

MANIFOLD LEARNING FOR MULTI-SENSOR, MULTI-RESOLUTION FUSION WITH  
IMPRECISE DATA

By

CONNOR H. MCCURLEY

A ORAL QUALIFYING EXAM PROPOSAL PRESENTED TO THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2019

© 2019 Connor H. McCurley

## TABLE OF CONTENTS

	<u>page</u>
LIST OF TABLES . . . . .	4
LIST OF FIGURES . . . . .	5
CHAPTER	
1 INTRODUCTION . . . . .	6
2 BACKGROUND . . . . .	9
3 PROBLEM DESCRIPTION . . . . .	10
4 TECHNICAL APPROACH . . . . .	11
5 EXPERIMENTAL DESIGN . . . . .	12
6 PRELIMINARY WORK . . . . .	13
7 FUTURE TASKS . . . . .	14
8 CONCLUSIONS . . . . .	15
APPENDIX	
REFERENCES . . . . .	16

## LIST OF TABLES

Table

page

## LIST OF FIGURES

Figure

page

## CHAPTER 1 INTRODUCTION

During this project, techniques will be explored for use in multi-sensor target detection, target classification, and information fusion given uncertain and imprecise groundtruth. These methods will be developed as universal approaches for fusion and will be evaluated on a variety of sensor modalities, including: mid-wave IR, visible, hyper-spectral and multi-spectral imagery, as well as LiDaR, ground-penetrating radar and electromagnetic induction sensors. The aim of this project is to develop fusion methods which can address mis-registration between sensor sources as well as uncertainty and imprecision in training data groundtruth while demonstrating robustness towards outlying and adversarial data points.

Multi-sensor fusion methods aim to amalgamate data collected from multiple information sources to reduce uncertainty and provide a greater level of understanding than can be obtained from the modalities, individually **CITE Xiaoxiao Paper**. Fusion of multiple sensor sources providing complimentary or reinforcing information is often paramount to the success of remote sensing applications. **provide intuitive example**

Information fusion approaches make two assumptions, (1) the data to be fused are homogeneous (of the same data type and with the same resolution) and (2) training labels are available for each data point **CITE**. If fusing multiple heterogeneous sources (varying types and resolutions), it is assumed that individual data points can be matched together **CITE**. In other words, standard methods require data from  $m$  different sensors to produce data with one-to-one correspondence, or that some form of pre-processing can transform all sources to the same resolution and perform matching **CITE**.

Two problems arise from these assumptions. First, the assumption concerning homogeneous data sources does not generally hold in real remote sensing applications for sensor fusion. In existing approaches, is extracted through sensor-specific pipelines before fusion on the sample, feature, or decision level. There is an assumption under this approach that each pipeline extracts information with an adequate level of precision. However, overall performance

is sensitive to each sensors' processing and information is often lost through co-registration. As an example, consider a self-driving car outfitted with an RGB camera and LiDAR (light detection and ranging) system. The goal of this computer vision system is to understand the scene around the vehicle so it may navigate safely. Classical approaches would perform pre-processing, feature extraction, and segmentation/ depth-map estimation on each sensor, individually. Co-registration would then be performed by mapping a single measurement from the dense LiDAR point-cloud to every pixel. However, if the wrong LiDAR point is assigned to a pixel, even just a small shift, the car's depth perception can be thrown off. This is information loss we cannot afford. A more intuitive method for fusion would be to first project each sensors' collection into a combined representation space before running on a unified pipeline such that the entirety of each sensor's collection is utilized.

Additionally, even assuming that homogeneously registered data is available for fusion or that there is a noiseless way to transform heterogeneous data, standard supervised learning methods require accurate labels for each training data point **CITE Xiaoxiao Dissertaion**. However, data-point specific labels are often unavailable or difficult and expensive to obtain **CITE Xiaoxiao dissertation reference 10**. Consider a real scene understanding task where an image is segmented into separate regions containing sky, grass, armored vehicle, mountains, etc. Existing segmentation techniques require pixel-level labels for each image in a training set, thus taking a significant amount of time and effort for a large dataset.

To address these two problems, I propose to investigate the following research questions:

1. Can manifold learning be extended to operate under the multiple-instance framework? If so, what is the "best" way to construct the data manifold? Can manifold learning provide robustness to outlying and adversarial exemplars?
2. Can we construct a joint-representation space for multiple sensors such that there is no loss of information amongst any of the modalities? What is the "best" way to construct this representation? Do we use raw data or perform feature extraction before combination?
3. Can we perform detection/ segmentation using a single, sensor-agnostic processing pipeline on the unified representation?

Experiments will be conducted on both synthetic data and real applications such as plant phenotyping, target detection and scene understanding in remote sensing imagery. Initial results demonstrate the aptitude of the proposed approaches and suggest further development and evaluation of these methods.



## CHAPTER 2 BACKGROUND

## CHAPTER 3

### PROBLEM DESCRIPTION

## CHAPTER 4

### TECHNICAL APPROACH

## CHAPTER 5

### EXPERIMENTAL DESIGN

CHAPTER 6  
PRELIMINARY WORK

## CHAPTER 7

### FUTURE TASKS

## CHAPTER 8

### CONCLUSIONS

## REFERENCES