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

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 [1, 2]. Fusion of multiple sensor sources providing complimentary or reinforcing information is often paramount to the success of remote sensing applications. For example, consider the scenario described by Du in [3] where hyperspectral (HSI) and LiDAR (light detection and ranging) sensors are used for scene understanding. Hyperspectral cameras can provide a broad range of spectral information about the materials present in a scene, while LiDAR can provide depth information. If a road and a building rooftop are built with the same material (e.g. asphalt), hyperspectral information alone may not be able to discriminate between the two. However, incorporating elevation information provided by the corresponding LiDAR depth-map would greatly facilitate the classification problem. Alternatively, paved and dirt biking trails might lay at the same elevation, and using LiDAR data alone may not be sufficient to distinguish between the two types of roads. However, the spectral characteristics given by the HSI camera could easily determine an area of ground covered in dirt and another filled by asphalt or cement. This example demonstrates how it could be useful to incorporate and fuse the information provided by multiple sensor modalities to reduce measurement uncertainty or to obtain a more complete understanding of a scene, thus aiding in classification or segmentation decisions.

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 [4]. 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 [1, 5]. Assumption (2) states that precise training labels are available for each data point [3].

Two problems arise from these assumptions. First, when working with sensors operating at varying spatial, spectral, or temporal resolutions, it is usually non-trivial to convert all data to the same resolution or to map to the same grid and existing co-registration approaches often result in loss of sensor-specific information [4, 6]. Referring to the previous scene understanding example, it is intuitive that the addition of external meta-data could supply context which has the ability to aid in decision making or provide confidence bounds on each sensor's measurement capabilities. This addition, of course, adds a new level of difficulty to the fusion process. Instead of just mapping two sensors of varying resolutions to the same grid, we now have to incorporate high-level information such as time-of-day, weather, environmental-setting (e.g. urban or forest), etc., all of which likely operate with different spatial or temporal rates and probably do not align easily. Typical approaches would throw out data points to match the lowest sampling rate. However, this mapping is often noisy and results in loss of sensor-specific information. This loss is often detrimental since classification improvement may only be achieved if the total information uncertainty in a problem is reduced [7]. Moreover, it becomes difficult to distinguish between errors which evolve during the registration process and actual physical differences in a scene [8].

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. However, data-point specific labels are often unavailable or difficult and expensive to obtain [9, 10]. 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. Therefore, fusion methods should be developed which consider this label ambiguity.

To address these two problems, I propose the following. During this project, techniques will be explored for use in multi-sensor, multi-resolution 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. Data manifolds are inherently robust to outlying exemplars. However, current manifold learning procedures do not consider data which is weakly or ambiguously labeled. To address this gap in the literature, a method for manifold construction will be developed under the multiple-instance learning framework. How does this method compare to both traditional and state-of-the-art manifold learning approaches in terms of outlier detection?

2. Can a joint-representation space be developed between multiple sensors such that there is less sensor-specific information lost when compared to current fusion approaches? Is there an appropriate metric to measure loss of information through fusion? If not, what is an appropriate measure?

3. Fusion methods such as the Choquet integral, deep learning and hierarchical mixture of experts often utilize individual processing pipelines on each sensor before combining on the sample, feature or decision level. Will a single, sensor-agnostic processing pipeline on the unified representation (developed in Objective 2) obtain comparable detection/segmentation results to these alternative approaches?

Experiments will be conducted on both synthetic data and real applications such as target detection and scene understanding in remote sensing imagery, plant phenotyping, and semantic segmentation. Datasets will include the DSIAC ATR Algorithm Development Database, MUUFL Gulfport, and Danforth Plant Science Center Manifold Learning datasets. Initial results demonstrate the aptitude of the proposed approaches and suggest further development and evaluation of these methods.

CHAPTER 2
BACKGROUND

2.1 Sensor Fusion

2.1.1 Information Loss

2.1.2 Classic Approaches

2.1.2.1 Choquet Integral

2.1.2.2 Hierarchical Mixture of Experts

2.1.2.3 Deep Learning

2.1.2.4 Graph-Based

2.1.3 Geocoding

2.1.4 Similarity Measures

2.1.5 Transformation, Interpolation, Re-sampling

2.1.6 Conflation

2.2 Manifold Learning

2.2.1 Classic Approaches

2.2.2 Competitive Hebbian Learning

2.2.3 Deep Learning

2.2.4 Manifold Regularization

2.3 Multiple Instance Learning

2.3.1 Multiple Instance Concept Learning

2.3.2 Multiple Instance Classification

2.3.3 Multiple Instance Regression

2.4 Outlier/ Adversarial Detection

CHAPTER 3
PROBLEM DESCRIPTION

CHAPTER 4
TECHNICAL APPROACH

CHAPTER 5
EXPERIMENTAL DESIGN

CHAPTER 6
PRELIMINARY WORK

CHAPTER 7
FUTURE TASKS

CHAPTER 8
CONCLUSIONS

REFERENCES

[1] J. K. Hackett and M. Shah, "Multi-sensor fusion: a perspective," in *Proceedings., IEEE International Conference on Robotics and Automation*, May 1990, pp. 1324–1330 vol.2.

[2] J. Zhang, "Multi-source remote sensing data fusion: Status and trends," *International Journal of Image and Data Fusion*, vol. 1, pp. 5–24, 03 2010.

[3] X. Du, "Multiple instance choquet integral for multiresolution sensor fusion," Ph.D. dissertation, Univ. of Missouri, Columbia, MO, Dec. 2017.

[4] H. Shen, X. Meng, and L. Zhang, "An integrated framework for the spatiotemporalspectral fusion of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7135–7148, Dec 2016.

[5] M. Butenuth, G. v. Gsseln, M. Tiedge, C. Heipke, U. Lipeck, and M. Sester, "Integration of heterogeneous geospatial data in a federated database," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 62, pp. 328–346, 10 2007.

[6] G. Brigot, E. Colin-Koeniguer, A. Plyer, and F. Janez, "Adaptation and evaluation of an optical flow method applied to coregistration of forest remote sensing images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 7, pp. 2923–2939, July 2016.

[7] D. Ruta and B. Gabrys, "An overview of classifier fusion methods," *Computing and Information Systems*, vol. 7, pp. 1–10, 01 2000.

[8] B. Zitov and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977 – 1000, 2003.

[9] X. Du and A. Zare, "Multi-resolution multi-modal sensor fusion for remote sensing data with label uncertainty," *CoRR*, vol. abs/1805.00930, 2018. [Online]. Available: http://arxiv.org/abs/1805.00930

[10] X. Du and A. Zare, "Multiple instance choquet integral classifier fusion and regression for remote sensing applications," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 5, pp. 2741–2753, May 2019.