1 Workshop Description

The explosive growth of emerging applied problems based on acquired massive data and its exploitation has challenged researchers and practitioners alike. The daunting size of the data in combination with its diversity and complexity have increasingly led to seeking intrinsic structure within the data to explain and parsimoniously reflect information useful to its exploitation. In particular, geometric structure which underlies data, has emerged as a promising, economical and effective tool to unveil new information about the data and shed new light on the problem.

Curves for instance, have over the last decade been of crucial importance in investigating localization and tracking issues in tumors in medical images, or in detecting and tracing the spatial evolution of holes in the ozone layer in environmental sciences. Statistical or other characteristics shared by a fraction of the data which translate into a coherent geometrical entity affords a simpler and perhaps more direct means of "divide to conquer" and interpret otherwise complex and daunting data. While such a curve-based approach has proved very successful, its application to higher dimensional data with more complex geometry is very limited. A major challenge comes from the extensive reliance upon the arclength parameterization in geometric statistical analysis of contours which does not easily generalize in computationally tractable ways to surfaces of higher dimension.

Difficult challenges arise in many applications where the curse of dimensionality very quickly becomes a the limiting issue, starting with 3D bodies to other processes lying in higher dimensions but yet associated with common characteristics which may be used to advantage. One such example, relevant in security applications is the characteristic space where human face data lie. Such data measurably lies in thousands of dimensions, while the information-bearing submanifold dimension is in reality but a fraction of that. Such em-

bedding clearly has a significant impact on the subsequent computational load, and is due to Whitney Embedding theorem in differential geometry. Such a theoretical framework in addition illustrates the achievable guidance for optimal implementation and performance of an algorithm to address a practical problem. Other connections which arise in the course of further exploring high dimensional data may be inspired from characteristics like curvature has been advocated and utilized as an invariant for curves, other invariants including topological may be constructed for manifolds. These may in turn play a crucial role in characterizing high dimensional data and their potential comparison to other data structures. Such comparison of data classes which are typically subjected to transformations would require one to look at algebraic structure-preserving morphing formalized in category theory.

Much of this formalism is learned and known in complete isolation even within the mathematical community, and yet applications call for the complete spectrum of methodologies and formalism to synthesize a coherent picture and provide a sound formalism to develop robust, efficient and effective solutions to existing and many upcoming challenges which are substantially data heavy.

2 Organizers

The proposed workshop has the following organizational structure:

Co-General Chairs:

Hamid Krim, Professor, Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC

Michael Kirby, Professor, Department of Mathematics and Department of Computer Science, Colorado State University, Fort Collins, CO

Anthony Yezzi, Professor, School of Electrical and Computer Engineering, Georgia Tech, Atlanta, GA

Local Arrangements:

Michael Kirby, Professor, Mathematics Department, Colorado State University,Fort Collins, CO

All contributors will be by invitation only. Research Directors from other federal agencies may also be attending the meeting.

3 Venue and Facilities for the workshop

Beaver Run Resort and Conference Center
Breckenridge, CO 80424-2115 www.beaverrun.com

4 Relevance of Research Topics and Rationale for a Workshop

Dealing with very high dimensional data has become extremely relevant in imaging applications due to the extensive information capacity of a single digital image and even more so for sequences of digital images. Moreover, advances in computer vision techniques over two decades have yielded powerful algorithms for segmentation, contour detection, surface reconstruction, and motion estimation which represent geometric information with the same high dimensional characteristics as the input image data from which they are derived. However, traditional techniques for processing "non-geometric" high dimensional data (sampled functions over Euclidean spaces for example) fail to preserve important geometric and sometimes even topological properties of structures such as curves, surfaces, and graphs. As such, it will become increasingly important to seek new methodologies which are "geometrically intelligent" for processing high dimensional data with relevant and interesting geometric content.

5 Emerging Challenges in Application Areas

High bandwidth data processing problems are one of the major challenges facing our data rich society. From scientific and military applications to commercial and financial applications researchers are seeking information hidden in massive data sets. Indeed, national security interests and other national

priorities (global weather shifting, emerging sources of energy, natural disaster management) hinge on our ability to collect, process and react to massive quantities of information. We envision that the topics developed in this workshop will have impact on a variety of applications including, but not limited to,

- pattern recognition
- target tracking
- video processing
- image analysis
- data compression
- classification
- severe weather prediction
- bioinformatics
- web and text mining
- segmentation
- low dimensional dynamical modeling
- signal processing
- data security