CSc I6730 Data Reduction in the Physical Sciences

Prof. Irina Gladkova

Phone: (212) 650-6135, e-mail: gladkova@cs.ccny.cuny.edu

Course Description: This course will cover both lossless and lossy compression techniques with application to satellite data of various kind. We will introduce and discuss concepts from Shannon's treatment of information theory: the basic notions of entropy, relative entropy, and mutual information, and show how they arise as natural answers to questions of data compression. We will consider different approaches to compression ranging from entropy coding (Huffman, Shannon-Fano, Rice) to transform coding (discrete cosine, Karhunen-Loéve, wavelet). Current and developing recommendations of the Consultative Committee for Space Data Systems (CCSDS) for the satellite data compression will also be presented in this course. Projects and presentations will be based on the variety of earth science data from both polar and geostationary orbiting satellites.

Required text: Khalid Sayood, Introduction to Data Compression, Third Edition, Morgan Kaufmann Series in Multimedia Information and Systems, ISBN-10: 012620862X

Supplementary text:
Nicholas Short, Remote Sensing Tutorial, available at http://rst.gsfc.nasa.gov/
Claude E. Shannon, A mathematical theory of communication. Bell System Technical Journal (available at http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf)

Grading: Final Exam (40%), project (40%) and group presentations (20%).
The students will be evaluated based on presentations, report based projects, and a final exam. Some of these activities will be group based, for instance groups of students could be asked to prepare presentations, as well as being prepared to answer questions individually of the group. Example presentations may require reading research papers and technical reports and presentation topics may cover Satellite instruments, data, or algorithms related to the course. Projects will focus on hands on experience with algorithms, for example comparing performance attributes of these algorithm applied to Satellite data covered in the course.

Course Outcomes:

1) knowledge of data compression methods
2) an ability to understand the relative strengths of compression techniques as applied to satellite data
3) acquire experience communicating scientific results effectively
4) acquire experience working effectively as part of the team

Topics covered:

- Entropy, Relative Entropy, and Mutual Information
- Compression Techniques: Lossy, Lossless, Measures of Performance
- Uniquely decodable, prefix codes
- Huffman, Shannon-Fano Codes
- Golomb, Rice Codes
- Current CCSDS Recommendation for Lossless Compression of Satellite Data
- Arithmetic Code
- Adaptive Entropy Codes
- Lempel-Ziv Code
- Burrows-Wheeler Transform
- Mathematical Preliminaries: Vector Spaces, Linear Transforms
- Vector Quantization, Linde-Buzo-Gray (LGB) Algorithm
- Transform Coding
- DCT transform, JPEG
- Karhunen-Loéve Transform, PCA
- Wavelet-based compression
- CCSDS Recommendation for future Satellite Data compression

Data: Projects and presentations will be based on a representative selection of earth science data from both polar and geostationary orbiting satellites such as:

Advanced Very High Resolution Radiometer (AVHRR),

Multi-angle Imaging SpectroRadiometer (MISR),

Current geo-stationary (GOES) imager,

Moderate Resolution Imaging Spectroradiometer (MODIS) imager,

Atmospheric Infrared Sounder (AIRS).