

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	5
LIST OF FIGURES	6
CHAPTER	
1 INTRODUCTION	7
2 BACKGROUND	12
2.1 Manifold Learning	13
2.1.1 Classic Approaches	13
2.1.1.1 Principal Component Analysis	13
2.1.1.2 Multidimensional Scaling	13
2.1.1.3 Isometric Feature Mapping	13
2.1.1.4 Locally Linear Embedding	13
2.1.1.5 Laplacian Eigenmaps	13
2.1.1.6 Diffusion Maps	13
2.1.1.7 Latent Variable Models	13
2.1.1.8 Competitive Hebbian Learning	13
2.1.1.9 Global vs Local Methods	13
2.1.2 Supervised and Semi-Supervised Approaches	13
2.1.3 Manifold Regularization	13
2.2 Multiple Instance Learning	13
2.2.1 Multiple Instance Concept Learning	13
2.2.2 Multiple Instance Classification	13
2.2.3 Multiple Instance Regression	13
2.3 Sensor Fusion	13
2.3.1 Classic Approaches	16
2.3.2 Feature-level Fusion	16
2.3.3 Manifold Alignment	16
2.4 Data Processing on Graphs	16
2.4.1 Classic Approaches	16
2.4.2 Graph Convolutional Neural Networks	16
3 PROBLEM DESCRIPTION	17
4 EXPERIMENTAL DESIGN	18
5 PRELIMINARY WORK	19
6 FUTURE TASKS	20

7	CONCLUSIONS	21
	APPENDIX	
	REFERENCES	22

LIST OF TABLES

Table

page

LIST OF FIGURES

Figure

page

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.

Additionally, data fusion can provide confidence bounds on sensor's collections [1, 4]. For example, HSI sensors measure the radiance across a broad range of wavelengths at varying ground locations. The measured radiance is a combination of radiation that is reflected and/or emitted by materials on the ground. This inherently implies that environmental characteristics such as atmospheric conditions and diurnal cycle affect the amount of radiance that can be captured by the sensor [5]. Therefore, incorporation of meta-data such as time-of-day and weather conditions can provide bounds on the uncertainty contained in a given measurement and reinforce when to trust readings from particular spectral bands (i.e. IR during night). These examples demonstrate 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 [6]. 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, 7]. 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 [6, 8]. 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 since high-level information such as time-of-day, weather, environmental-setting (e.g. urban or forest), etc., would likely be collected at different spatial or temporal rates and would probably not align easily. Typical approaches would address this problem by throwing 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 [9]. Moreover, it becomes difficult to distinguish between errors which evolve during the registration process and actual physical differences in a scene [10].

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 [11, 12]. 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. Even if the true size and location of a target is known, estimating the spatial expanse of the response is nontrivial and subject to inaccuracies which could be detrimental to the algorithmic training process. Additionally, labeling high-resolution imagery on the sample-level is tedious, time-consuming, and costly in terms of resources. There is often an unrealistic assumption that labelers are domain experts. Even among experts, label boundaries are often subjective and varying levels of uncertainty in the true class exist between different labelers. 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.

One method for fusion of heterogeneous data sources which has gained popularity in previous years is *manifold alignment* (MA) [13]. The goal of MA is to match two or more seemingly disparate datasets by mapping them to a joint latent space, while both preserving the qualities of each dataset and highlighting their similarities [14, 15, 16]. This mapping creates a joint feature space which can then be utilized by more traditional learning algorithms [13]. The underlying assumption of MA is that, while datasets might capture different qualities presented in a scene, objects under observation should demonstrate similar latent properties independent of modality. Thus, a single manifold can be discovered from multiple sensors which accurately describes the information present in a scene. [17]. While MA has been explored for domain adaptation and transfer learning, few studies have investigated MA applied to remote sensing applications [18, 19, 20]. Furthermore, MA suffers from the same problems mentioned previously. Traditional methods for MA assume that accurate correspondence between datasets is available [14]. While unsupervised MA techniques have been proposed, results demonstrate the advantage of incorporating correspondence points [18, 16, 13]. It was shown in [19] that MA can be extended to incorporate label correspondence instead of instance-level pairs. However, in many remote sensing scenarios we do not know the

correspondence between data points and accurate labels for supervised learning might not be available. This uncertainty should be incorporated in the model [21]. In agreement, [15] recently noted that completely accurate correspondence might not be needed for MA. This poses the question, *can we take advantage of weak labels to generate correspondence points?* Additionally, *if the purpose of fusion is to ultimately perform a classification task, this suggests that label information should be used to drive class representations further from each other in the latent space* [20].

To address these two points, 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, LiDAR and more. 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 a joint-representation space be developed between multiple sensors through incorporation of weakly and ambiguously labeled data? A method for generating correspondence points from weak labels will be proposed and a manifold alignment method will be adapted to incorporate the weak correspondences. How does this data alignment compare to state-of-the-art manifold and graph isomorphism-based alignment methods?
2. Supervised and semi-supervised manifold learning has proven effective at discovering low-dimensional data representations which provide adequate class separation in the latent space. 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 weakly-supervised manifold learning will be developed and applied in conjunction to the manifold alignment procedure proposed in Objective 1. How does this method of manifold construction compare to state-of-the-art manifold learning techniques?
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 the previous objectives) 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

This chapter provides a literature review on Manifold Learning, including classic approaches, supervised and semi-supervised methods and uses of manifolds for functional regularization. A review of the Multiple Instance Learning framework for learning from weak and ambiguous labels is provided. Additionally, this chapter reviews the existing literature in sensor fusion, focusing heavily on manifold alignment and fusion of heterogeneous and multi-resolution data types. A brief review on data processing over graphs is also included. Reviews describe basic terminology and definitions. Foundational approaches are described and advances are addressed.

2.1 Manifold Learning

2.1.1 Classic Approaches

2.1.1.1 Principal Component Analysis

2.1.1.2 Multidimensional Scaling

2.1.1.3 Isometric Feature Mapping

2.1.1.4 Locally Linear Embedding

2.1.1.5 Laplacian Eigenmaps

2.1.1.6 Diffusion Maps

2.1.1.7 Latent Variable Models

2.1.1.8 Competitive Hebbian Learning

2.1.1.9 Global vs Local Methods

2.1.2 Supervised and Semi-Supervised Approaches

2.1.3 Manifold Regularization

2.2 Multiple Instance Learning

2.2.1 Multiple Instance Concept Learning

2.2.2 Multiple Instance Classification

2.2.3 Multiple Instance Regression

2.3 Sensor Fusion

Sensor fusion is the process of combining multiple sensors to provide complimentary or reinforcing information about a scene under observation [1, 2]. Sensor fusion can operate on three levels: data or sample level, feature level, and decision level. This means that raw data, extracted features, or classifier prediction outputs can all be leveraged through a fusion process [9, 2]. While the process of sensor fusion is, in fact, general, this dissertation focuses primarily on combining sensors of varying types and resolutions with the goal of providing support for classification and detection decisions. In this context, sample-level fusion aims to combine data from multiple sources into single-resolution data which are expected to be more informative than any of the inputs, individually. Fusion on this level

involves applying a transformation on the data directly to project, map or co-register data obtained from multiple sources. Feature-level fusion combines attributes computed from each sensors' collection, such as shape, texture, frequency, local statistics, or other discriminative characteristics exhibited in the data. Feature-level fusion transforms the data into a single representation space which can be used instead of the original spaces for further processing. Decision-level fusion takes outputs from multiple classifier or real-valued prediction models and derives a singular, fused output. Decision-level fusion aims to provide a more accurate and/or descriptive classification or regression result through selection, transformation, or consensus of the models being combined [1, 2, 3, 22, 23, 11]. Sensor fusion has been used in a wide expanse of applications, including: remote sensing and hyperspectral image registration, classification and visualization, [24, 15, 20, 25], object detection [26], explosive hazard detection [27, 28, 29, 30, 31], handwriting recognition [32], pose estimation [33], automatic target recognition in mid-wave infrared/ sonar imagery [17], land-use classification [13], super-resolution, medical and fault diagnosis, defense, meteorology, environmental monitoring, and economic forecasting [2, 10].

Challenges in sensor fusion Features might not have the same dimensionality between sensors. The number/ length of features/ physical qualities vary between modalities. Data from different sensors must be put into equivalent forms to allow for fusion. In order for data from multiple sources to be fused, there must be some method to relate data points from one sensor with corresponding data points from the other sensors. Before sensor measurements can be combined, we must ensure that the measurements represent the same physical entity. Therefore, we need to check the consistency of sensor measurements. Since each sensor is sensitive to a different modality, multiple sensors not only can provide multiple views of objects, but they can also impose more constraints to reduce the search space during matching.

Additionally, data fusion can provide confidence bounds on sensor's collections [1, 4]. For example, HSI sensors measure the radiance across a broad range of wavelengths at varying

ground locations. The measured radiance is a combination of radiation that is reflected and/or emitted by materials on the ground. This inherently implies that environmental characteristics such as atmospheric conditions and diurnal cycle affect the amount of radiance that can be captured by the sensor [5]. Therefore, incorporation of meta-data such as time-of-day and weather conditions can provide bounds on the uncertainty contained in a given measurement and reinforce when to trust readings from particular spectral bands (i.e. IR during night). These examples demonstrate 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 [6]. 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, 7]. 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 [6, 8]. 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 since high-level information such as time-of-day, weather, environmental-setting (e.g. urban or forest), etc., would likely be collected at different spatial or temporal rates and would probably not align easily. Typical approaches would address this problem by throwing 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 [9]. Moreover, it becomes difficult to distinguish between errors which evolve during the registration process and actual physical differences in a scene [10].

2.3.1 Classic Approaches

2.3.2 Feature-level Fusion

2.3.3 Manifold Alignment

Manifold alignment is the process of matching two or more seemingly disparate datasets by mapping them to a joint latent space, while both preserving the qualities of each dataset and highlighting their similarities [14, 15, 16].

2.4 Data Processing on Graphs

2.4.1 Classic Approaches

2.4.2 Graph Convolutional Neural Networks

CHAPTER 3

PROBLEM DESCRIPTION

CHAPTER 4

EXPERIMENTAL DESIGN

CHAPTER 5

PRELIMINARY WORK

CHAPTER 6

FUTURE TASKS

CHAPTER 7

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] M. Mohandes, M. Deriche, and S. O. Aliyu, "Classifiers combination techniques: A comprehensive review," *IEEE Access*, vol. 6, pp. 19 626–19 639, 2018.
- [5] A. Zare, "Hyperspectral endmember detection and band selection using bayesian methods," Ph.D. dissertation, Univ. of Florida, Gainesville, FL, 2008.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] D. Ruta and B. Gabrys, "An overview of classifier fusion methods," *Computing and Information Systems*, vol. 7, pp. 1–10, 01 2000.
- [10] B. Zitov and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977 – 1000, 2003.
- [11] 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>
- [12] 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.
- [13] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. X. Zhu, "Learnable manifold alignment (lema): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 147, pp. 193 – 205, 2019.

- [14] C. Wang, P. Krafft, and S. Mahadevan, "Manifold alignment," 12 2011.
- [15] D. Liao, Y. Qian, J. Zhou, and Y. Y. Tang, "A manifold alignment approach for hyperspectral image visualization with natural color," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3151–3162, June 2016.
- [16] J. S. Stanley III, G. Wolf, and S. Krishnaswamy, "Manifold alignment with feature correspondence," *CoRR*, vol. abs/1810.00386, 2018. [Online]. Available: <http://arxiv.org/abs/1810.00386>
- [17] D. Shen, P. Zulch, M. Disasio, E. Blasch, G. Chen, Z. Wang, J. Lu, and R. Niu, "Manifold learning algorithms for sensor fusion of image and radio-frequency data," in *2018 IEEE Aerospace Conference*, March 2018, pp. 1–9.
- [18] C. Wang and S. Mahadevan, "Multiscale manifold alignment," 09 2010.
- [19] —, "Heterogeneous domain adaptation using manifold alignment," in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two*, ser. IJCAI'11. AAAI Press, 2011, pp. 1541–1546. [Online]. Available: <http://dx.doi.org/10.5591/978-1-57735-516-8/IJCAI11-259>
- [20] H. L. Yang and M. M. Crawford, "Spectral and spatial proximity-based manifold alignment for multitemporal hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 51–64, Jan 2016.
- [21] A. Damianou, N. D. Lawrence, and C. H. Ek, "Manifold alignment determination: finding correspondences across different data views," 01 2017.
- [22] S. Tulyakov, S. Jaeger, V. Govindaraju, and D. Doermann, *Review of Classifier Combination Methods*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 361–386. [Online]. Available: https://doi.org/10.1007/978-3-540-76280-5_14
- [23] X. Du, A. Zare, J. M. Keller, and D. T. Anderson, "Multiple instance choquet integral for classifier fusion," in *2016 IEEE Congress on Evolutionary Computation (CEC)*, July 2016, pp. 1054–1061.
- [24] Y. Zhang, X. Zheng, G. Liu, X. Sun, H. Wang, and K. Fu, "Semi-supervised manifold learning based multigraph fusion for high-resolution remote sensing image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 2, pp. 464–468, Feb 2014.
- [25] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "Cospace: Common subspace learning from hyperspectral-multispectral correspondences," *CoRR*, vol. abs/1812.11501, 2018. [Online]. Available: <http://arxiv.org/abs/1812.11501>
- [26] M. A. Davenport, C. Hegde, M. F. Duarte, and R. G. Baraniuk, "Joint manifolds for data fusion," *IEEE Transactions on Image Processing*, vol. 19, no. 10, pp. 2580–2594, Oct 2010.

- [27] P. Gader, A. Mendez-Vasquez, K. Chamberlin, J. Bolton, and A. Zare, "Multi-sensor and algorithm fusion with the choquet integral: applications to landmine detection," in *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium*, vol. 3, Sep. 2004, pp. 1605–1608 vol.3.
- [28] J. E. B. A. Z. B. A. Ryan E. Smith, Derek T. Anderson, "Aggregation of choquet integrals in gpr and emi for handheld platform-based explosive hazard detection," in *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XXII*, 1018217, vol. 10182, May 2017. [Online]. Available: <https://doi.org/10.1117/12.2263005>
- [29] J. M. Keller, S. Auephanwiriyakul, and P. D. Gader, "Experiments inf predictive sensor fusion," in *Detection and Remediation Technologies for Mines and Minelike Targets VI*, ser. Proc.SPIE, vol. 4394, Oct. 2001, pp. 1047 – 1058.
- [30] A. H. Gunatilaka and B. A. Baertlein, "Feature-level and decision-level fusion of noncoincidently sampled sensors for land mine detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 577–589, June 2001.
- [31] H. Frigui, L. Zhang, and P. D. Gader, "Context-dependent multisensor fusion and its application to land mine detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2528–2543, June 2010.
- [32] P. D. Gader, M. A. Mohamed, and J. M. Keller, "Fusion of handwritten word classifiers," *Pattern Recognition Letters*, vol. 17, no. 6, pp. 577 – 584, 1996.
- [33] R. Navaratnam, A. W. Fitzgibbon, and R. Cipolla, "The joint manifold model for semi-supervised multi-valued regression," in *2007 IEEE 11th International Conference on Computer Vision*, Oct 2007, pp. 1–8.