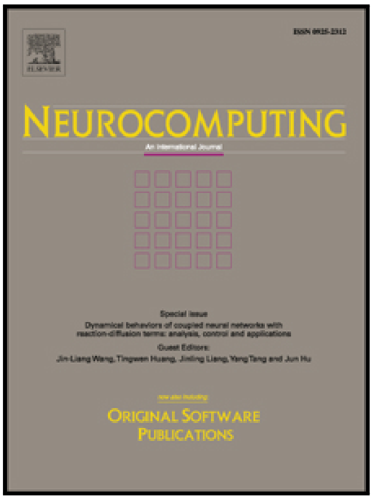
Communicated by Prof. Liu Guangcan

Accepted Manuscript



Fast Subspace Segmentation via Random Sample Probing

Yang Li, Yubao Sun, Qingshan Liu, Shengyong Chen

PII: S0925-2312(18)31016-6
DOI: https://doi.org/10.1016/j.neucom.2018.08.054
Reference: NEUCOM 19902

To appear in: *Neurocomputing*

Received date: 6 March 2018
Revised date: 26 August 2018
Accepted date: 28 August 2018

Please cite this article as: Yang Li, Yubao Sun, Qingshan Liu, Shengyong Chen, Fast Subspace Segmentation via Random Sample Probing, *Neurocomputing* (2018), doi: https://doi.org/10.1016/j.neucom.2018.08.054

Fast Subspace Segmentation via Random Sample Probing

Yang Li[a], Yubao Sun[b], Qingshan Liu[c], Shengyong Chen[d],

[a]*Y. Li is with the Key Laboratory of Computer Vision and System, Ministry of Education, Department of Computer Science and Engineering, Tianjin University of Technology, No 319 Binshui West Road, Tianjin, China 300384. E-mail: 163102403@stud.tjut.edu.cn.*
[b]*Y. Sun is with B-DAT and CICAEET, School of Information and Control, Nanjing University of Information Science and Technology, 219 Ningliu Road, Nanjing, Jiangsu, China 210044. E-mail: ybsun@nuist.edu.cn.*
[c]*Q. Liu is with B-DAT, School of Information and Control, Nanjing University of Information Science and Technology, Nanjing, China 210044. Email: qsliu@nuist.edu.cn.*
[d]*S. Chen is with the Key Laboratory of Computer Vision and System, Ministry of Education, Department of Computer Science and Engineering, Tianjin University of Technology, No 319 Binshui West Road, Tianjin, China 300384. E-mail: csy@tjut.edu.cn.*

Abstract

Subspace segmentation is to group a given set of n data points into multiple clusters, with each cluster corresponding to a subspace. Prevalent methods such as Sparse Subspace Clustering (SSC) and Low-Rank Representation (LRR) are effective in terms of segmentation accuracy, but computationally inefficient while applying to gigantic datasets where n is very large as they possess a complexity of $O(n^3)$. In this paper, we propose an iterative method called Random Sample Probing (RANSP). In each iteration, RANSP finds the members of one subspace by randomly choosing a data point (called "seed") at first, and then using Ridge Regression (RR) to retrieve the other points that belong to the same subspace as the seed. Such a procedure is repeated until all points have been classified. RANSP has a computational complexity of $O(n)$ and can therefore handle large-scale datasets. Experiments on synthetic and real datasets confirm the effectiveness and efficiency of RANSP.

Keywords: clustering, subspace segmentation, large-scale, random sample probing

1. Introduction

Subspace segmentation assumes that the data points can be clustered into different groups and every group can be fitted with a low dimensional subspace, which has been widely used in serval areas such as machine learning, signal processing, data mining, and machine vision [1, 2, 3, 4, 5, 6].

In the past several years, subspace segmentation have attracted much attention, and many statistical methods have been proposed and investigated in the literature. To deal with low-dimensional data (e.g., 3D data in computer graphics), Random Sample Consensus (RANSAC) [7], which estimate the model of subspaces is an iterative fashion, is probably the most widely used method due to its high efficiency and effectiveness. However, RANSAC is inappropriate for high-dimensional data, because its time consumption will increase exponentially as the dimension of data. Moreover, the segmentation accuracy produced by RANSAC may drop fast when the data dimension goes high. K-Subspace [8], a generalization of the well-known K-Means [9] method, may apply to high-dimensional data, but suffers from the problem of local minimum and thus can only achieve limited performance in many subspace