CERTIFICATE

This is to certify that, this thesis entitled “**An Approach to Data compression Technique**” which is submitted by Tarun Kumar Chaudhary, roll number 17/SE/2009; in partial fulfilment of the requirement for the award of degree M. Tech. in Software Engineering to Delhi Technological University, Bawana Road Delhi; is a record of the candidate’s own work carried out by him under my supervision. The matter embodied in this thesis is original as per my knowledge and has not been submitted for the award of any other degree.

June 2011

Guide: **Mr. R. K. Yadav**

Assistant Professor, Department of Computer Engineering

Delhi Technological University

Bawana road, delhi- 110042

ACKNOWLEDGEMENT

This project named “**An Approach to Data Compression Technique”** undertaken at “Delhi Technological University” has been a useful and enlightening experience. I am very thankful to **Mr. R. K. Yadav**, Assistant Professor of Computer Engineering Department, for his guidance and tremendous support. His cooperation and suggestions have made the timely completion of our project possible.

I am also thankful to **Dr. Daya Gupta** H.O.D, Computer Engineering department, for her support and guidance.

I would also like to extend my sincere thanks to all the faculty and Instructor of our Computer Engineering Department, who took interest and made me available the necessary requirement for my project to complete in a better way.

**Date:**

#### **Name of Student**

Tarun Kumar Chaudhary

Reg. No. 12/MT/SE/FT

Roll No. 17/SE/09

TABLE OF CONTENT

1. CERTIFICATE………………………………………………………………………………………………………….i
2. ACKNOWLEDGEMENT………………………………………………………………………………………….ii
3. TABLE OF CONTENT…………………………………………………………………………………………….iii
4. LIST OF FIGURES…………………………………………………………………………………………………...v
5. LIST OF TABLES…………………………………………………………………………………………………….vi
6. ABSTRACT…………………………………………………………………………………………………………..vii
7. **INTRODUCTION……………………………………………………………………………………..1**
   1. Compression Technique…………………………………………………………………….3
      1. Lossless Algorithm ……………………………………………………………………3
      2. Lossy algorithms……………………………………………………………………….3
   2. Motivation…………………………………………………………………………………………5
   3. Goals of the Thesis……………………………………………………………………………..6
   4. Organization of the Thesis………………………………………………………………….7
8. **Literature Survey………………………………………………………..…………………………8**
   1. Data compression Terms……………………………………………………………………10
   2. Categorization of Data compression algorithm…………………………………15
   3. Entropy……………………………………………………………………………………………..16
   4. Complexity ……………………………………………………………………………………….16
   5. Previously Defined algorithms…………………………………………………………..17
      1. Semantic Dependent Methods…………………………………………………17
         1. Run length coding……………………………………………………….17
         2. Difference mapping…………………………………………………….19
      2. Dictionary substitution…………………………………………………………….19
         1. Lempel-ZivWelch Codes………………………………………………….20
      3. Entropy Coding…………………………………………………………………………23
         1. Shannon-Fano Coding…………………………………………..23
         2. Static Huffman Coding…………………………………………..25
         3. Adaptive Huffman Coding……………………………………..28
         4. Arithmetic Coding…………………………………………………30
         5. Range encoding…………………………………………………….32
   6. Comparison between different compression algorithms……………………34
9. **Data Compression Algorithm: Proposed………………………………………………..36**
   1. Previous Techniques………………………………………………………………………….37
   2. Proposed Technique………………………………………………………………………….37
      1. Example related to proposed algorithm……………………………………37
      2. About the algorithm………………………………………………………………….38
   3. Proposed Data Compression algorithm………………………………………………42
      1. Algorithm Compression…………………………………………………………….42
      2. Algorithm Decompression…………………………………………………………45
10. **Result and Performance…………………………………………………………………………47**
    1. Measure of Performance……………………………………………………………………48
    2. Results……………………………………………………………………………………………...49
11. **Limitation…………………………………………..…………………………………………………52**
12. **Conclusion…………………………………………………………………………………………….54**
13. **Future Scope…………………………………………………………………………………..…….56**
14. **References…………………………………………………………………………………………….57**

LIST OF FIGURES

1. Compression and Reconstruction……………………………………………………….9
2. Compression in cascade………………………………………………………………………12
3. Decompression in cascade………………………………………………………………….12
4. Scan used in the run length coding for images…………………………………….18
5. A Shannon-Fano Code…………………………………………………………………………24
6. A Shannon-Fano Code for EXAMPLE..………………………………………………....25
7. The Huffman process……………………………………………………………………..…..26
8. Huffman Coding procedure…………………………………………………………………27
9. The Escape code………………………………………………………………………………...30
10. Arithmetic Coding………………………………………………………………………………31
11. Flow chart Data compression……………………............................................44
12. Flow chart Data decompression………………………………………………………….46
13. Graph between uncompressed file size Vs compression ratio…………….50
14. Graph Compression ratio Vs Compression factor………………………………..50

LIST OF TABLES

1. LZW Dictionary……………………………………………………….21
2. LZW dictionary entry…………………………………………………22
3. Comparison between different compression algorithms…….......34
4. Different Data Compression Algorithm……………………………35
5. Evaluations of four lossless compressions methods………………35
6. Results………………………………………………………………….49

Abstract

This report surveys a variety of data compression methods. Lots of data has been already stored in the physical media. But it is not the end, it keeps increasing with time. Therefore collecting, managing and storing are very important. The aim of data compression is to reduce redundancy in stored or communicated data, thus increasing effective data density. Data compression has important application in the areas of file storage and distributed systems.

Concepts from information theory, as they relate to the goals and evaluation of data compression methods, are discussed briefly. Most of the data compression algorithm developed on entropy based and dictionary based and some of the algorithms are based on the combination.

In growing data environment many data compression came. Many of them work well in their field. Some of them are lossless and some are loosy. Lossy can be applied only on the image and video compression. And lossless can be used where we want our data to be fully recovered, for example text. Some of the data compression algorithm evolved over the time is run length coding, Huffman coding, Shannon-Fano coding, arithmetic coding, LZW coding. For text and video compression we have JPEG, MPEG formats that stores data in very compressible manner. But increasing data forces the researchers to improve their data compression algorithm. That makes us to motivate to design a new algorithm.

This thesis presents a completely new idea and its implemented software of data compression. This data compression uses a very different representation system. Algorithm depends on converting conventional power of 2 bit representation into compressed form of power of 3, bit representation. The representation in power of 3 is another form to represent any number like power of 2 with some slight change. This algorithm is used to applying recursively on partially compressed data that makes it to give theoretically high compression ratio.

By using this representation we are going to show here how the new representation will be used and how it can help in the data compression. And also we have presented an algorithm on the basis of this design. The implemented software application shows how good it is. This algorithm has the capability to give a very highly compressive data. So it has the great chance to be used in many data repository and communication system. And also this algorithm requires lots of work to be done by researchers.