UROP Project Proposal

**Project Title:** Investigate various methods of lossy compression, which may allow for a more efficient storage mechanism.

**Principal Investigator:** Hemal A Lal

**Faculty Sponsor:** Dr. Carolyn Crouch

**Budget:** $1,700 ($1,400: Student Salary + $300: Equipment).

**Projected Period:** 21 weeks.

1. Project Description

Today’s Intelligent Transportation System (ITS) sensor network, such as traffic sensor network, generates very large amounts of data [1,2]. These data have been the subject of much interest and analysis in both academia and industry [3]. As huge amounts of data have been stored in traffic and transportation databases, data warehouses, geographic information systems, and other information repositories, archiving and retrieving of these data are becoming increasingly difficult [3,4,5]. In this project I plan to apply the lossy compression algorithms to the domain set, which in this case is the traffic data. In a lossy compression, some of the specifics of the data are lost, but the compressed result should nevertheless be close to the original data (and hopefully be much smaller in size).

While a variety of lossy compression schemes have been developed for certain forms of digital data (e.g., images, audio, video), the area of lossy compression techniques for arbitrary data tables is relatively unexplored. Nevertheless, such techniques are clearly motivated by the ever-increasing data collection rates of modern enterprises and the need for effective, guaranteed-quality approximate answers to queries over massive relational data sets [4].

There will be three major phases in the project. I will learn about the various lossy compression algorithms, their advantages and disadvantages, and filter them in respect to the
domain. Second, I will employ these filtered algorithms with other requirements over the domain producing new data sets. The final phase will be to compare the resultant data sets with the original data sets for size, efficiency, and reliability and see if the resultant losses are acceptable.

The end product of this project will be to determine if it is possible to produce a new compressed data set, which is similar to the original one but is much smaller in size.

2. Timetable

I am planning to complete this project in about 21 weeks.

(a) Week 1-3: Background Research and evaluating the data
(b) Week 4-8: Learning general algorithms for lossy compression
(c) Week 9-13: Applying the learned algorithms to domain set
(d) Week 14-19: Analyzing the results with respect to general use of the data
(e) Week 20-21: Preparation of Final Report

3. Budget

The budget for the project includes the following expenses:

1. The research budget for student salary is about $1400.00

2. Equipment:

   (a) Software: $200.00 commercial compression software.
   (b) Relevant text books and supplies: $100.00

Thus the budget for equipment is $300.00

4. Faculty Sponsor

The research will be under observation and sponsored by Associate Professor Carolyn Crouch, Department of Computer Science at University of Minnesota Duluth. Her research
interests are text analysis and processing, data mining, and web-based retrieval. Her interests provide an excellent background for this project.

5. Educational Objectives

My interest in data archival and retrieval and a thesis by my friend Nirish Dhruv interested me in the research done by Dr. Carolyn Crouch. The main educational objective is conduct research and to gain an understanding of the procedures involved with data compression in general and lossy compression in particular. Further objectives include comparison of the advantages and disadvantages found with the different lossy compression algorithms.

6. Reference


