The unprecedented growth of data in volume and dimension has led to an increased number of computationally-demanding and data-driven decision-making methods in many disciplines, such as computer vision, genomics, finance, etc. Research on big data aims to understand and describe trends in massive volumes of high-dimensional data. High volume and dimension are the determining factors in both computational and time complexity of algorithms. The challenge grows when the data are formed of the union of group-structures of different dimensions embedded in a high-dimensional ambient space. To address the problem of high volume, we propose a sampling method referred to as the Sparse Withdrawal of Inliers in a First Trial (SWIFT), which determines the smallest sample size in one grab so that all group-structures are adequately represented and discovered with high probability. The key features of SWIFT are: (i) sparsity, which is independent of the population size; (ii) no prior knowledge of the distribution of data, or the number of underlying group-structures; and (iii) robustness in the presence of an overwhelming number of outliers. We report a comprehensive study of the proposed sampling method in terms of accuracy, functionality, and effectiveness in reducing the computational cost in various applications of computer vision. In the second part of this dissertation, we study dimensionality reduction for multi-structural data. We propose a probabilistic subspace clustering method that unifies soft- and hard-clustering in a single framework. This is achieved by introducing a delayed association of uncertain points to subspaces of lower dimensions based on a confidence measure. Delayed association yields higher accuracy in clustering subspaces that have ambiguities, i.e. due to intersections and high-level of outliers/noise, and hence leads to more accurate self-representation of underlying subspaces. Altogether, this dissertation addresses the key theoretical and practically issues of size and dimension in big data analysis.

Major: Computer Science

Educational Career:
Bachelor's of Computer Engineering - Software, BS, 2004, Kharazmi University of Tehran
Master's of Computer Engineering - Artificial Intelligence, MS, 2007, Amirkabir University of Technology Tehran
Master's of Computer Science, MS, 2012, University of Nevada, Reno

Committee in Charge:
Hassan Foroosh, Chair, Computer Science
Marianna Pensky, Co-Chair, Mathematics
Boqing Gong, Computer Science
Guojun Qi, Computer Science

Approved for distribution by Hassan Foroosh, Committee Chair, on March 6, 2018.

The public is welcome to attend.