar}$ is the confidence value of the LiDAR points and d_{lidar} is the Euclidean distance between all the LiDAR points in the scene and the peak height value. In this way, the LiDAR map have high confidence on points with similar height to training building points and have low confidence on points with distinctly different heights than buildings. Figure 9b and Figure 9c shows the top two confidence maps computed from a rasterized LiDAR map. As can be seen, the two LiDAR confidence maps highlight different buildings in the scene and are useful fusion sources for the building detection problem.

4.1.4. MIMRF Fusion Results on Building Detection

The ACE detection map and two LiDAR confidence maps are used for fusion. Note that for the proposed MIMRF algorithm, the raw LiDAR points were used instead of the

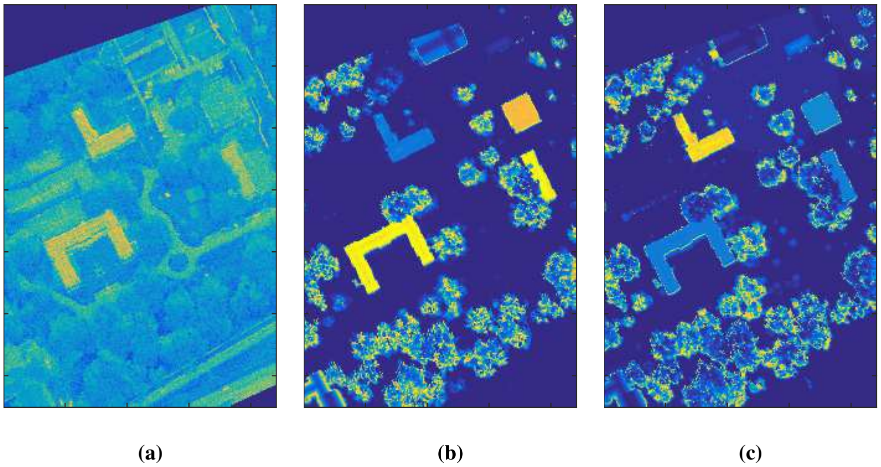


Figure 9: The confidence maps from HSI and LiDAR data for building detection in the MUUFL Gulfport. These three sources are used for fusion. (a) ACE detection map from HSI data. (b)(c) LiDAR building detection map from two LiDAR flights. The colorbar can be seen in Figure 5c.

rasterized LiDAR imagery. The rasterized LiDAR imagery are only used in comparison methods since these methods cannot handle multi-resolution multi-modal fusion.

The proposed MIMRF algorithm was compared with various methods to prove its effectiveness in multi-resolution, multi-modal fusion with uncertain labels. First, the proposed MIMRF algorithm is compared with results of individual sources, before fusion (i.e., the ACE confidence map and the LiDAR height confidence maps). Then, the proposed MIMRF is compared with popular fusion and decision-making methods including the Support Vector Machine (SVM) and the min/max/mean operators. Both SVM and the aggregation operators only work with images with pixel-to-pixel correspondence and cannot deal with multi-resolution fusion. The CI-QP [47] approach is also used as a comparison fusion method. The CI-QP approach uses the Choquet integral to perform fusion, but CI-QP does not support MIL-type learning with label uncertainties. The mi-SVM [45] method is also used in comparison since it is an alternative MIL fusion approach, but mi-SVM does not support multi-resolution fusion either. In addition, the previously proposed MICI classifier fusion algorithm [53] is also used as comparison. MICI is a MIL extension on Choquet integral and can handle

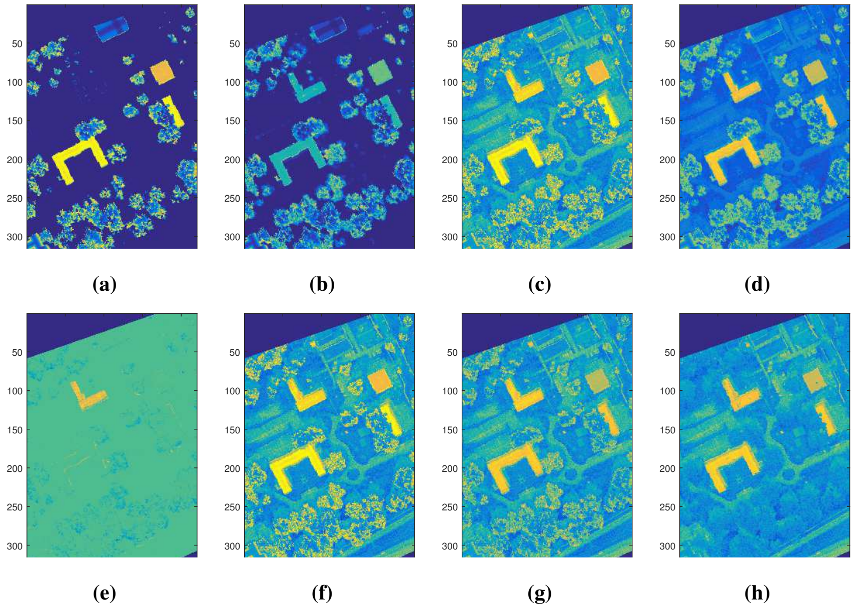
Table 1: The characteristics of comparison methods. “Fusion” means fusion method. “CI” means the method uses the Choquet integral as a fusion tool. “MIL” means the method supports the MIL framework and can handle label uncertainty. “MR” means the method supports multi-resolution and multi-modal fusion. The table cell is marked “✓” if the method supports the heading conditions, and left blank if it does not work with the heading conditions. The top three rows indicate individual sources before fusion.

Comparison Methods	**Fusion**	**CI**	**MIL**	**MR**
ACE				
Lidar1				
Lidar2				
SVM/min/max/mean	✓			
mi-SVM	✓		✓	
CI-QP	✓	✓		
MICI	✓	✓	✓	
MIMRF (proposed)	✓	✓	✓	✓

label uncertainty, but MICI can only work with pixel-to-pixel correspondence as well and does not support multi-resolution fusion. The CI-QP, mi-SVM, and MICI methods only work with rasterized LiDAR imagery, while our proposed MIMRF can directly handle raw LiDAR point cloud data. Table 1 shows a comprehensive list of comparison methods and their differences.

Two-fold cross validation is performed on this data set, i.e. training on one flight and testing on another flight. An example of fusion results when training on campus 1 and testing on campus 2 across all methods are shown in Figure 10. Figures 11a and Figures 11b show the overall ROC curve on building detection with cross validation. In addition, the Area Under Curve (AUC) results from the ROC curves were computed to provide a quantitative comparison. The first two columns of Table 2 shows the AUC results for building detection. The ACE, Lidar1, and Lidar2 rows are results from the individual sources before fusion; the methods below the dotted line are all comparison fusion results. As can be seen, the proposed MIMRF algorithm produces the best or second best ROC curve and AUC results.

Another evaluation metric, the Root Mean Square Error (RMSE), is used for comparison as well. The AUC evaluates how well the method detects the buildings (the



(a) (b) (c) (d)

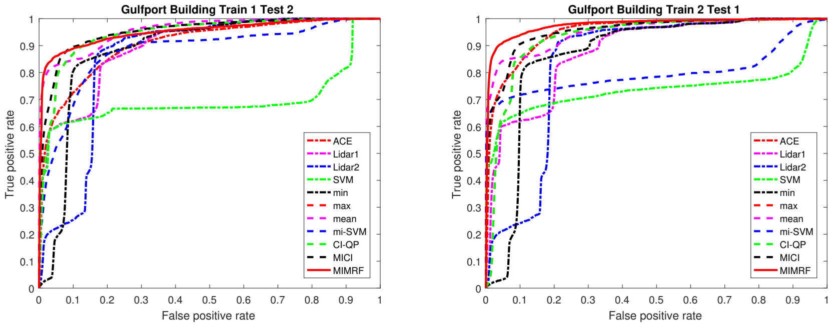
(e) (f) (g) (h)

Figure 10: The fusion results for building detection in the MUUFL Gulfport data set. Train on campus 1 and test on campus 2. Fusion results by (a) SVM; (b) Min operator; (c) Max opeator; (d) Mean operator; (e) mi-SVM; (f) CI-QP; (g) MICI; (h) The proposed MIMRF algorithm.

higher AUC the better) and the RMSE shows how the detection results on both the building and non-building points differ from the ground truth (the lower the RMSE the better). Table 3 shows the RMSE comparison results between MICI and MIMRF methods (the top methods with high AUC results). As can be seen, the proposed MIMRF has higher AUC performance and lower RMSE than comparison methods.

4.1.5. MIMRF Results on Edge Areas

As discussed in Section 4.1.2, building detection is a challenging tasks especially along the building edges. The edge areas (including but not limited to building edges) refer to transition areas and boundaries regions where there is a sudden change in altitude. These edge areas are difficult to label and are prone to noise and misalignment in image registration and rasterization. One of the biggest difference between the proposed MIMRF algorithm and previous methods is that the proposed MIMRF can directly use raw LiDAR point cloud data while previous image fusion methods require



(a)

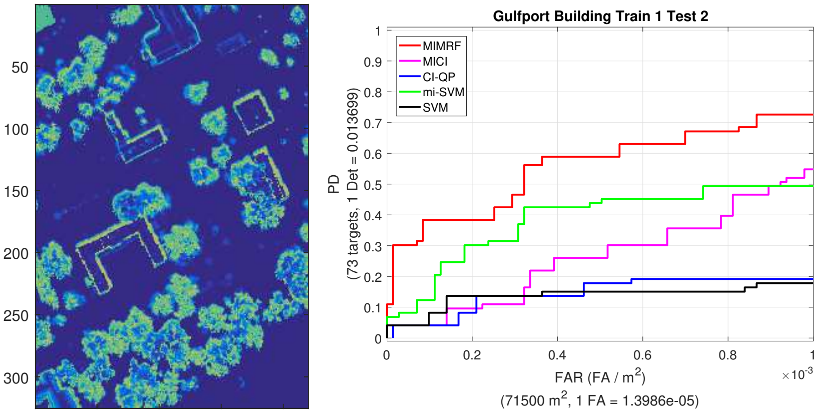
(b)

Figure 11: The overall ROC curve for building detection for MUUFL Gulfport data. (a) Train on campus 1, test on campus 2; (b) Train on campus 2, test on campus 1.

rasterization. It would be interesting, therefore, to investigate the performance of all fusion methods on such edge areas, where the rasterization may be noisy, inaccurate, or misaligned.

The rasterized LiDAR imagery in our MUUFL data set was provided by third party companies (3001 Inc. and Optech Inc.) with unknown rasterization techniques. We call the provided LiDAR rasterization “Optech LiDAR map”. Recall that the proposed MIMRF algorithm can automatically select the “correct” LiDAR points for multi-resolution fusion with its objective function. Thus, we can plot the LiDAR points selected by the proposed MIMRF into a 325×220 map. We call this map a “ selected LiDAR map” by the proposed MIMRF algorithm. Figure 12a show the difference between the selected LiDAR map and the Optech rasterized LiDAR map. As can be seen, the differences are mainly along the edge areas, such as building edges or the edges of tree canopy. These pixels on the boundary are likely where the rasterization is inaccurate or misaligned, due to the drastic changes in elevation between an object and its surrounding pixels. This difference map is referred to as an “edge map” in the following dicussions.

We also generated three alternative rasterized LiDAR maps by taking the min/max/ mean of LiDAR points within neighborhood pixels. We computed the difference between the min/max/mean LiDAR map and the given Optech LiDAR map. Any pixel



(a) **(b)**

Figure 12: The LiDAR edge map and the scoring results on the edge map in the MUUFL Gulfport data set. (a) The differences between the rasterized LiDAR imagery and the selected lidar points by the proposed MIMRF, i.e. the LiDAR edge map. (b) One example of the ROC curve results for building detection scoring on the LiDAR edge map.

that has a difference value above a threshold is determined as "edge" pixels. We then look specifically at the detection results on those edge pixels. The purpose of this experiment is to show that the proposed MIMRF algorithm, which uses the raw LiDAR point cloud data, can select the correct LiDAR points from the point cloud and produce better performance especially in those edge areas when compared with other fusion methods that use the (inaccurate) rasterized imagery.

Figure 12b shows an example of the ROC curve results scored only on the edge pixels. Table 4 and Table 5 present the AUC results with FAR up to $10^{-3}/m^2$ and the RMSE of the fusion methods, scored only on the edges. In the tables, the "MIMRF diff map" refers to results scoring on the edges determined by the difference between the LiDAR points that were selected by the MIMRF method and the Optech rasterized LiDAR imagery. The "max/min/mean diff map" refers to results scoring on the edges between aggregating neighborhood LiDAR points using the max/min/mean operators and the Optech rasterized LiDAR imagery. As can be seen from the AUC and RMSE results, the proposed MIMRF algorithm has superior performance compared with other fusion methods, specifically on the edge areas.

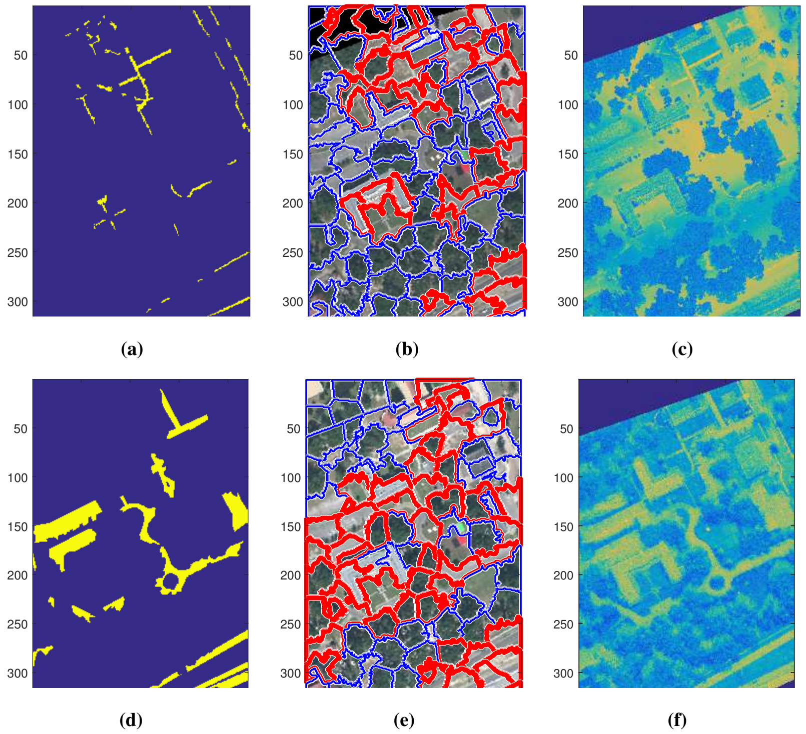
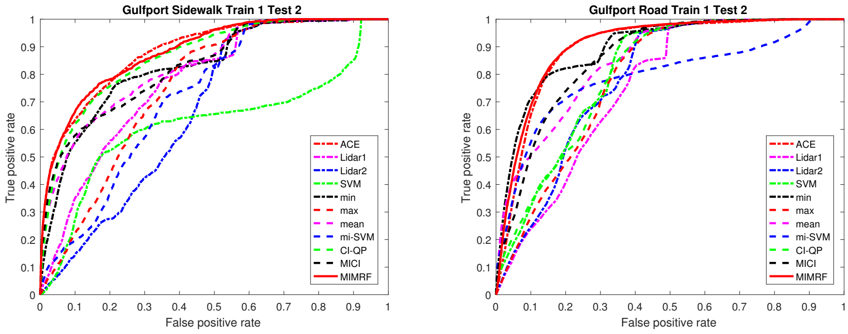


Figure 13: Results for the MUUFL Gulfport sidewalk and road detection experiments. (a)(d) The Ground Truth map of sidewalks and roads. (b)(e) The SLIC segmentation results on sidewalks and roads. Red marks positive training bags and blue marks negative bags for road detection experiment. (c)(f) One example of the MIMRF Fusion results for sidewalk and road detection, trained on campus 1 and tested on campus 2.

4.1.6. MUUFL Gulfport Sidewalk and Road Detection

We conducted similar experiments on other materials in the scene and present sidewalk and road detection results in this section. Sidewalks and roads are on ground surface level and do not have drastic altitude change as building edges. These additional experiments on sidewalk and road show the effectiveness of the proposed MIMRF algorithm in understanding other materials and objects in the scene.

Figure 13 shows the ground truth map and SLIC segmentation results for sidewalk



(a) **(b)**

Figure 14: The overall ROC curve for (a) sidewalk and (b) road detection for the MUUFL Gulfport data. Train on campus 1, test on campus 2.

and road detection experiments. Figure 13c and Figure 13f show the MIMRF fusion results for sidewalk and road detection. Figure 14 shows the cross-validated overall ROC curve results on sidewalk and road detection. The complete AUC results can be seen in Table 2. As can be seen, the proposed MIMRF algorithm can successfully detect sidewalks and roads in the scene as well. Note that since sidewalks and roads have similar altitude with ground surface, the LiDAR sources do not have significant impact. That is why the ACE map from the HSI data has comparable AUC performance to the fusion results in sidewalk and road. Still, the MIMRF achieved best or second best in half of the experiments.

4.2. Soybean and Weed Data Set

The proposed MIMRF algorithm was originally motivated by the HSI/LiDAR fusion problem for scene understanding. However, the proposed MIMRF can also be used as a general multi-resolution fusion framework for many applications. This section provides additional experimental results of the proposed MIMRF on an agricultural data set.

The proposed MIMRF was used to detect weed in a multi-resolution soybean and

Table 2: The AUC results of building, sidewalk, and road detection using MUUFL Gulfport HSI and LiDAR data. The best two results with the highest AUC were **bolded** and underlined, respectively. The standard deviation is in parentheses.

	Building Detection		**Sidewalk Detection**		**Road Detection**	
	Train1Test2	Train2Test1	Train1Test2	Train2Test1	Train1Test2	Train2Test1
ACE	0.906	0.952	**0.882**	**0.931**	0.896	0.902
Lidar1	0.897	0.880	0.772	0.769	0.752	0.748
Lidar2	0.856	0.839	0.670	0.669	0.784	0.779
SVM	0.694	0.738	0.622	0.663	0.806	0.396
min	0.885	0.867	0.830	0.885	0.896	**0.918**
max	0.943	0.931	0.754	0.754	0.785	0.779
mean	**0.957**	0.953	0.831	0.870	0.849	0.856
mi-SVM	0.881	0.800	0.721	0.904	0.791	0.817
CI-QP	0.943	0.931	0.767	0.918	0.801	0.815
MICI	0.952(0.000)	0.956(0.000)	0.838(0.009)	0.908(0.001)	0.873(0.011)	0.824(0.003)
MIMRF	0.952(0.000)	**0.977(0.000)**	0.854(0.019)	0.861(0.010)	**0.905(0.002)**	0.895(0.003)

Table 3: The RMSE results of MICI and MIMRF on building, sidewalk, and road detection. The best results with the lower RMSE are **bolded**. The standard deviation is in parentheses.

	Building Detection		**Sidewalk Detection**		**Road Detection**	
	Train1Test2	Train2Test1	Train1Test2	Train2Test1	Train1Test2	Train2Test1
MICI	0.403(0.002)	0.382(0.000)	0.485(0.002)	**0.466(0.002)**	0.480(0.009)	0.514(0.002)
MIMRF	**0.351(0.004)**	**0.331(0.001)**	**0.460(0.007)**	0.489(0.006)	**0.448(0.007)**	**0.478(0.008)**

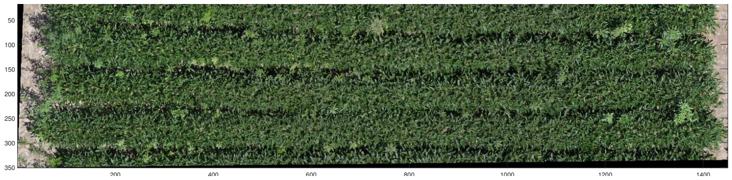


Figure 15: The RGB image of the soybean-weed data.

weed data set[1]. In the data set, a height map and a RGB image are provided over a patch of soybean field and the goal is to detect weed amongst the soybean plants. The

[1] This data set is provided by Precision Silver, LLC.

Table 4: The AUC and RMSE results of MICI and MIMRF on building detection, scored on the edge maps. Train on campus 1 and test on campus 2. The best two results with the highest AUC and lowest RMSE were **bolded** and underlined, respectively. The standard deviation is in parentheses.

	MIMRF diff map		max diff map		min diff map		mean diff map	
	AUC	RMSE	AUC	RMSE	AUC	RMSE	AUC	RMSE
SVM	0.420	0.040	0.113	0.067	0.141	0.125	0.078	0.063
mi-SVM	0.704	0.031	0.327	0.076	0.448	0.140	0.490	0.071
CI-QP	0.329	0.055	0.135	0.080	0.126	0.147	0.115	0.073
MICI	0.371(0.021)	0.046(0.000)	0.311(0.017)	0.068(0.001)	0.190(0.022)	0.125(0.002)	0.401(0.020)	0.062(0.001)
MIMRF	**0.776(0.004)**	**0.022(0.001)**	**0.458(0.022)**	**0.049(0.000)**	**0.614(0.025)**	**0.082(0.001)**	**0.619(0.027)**	**0.044(0.000)**

Table 5: The AUC and RMSE results of MICI and MIMRF on building detection, scored on the edge maps. Train on campus 2 and test on campus 1.

	MIMRF diff map		max diff map		min diff map		mean diff map	
	AUC	RMSE	AUC	RMSE	AUC	RMSE	AUC	RMSE
SVM	0.513	0.058	0.413	0.101	0.537	0.176	0.390	0.109
mi-SVM	0.695	**0.021**	0.488	**0.018**	0.577	**0.031**	0.528	**0.017**
CI-QP	0.094	0.104	0.096	0.113	0.027	0.202	0.000	0.119
MICI	0.683(0.004)	0.077(0.000)	0.451(0.008)	0.091(0.000)	0.375(0.009)	0.162(0.000)	0.448(0.009)	0.096(0.000)
MIMRF	**0.794(0.007)**	0.035(0.000)	**0.529(0.003)**	0.061(0.000)	**0.649(0.007)**	0.103(0.000)	**0.638(0.005)**	0.064(0.000)

height map is 351×1450 in size and the RGB map is 1404×5864 in size. Figure 15 shows the RGB map over the scene.

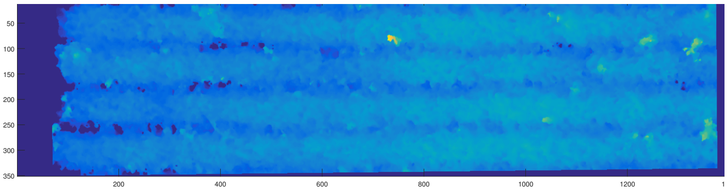


Figure 16: The height map of the soybean-weed data.

Three fusion sources were used for the proposed algorithm for weed detection. Figure 16 shows the height map over the soybean-weed field. Some weeds in the scene are slightly higher than soybean plants, indicating that height is an important feature for weed detection. To extract height features, four Gabor filters at angles 0°, 45°, 90°, and 135° were applied to the height map and the sum of the filtered images was used as one of the fusion sources, as shown in Figure 19.

We also observed that, in this data set, the weed plants have lighter-colored pixels than the soybean plants. That is to say, lightness and color are useful sources of weed detection as well. We transformed the RGB values into LAB space and the L- and B-band imagery were used as the other two sources for fusion. The L dimension provide information about lightness and the B dimension is the color opponent for blue-yellow space. Figure 17 and Figure 18 show the L and B band images, where weed pixels are highlighted.

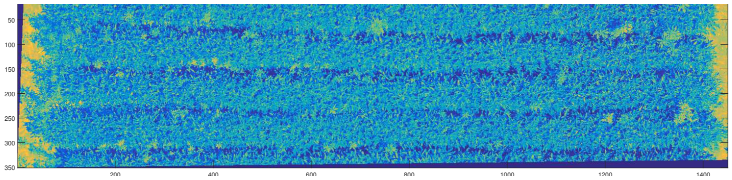


Figure 17: The L-band image of the soybean-weed data.

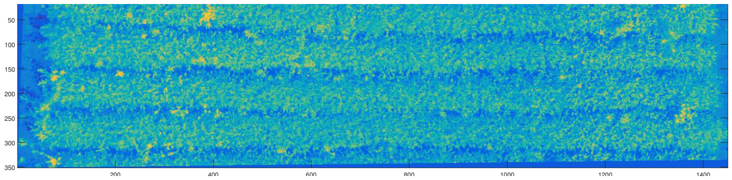


Figure 18: The B-band image of the soybean-weed data.

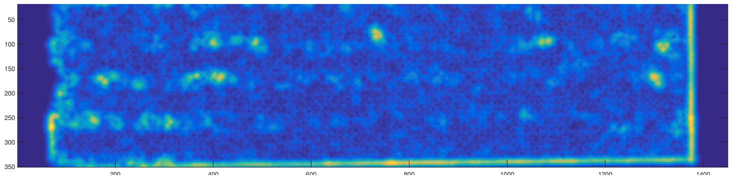


Figure 19: The Gabor filtered image of the soybean-weed data height map.

Figure 20 shows the (manual) ground truth map for weed in this data set. Figure 21 shows the SLIC segmentation results. Similar to Section 4.1.2, each superpixel in this segmentation was regarded as a “bag” and the colors mark the bag-level training labels. Figure 22 shows the confidence maps after the proposed MIMRF fusion. Figure 23 shows the overall ROC curve result in one run for all the comparing fusion methods. As can be seen, the proposed MIMRF method can effectively detect weed in the scene and produce an overall better ROC curve performance.

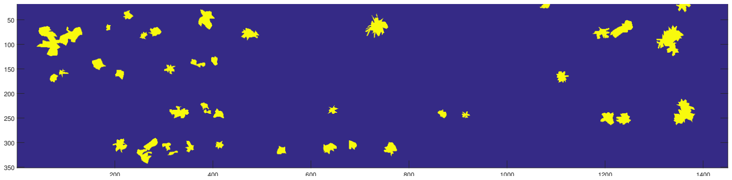


Figure 20: The Ground Truth map of weed in the soybean-weed data. The deep blue is the background (soybean plants) and the yellow marks the target (weed).

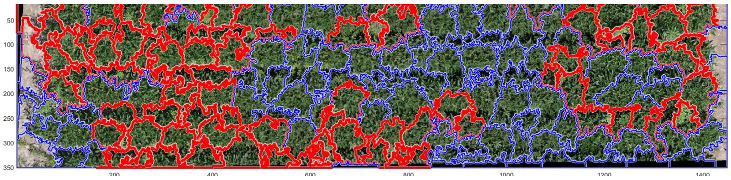


Figure 21: The SLIC segmentation map of the soybean-weed data. Red marks positive training bags and blue marks negative bags for weed detection experiment.

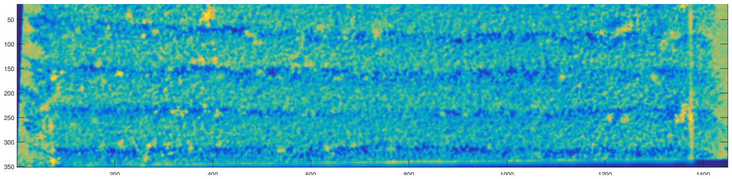


Figure 22: The confidence map obtained from the MIMRF fusion for the soybean-weed data.

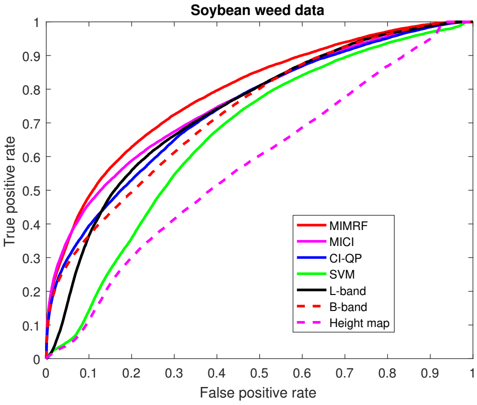


Figure 23: The overall ROC curve results for weed detection in the soybean-weed data across comparison methods.

5. Conclusion

In this paper, we proposed a novel Multiple Instance Multi-Resolution Fusion (MIMRF) framework that can perform multi-resolution, multi-modal sensor fusion on remote sensing data with label uncertainty. The proposed MIMRF is unique in that it directly fuses multi-resolution, multi-modal sensor outputs without the need of rasterization or alignment. Specifically, in the HSI/LiDAR fusion problem, the proposed MIMRF can directly handle raw LiDAR point cloud data instead of requiring rasterized LiDAR imagery. In addition, the proposed MIMRF is effective at selecting the correct LiDAR points and show superior performance especially at edge areas where tradition methods are likely to be wrong. Additionally, the proposed MIMRF does not require accurate pixel-wise training labels and can handle bag-level labels. Although the proposed MIMRF was originally motivated by the hyperspectral imagery and LiDAR point cloud fusion problem in remote sensing, the method is a general framework can be applied to many multi-resolution and multi-modal fusion applications with uncertian labels, such as precision agriculture.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant IIS-1723891-CAREER: Supervised Learning for Incomplete and Uncertain Data. The authors would also like to thank Precision Silver LLC for providing the soybean and weed data set.

References

References

[1] P. Gader, A. Mendez-Vasquez, K. Chamberlin, J. Bolton, A. Zare, Multi-sensor and algorithm fusion with the choquet integral: applications to landmine detection, in: IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), Vol. 3, 2004, pp. 1605–1608.

[2] P. Gader, A. Zare, R. Close, J. Aitken, G. Tuell, Muufl gulfport hyperspectral and lidar airborne data set, Tech. Rep. Rep. REP-2013-570, University of Florida, Gainesville, FL (Oct. 2013).

[3] X. Du, A. Zare, Technical report: scene label ground truth map for muufl gulfport data set, Tech. Rep. Tech. Rep. 20170417, University of Florida, Gainesville, FL. (2017).

[4] C. Pohl, J. L. V. Genderen, Multisensor image fusion in remote sensing: Concepts, methods and applications, Int. J. Remote Sens. 19 (5) (1998) 823–854.

[5] M. Liggins II, D. Hall, J. Llinas, Handbook of multisensor data fusion, CRC Press, 2008.

[6] J. Zhang, Multi-source remote sensing data fusion: status and trends, Int. J. Image and Data Fusion 1 (1) (2010) 5–24. `arXiv:http://dx.doi.org/10.1080/19479830903561035`.

[7] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, J. Chanussot, Hyperspectral remote sensing data analysis and future challenges, IEEE Geosci. Remote Sens. Mag. 1 (2) (2013) 6–36.

[8] Y. Li, E. B. Olson, Extracting general-purpose features from lidar data, in: IEEE Int. Conf. Robotics and Automation (ICRA), 2010, pp. 1388–1393.

[9] M. Khodadadzadeh, J. Li, S. Prasad, A. Plaza, Fusion of hyperspectral and lidar remote sensing data using multiple feature learning, IEEE J. Sel. Topics. Appl. Earth Observ. 8 (6) (2015) 2971–2983.

[10] B. Rasti, P. Ghamisi, J. Plaza, A. Plaza, Fusion of hyperspectral and lidar data using sparse and low-rank component analysis, IEEE Trans. Geosci. Remote Sens. 55 (11) (2017) 6354–6365.

[11] R. Luo, W. Liao, H. Zhang, L. Zhang, P. Scheunders, Y. Pi, W. Philips, Fusion of hyperspectral and lidar data for classification of cloud-shadow mixed remote sensed scene, IEEE J. Sel. Topics. Appl. Earth Observ. 10 (8) (2017) 3768–3781.

[12] A. Sampath, J. Shan, Segmentation and reconstruction of polyhedral building roofs from aerial lidar point clouds, IEEE Trans. Geosci. Remote Sens. 48 (3) (2010) 1554–1567.

[13] S. Cao, X. Zhu, Y. Pan, Q. Yu, A stable land cover patches method for automatic registration of multitemporal remote sensing images, IEEE J. Sel. Topics. Appl. Earth Observ. 7 (8) (2014) 3502–3512.

[14] G. Brigot, E. Colin-Koeniguer, A. Plyer, F. Janez, Adaptation and evaluation of an optical flow method applied to coregistration of forest remote sensing images, IEEE J. Sel. Topics. Appl. Earth Observ. 9 (7) (2016) 2923–2939.

[15] Y. Li, H. Wu, Adaptive building edge detection by combining lidar data and aerial images, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 37 (Part B1) (2008) 197–202.

[16] T. G. Dietterich, R. H. Lathrop, T. Lozano-Pérez, Solving the multiple instance problem with axis-parallel rectangles, Artif. Intell. 89 (1-2) (1997) 31–71.

[17] H. Yang, Q. Du, B. Ma, Decision fusion on supervised and unsupervised classifiers for hyperspectral imagery, IEEE Geosci. Remote Sens. Lett. 7 (4) (2010) 875–879.

[18] M. Dalponte, L. Bruzzone, D. Gianelle, Fusion of hyperspectral and lidar remote sensing data for classification of complex forest areas, IEEE Trans. Geosci. Remote Sens. 46 (5) (2008) 1416–1427.

[19] H. Shen, X. Meng, L. Zhang, An integrated framework for the spatio-temporal-spectral fusion of remote sensing images, IEEE Trans. Geosci. Remote Sens. 54 (12) (2016) 7135–7148.

[20] S. Hinz, A. Baumgartner, Automatic extraction of urban road networks from multi-view aerial imagery, ISPRS J. Photogrammetry and Remote Sensing 58 (1) (2003) 83–98.

[21] K. Stankov, D.-C. He, Detection of buildings in multispectral very high spatial resolution images using the percentage occupancy hit-or-miss transform, IEEE J. Sel. Topics. Appl. Earth Observ. 7 (10) (2014) 4069–4080.

[22] R. L. King, A challenge for high spatial, spectral, and temporal resolution data fusion, in: IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), Vol. 6, IEEE, 2000, pp. 2602–2604.

[23] A. Y.-M. Lin, A. Novo, S. Har-Noy, N. D. Ricklin, K. Stamatiou, Combining geoeye-1 satellite remote sensing, uav aerial imaging, and geophysical surveys in anomaly detection applied to archaeology, IEEE J. Sel. Topics. Appl. Earth Observ. 4 (4) (2011) 870–876.

[24] M. Ehlers, S. Klonus, P. Johan Åstrand, P. Rosso, Multi-sensor image fusion for pansharpening in remote sensing, International J. Image and Data Fusion 1 (1) (2010) 25–45.

[25] C. Shi, F. Liu, L. Li, L. Jiao, Y. Duan, S. Wang, Learning interpolation via regional map for pan-sharpening, IEEE Trans. Geosci. Remote Sens. 53 (6) (2015) 3417–3431.

[26] P. Liu, L. Xiao, J. Zhang, B. Naz, Spatial-hessian-feature-guided variational model for pan-sharpening, IEEE Trans. Geosci. Remote Sens. 54 (4) (2016) 2235–2253.

[27] Q. Du, N. H. Younan, R. King, V. P. Shah, On the performance evaluation of pansharpening techniques, IEEE Geosci. Remote Sens. Lett. 4 (4) (2007) 518–522.

[28] D. M. McKeown, S. D. Cochran, S. J. Ford, J. C. McGlone, J. A. Shufelt, D. A. Yocum, Fusion of hydice hyperspectral data with panchromatic imagery for cartographic feature extraction, IEEE Trans. Geosci. Remote Sens. 37 (3) (1999) 1261–1277.

[29] G. A. Licciardi, M. M. Khan, J. Chanussot, A. Montanvert, L. Condat, C. Jutten, Fusion of hyperspectral and panchromatic images using multiresolution analysis

and nonlinear pca band reduction, EURASIP J. Advances in Signal processing 2012 (1) (2012) 1–17.

[30] R. B. Gomez, A. Jazaeri, M. Kafatos, Wavelet-based hyperspectral and multispectral image fusion, in: Aerospace/Defense Sensing, Simulation, and Controls, International Society for Optics and Photonics, 2001, pp. 36–42.

[31] Z. Chen, H. Pu, B. Wang, G.-M. Jiang, Fusion of hyperspectral and multispectral images: A novel framework based on generalization of pan-sharpening methods, IEEE Geosci. Remote Sens. Lett. 11 (8) (2014) 1418–1422.

[32] V. De Silva, J. Roche, A. Kondoz, Fusion of lidar and camera sensor data for environment sensing in driverless vehicles, arXiv preprint arXiv:1710.06230.

[33] W. Maddern, P. Newman, Real-time probabilistic fusion of sparse 3d lidar and dense stereo, in: IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS), IEEE, 2016, pp. 2181–2188.

[34] O. Maron, A. L. Ratan, Multiple-instance learning for natural scene classification, in: Proc. 15th Int. Conf. Mach. Learn., 1998, pp. 341–349.

[35] Z.-H. Zhou, M.-L. Zhang, Multi-instance multi-label learning with application to scene classification, in: Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2006, pp. 1609–1616.

[36] S. Ali, M. Shah, Human action recognition in videos using kinematic features and multiple instance learning, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2) (2010) 288–303.

[37] C. Zhang, J. C. Platt, P. A. Viola, Multiple instance boosting for object detection, in: Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2005, pp. 1417–1424.

[38] B. Babenko, M.-H. Yang, S. Belongie, Visual tracking with online multiple instance learning, in: IEEE Conf. Computer Vision and Pattern Recognition (CVPR), IEEE, 2009, pp. 983–990.

[39] B. Babenko, M.-H. Yang, S. Belongie, Robust object tracking with online multiple instance learning, IEEE Trans. Pattern Anal. Mach. Intell. 33 (8) (2011) 1619–1632.

[40] X. Du, A. Zare, J. T. Cobb, Possibilistic context identification for sas imagery, in: Proc. SPIE 9454, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XX, Vol. 9454, 2015.

[41] P. Torrione, C. Ratto, L. M. Collins, Multiple instance and context dependent learning in hyperspectral data, in: 1st Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), IEEE, 2009, pp. 1–4.

[42] L. Xu, D. A. Clausi, F. Li, A. Wong, Weakly supervised classification of remotely sensed imagery using label constraint and edge penalty, IEEE Trans. Geosci. Remote Sens. 55 (3) (2017) 1424–1436.

[43] A. Zare, C. Jiao, T. Glenn, Discriminative multiple instance hyperspectral target characterization, IEEE Trans. Pattern Anal. Mach. Intell. PP (99) (2017) 1–12.

[44] L. Cao, F. Luo, L. Chen, Y. Sheng, H. Wang, C. Wang, R. Ji, Weakly supervised vehicle detection in satellite images via multi-instance discriminative learning, Pattern Recognition 64 (2017) 417 – 424.

[45] S. Andrews, Support vector machines for mulitple-instance learning, in: Ann. Conf. Neural Inf. Proc. Systems (NIPS), 2002.

[46] G. Choquet, Theory of capacities, in: Annales de l'institut Fourier, Vol. 5, 1954, pp. 131–295.

[47] M. Grabisch, The application of fuzzy integrals in multicriteria decision making, European J. Operational Research 89 (3) (1996) 445–456.

[48] M. Grabisch, A new algorithm for identifying fuzzy measures and its application to pattern recognition, in: Int. Joint Conf. 4th IEEE Int. Conf. Fuzzy Systems and 2nd Int. Fuzzy Eng. Symp., Vol. 1, 1995, pp. 145–150.

[49] C. Labreuche, M. Grabisch, The choquet integral for the aggregation of interval scales in multicriteria decision making, Fuzzy Sets and Systems 137 (1) (2003) 11 – 26.

[50] A. Mendez-Vazquez, P. D. Gader, J. M. Keller, K. Chamberlin, Minimum classification error training for choquet integrals with applications to landmine detection, IEEE Trans. Fuzzy Systems 16 (1) (2008) 225–238.

[51] A. Mendez-Vazquez, P. Gader, Learning fuzzy measure parameters by logistic lasso, in: Proc. Annual Conf. North American Fuzzy Information Processing Society (NAFIPS), 2008, pp. 1–7.

[52] P. D. Gader, J. M. Keller, B. N. Nelson, Recognition technology for the detection of buried land mines, IEEE Trans. Fuzzy Systems 9 (1) (2001) 31–43.

[53] X. Du, A. Zare, J. Keller, D. Anderson, Multiple instance choquet integral for classifier fusion, in: IEEE Congress on Evolutionary Computation (CEC), Vancouver, BC, 2016, pp. 1054–1061.

[54] Q. Wang, C. Zheng, H. Yu, D. Deng, Integration of heterogeneous classifiers based on choquet fuzzy integral, in: 7th Int. Conf. Intelligent Human-Machine Systems and Cybernetics, Vol. 1, 2015, pp. 543–547.

[55] J. Fodor, J.-L. Marichal, M. Roubens, Characterization of the ordered weighted averaging operators, IEEE Trans. Fuzzy Systems 3 (2) (1995) 236–240.

[56] O. Maron, Learning from ambiguity, Ai technical report 1639, Massachusetts Institute of Technology (1998).

[57] J.-L. Marichal, An axiomatic approach of the discrete choquet integral as a tool to aggregate interacting criteria, IEEE Trans. Fuzzy Systems 8 (6) (2000) 800–807.

[58] M. Sugeno, Theory of fuzzy integrals and its applications, Ph.D. thesis, Tokyo Institute of Technology (1974).

[59] M. Fitting, E. Orlowska (Eds.), Beyond Two: Theory and Applications of Multiple-Valued Logic, Springer, 2003.

[60] J. M. Keller, D. Liu, D. B. Fogel, Fundamentals of computational intelligence: Neural networks, fuzzy systems and evolutionary computation, 1st Edition, IEEE Press Series on Computational Intelligence, John Wiley & Sons, Inc., 2016.

[61] A. Mendez-Vazquez, Information fusion and sparsity promotion using choquet integrals, Ph.D. thesis, University of Florida (2008).

[62] J. Nocedal, S. Wright, Numerical Optimization, Springer-Verlag New York, 2006.

[63] X. Du, Multiple instance choquet integral for multiresolution sensor fusion, Ph.D. thesis, University of Missouri (2017).

[64] X. Du, A. Zare, Multiple instance choquet integral classifier fusion and regression for remote sensing applications.

[65] N. Johnson, S. Kotz, N. Balakrishnan, Continuous Univariate Distributions, 2nd Edition, Vol. 1, Wiley-Interscience, 1994.

[66] A. Zare, P. Gader, J. Aitken, R. Close, G. Tuell, T. Glenn, D. Dranishnikov, X. Du, Gatorsense/muuflgulfport: Release 01 (Feb. 2018). doi:10.5281/zenodo.1186326.

[67] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, Slic superpixels, Tech. rep., École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland (2010).

[68] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, Slic superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2274–2282.

[69] OSMcontributors, Open street map.
URL https://www.openstreetmap.org

[70] Google, Google earth.
URL https://www.google.com/earth/

[71] Google, Google maps.
URL `https://www.google.com/maps/`

[72] R. J. Schalkoff, Digital image processing and computer vision, Vol. 286, Wiley New York, 1989.

[73] L. L. Scharf, L. T. McWhorter, Adaptive matched subspace detectors and adaptive coherence estimators, in: Proc. 30th Asilomar Conf. Signals Syst., IEEE, 1996, pp. 1114–1117.

[74] S. Kraut, L. L. Scharf, R. W. Butler, The adaptive coherence estimator: a uniformly most-powerful-invariant adaptive detection statistic, IEEE Trans. Signal Proc. 53 (2) (2005) 427–438.

[75] N. Pulsone, M. A. Zatman, A computationally efficient two-step implementation of the glrt, IEEE Trans. Signal Proc. 48 (3) (2000) 609–616.