# 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

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1
# INTRODUCTION

–(**Motivation for Sensor Fusion**)

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**. Fusion of multiple sensor sources providing complimentary or reinforcing information is often paramount to the success of remote sensing applications. Take the task of automatic target recognition (ATR) as an example, where the goal is to locate armored vehicles in a scene subject to varying environmental conditions. Imagery can be obtained from multiple sensors such as mid-wave infrared (MWIR) and hyperspectral (HSI) cameras. Hyperspectral can provide a broad range of spectral information about the materials present in a scene, while MWIR imaging sensors can supply thermal information. If a building and vehicle are both constructed from the same material (i.e. metal), hyperspectral information alone may not be sufficient to tell them apart. However, MWIR supplying thermal information can easily distinguish a vehicle whose engine was recently running from a cold building. Alternatively, MWIR will likely fail during the day when temperatures are elevated for the entire scene, whereas HSI will likely operate better since it can obtain more reflectance. This example demonstrates how it would be useful to incorporate and fuse the information provided by multiple sensor modalities to reduce the uncertainty in a measurement or obtain a more complete understanding of a scene which can then be used to improve classification or segmentation decisions.

–(**Current Assumptions/ Issues**)

Information fusion approaches make two typical assumptions. (1) If fusing multiple heterogeneous sources (varying types and resolutions), it is assumed that individual data points can be co-registered (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**. Assumption (2) states that training labels are available for each data point **CITE**.

–(**Loss of information, noisy transformations**)

Two problems arise from these assumptions. First, when working with sensors operating at varying spatial and temporal resolutions, it is not necessarily feasible to convert all data to the same resolution or to map to the same grid **CITE** and existing co-registration approaches often result in loss of sensor-specific information **CITE**. In the previous ATR example, it was observed that the MWIR and HSI sensors would demonstrate varying levels of operation based on environmental conditions. It is then intuitive that external meta-data should also be incorporated to provide confidence bounds on each sensor's measurement capabilities. This, of course, adds an additional level of difficulty to the fusion process. Where just two sensors of varying resolutions were mapped to the same grid, we now have to incorporate time-of-day, weather, ground temperature, etc., all of which likely operate with different spatial or temporal rates and may or may not align easily. Typical approaches would throw out data points to match the lowest sampling rate **CITE**. This mapping is often noisy and results in loss of sensor-specific information.

–(**Physical limitations, cost of labeling**)

Additionally, even assuming that there is a noiseless/ lossless way to co-register heterogeneous data, standard supervised learning methods require accurate labels for each training data point**CITE**. However, data-point specific labels are often unavailable or difficult and expensive to obtain **CITE**. For example, when attempting to detect explosive ordinance with a hand-held electromagnetic induction sensor (metal detector), it is difficult to estimate the expected response from a target due to variations in target size, soil conditions, etc. Additionally, supplementary Global Positioning Systems (GPS) are only accurate on the level of several meters.

These effects make it very difficult (potentially even impossible) to obtain exact, sample-level labels for a training target, and thus add an additional level of geometric uncertainty to the sensor fusion process.

–(**Statement of Work**)

To address these two problems, I propose the following. 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. Roughly, the following research questions will be addressed during the scope of this project:

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? How do we obtain representative, quantitative evaluation of performance?

Experiments will be conducted on both synthetic data and real applications such as plant phenotyping, as well as 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