Signal Processing 126 (2016) 35–51

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Distributed information-based clustering of heterogeneous sensor data

Jia Chen, Ioannis D. Schizas *,[1]

University of Texas at Arlington, Department of Electrical Engineering, 416 Yates Street, Arlington, TX 76010, USA

ARTICLE INFO

Article history:
Received 14 June 2015
Received in revised form
4 December 2015
Accepted 28 December 2015
Available online 12 January 2016

Keywords:
Canonical correlation analysis
Heterogeneous sensor network
Clustering
Regularization

ABSTRACT

Heterogeneous sensing systems, consisting of sensors with different sensing capabilities, offer flexibility and provide multiple views of the sensed field by acquiring different types of measurements. The acquired sensor measurements are affected by different and unknown in number phenomena/sources of interest. To this end, a novel canonical correlation analysis (CCA) framework equipped with norm-one and norm-two regularization terms is designed to cluster the sensor data based on their information content. Block coordinate descent (BCD) is combined with the alternating direction method of multipliers (ADMM) framework to derive a centralized algorithm tackling the novel regularized CCA framework. Further, splitting of the regularized CCA into localized minimization subtasks across sensors enables distributed clustering of heterogeneous data based on their information content. Numerical tests demonstrate that the novel framework can achieve higher probability of correct clustering than existing alternatives.

1. Introduction

The utilization of sensors with different types of sensing capabilities in heterogeneous sensor networks provide different 'views' of the sensed field by acquiring different types of measurements. However, the sensed data oftentimes are collected in challenging environments, whose statistical structure is not known and maybe dynamically changing with time [29]. The acquired sensor measurements are affected by different and unknown in number phenomena of interest such as thermal and/or pollution sources [30,23,26], while other sensor data may just contain noise, e.g., a vehicle could be sensed by both a thermal sensor due to the heat it produces and a carbon monoxide sensor due to fuel emissions. Thus, sensor measurements can be clustered into different groups each of which will mainly contain information about the present field sources. Before applying any data processing task such as estimation and detection [20] it is essential to develop techniques to identify and separate these unknown groups of sensor measurements that have the same information content. Such schemes will prevent mixing sensor measurements with irrelevant information content before applying any statistical inference task, while further it will identify regions of interest in the sensed field, such as pollution sources [23,26], by interpreting the acquired data.

The focus is to match different types of sensor measurement groups based on their information source content. However, this is challenging due to the absence of sensor localization information due to the cost and energy considerations imposed by GPS equipment, see e.g., [12]. To this end, we develop here an algorithmic framework that has the ability to both identify, cluster and match different types of sensor data based on their information content. Clustering and matching of different types of

* Corresponding author.
E-mail addresses: jia.chen81@mavs.uta.edu (J. Chen), schizas@uta.edu (I.D. Schizas).

[1] Work in this paper is supported by the NSF Grant 1509780 and University of Texas at Arlington.

http://dx.doi.org/10.1016/j.sigpro.2015.12.017
0165-1684/

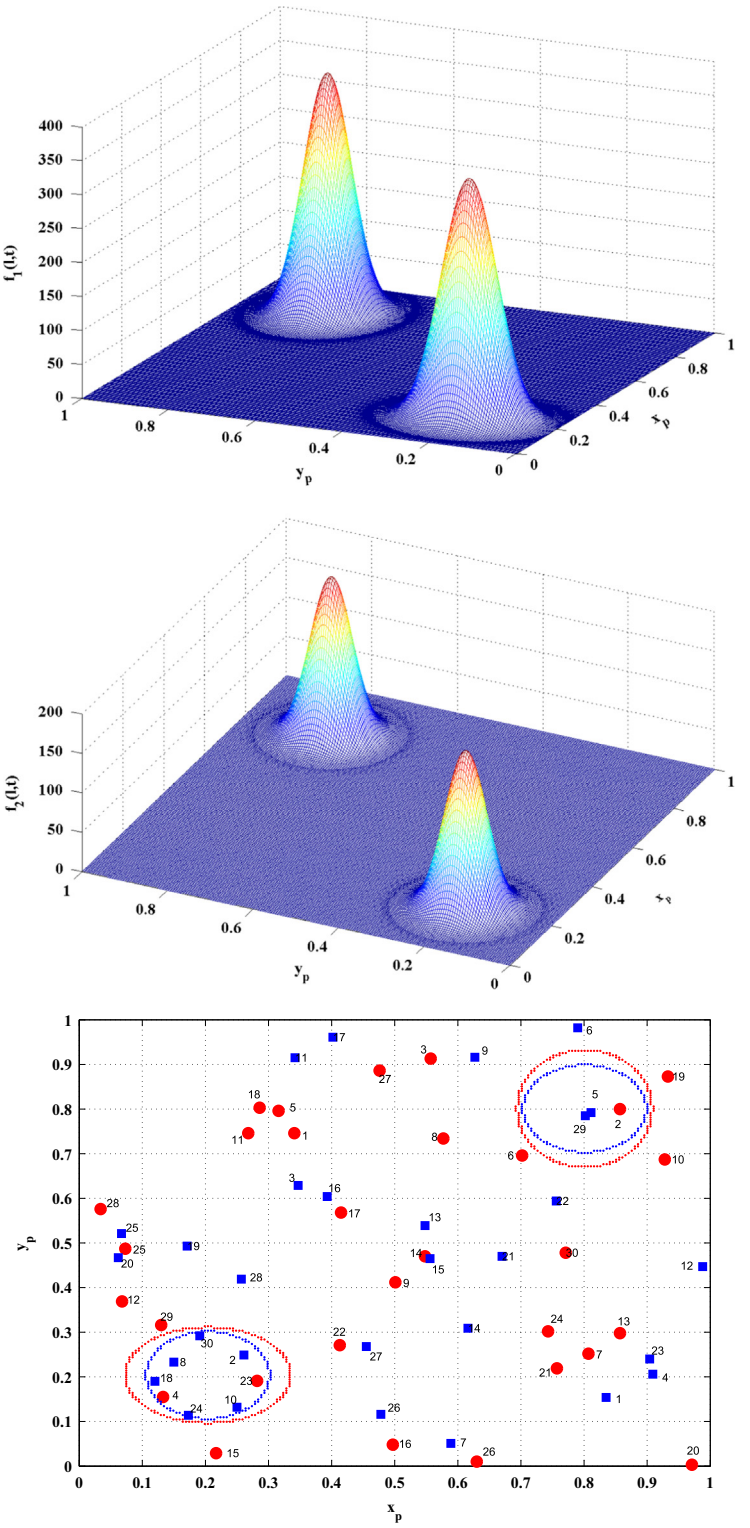

Fig. 1. Diffusion field examples (top, center) sensed by a heterogeneous sensor network with two types of sensing units (bottom). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

sensor measurements is extremely useful to learn the sensed field and categorize the data based on the information they contain.

Field sources in practice are localized and affect a small portion of the acquired sensor measurements. Further, the sensor measurements containing information about the