Improved Statistical Transform for Data lossless Compression

Ilias Seddik
University of Sciences and Technologies
USTOMB Oran, Algeria

Redouane Tlemsani
National Institute of
Telecommunications and ICT, Algeria

Khaled Belkadi
University of Sciences and Technologies
USTOMB Oran , Algeria

Abstract— **In this work we will consider solely the problem of lossless compression for many kinds of data with the same dictionary supervised algorithm with a reduced complexity. Data compression is taking an important part in computer research on applications like tele-video- conferencing, remote sensing, document and medical imaging, and facsimile transmission; while our stocking disk and our passing bands are limited. This field regroups two families: Lossy compression, with acceptable loss of information, and lossless compression without any loss. This paper proposes a method can be used to improve both families of methods, while decreasing the entropy of the treated data.**

Keywords- ***Compression; entrop; dictionary; coding; compression ratio.***

I. Introduction

Since the beginning of researches about data compression, more than half century ago, most of methods make a transform on the treated data to increase order or repetitions; in entropy coding [1]; within the latter, than apply a coding method that will really make the compression process using this increases.

Generally, we try to make a prediction on the next block of data according to its type [2] to achieve the first part.

For coding, we will make our tests with the Huffman method [3], which is a statistical method based on the repetition of characters, and the LZW method based on the coding of a concatenation of characters to compress the resulted dictionary [7]. The transform we will apply is supervised since it performs a first pass to evaluate the correlation between each character and its following, so that each character often behind another one will be coded by a little coefficient, corresponding to its rank among the other successors, in every succession of both of them. Therefore, our method is not itself a compression method, but it will increase the number of little coefficients, which will improve the efficiency of the statistical coding methods. We will focus on lossless compression which means that the compressed element can be reconstructed without any loss of information.

Note that the latest compression methods often focus on one type of data. Our method, although it is less efficient in most of cases than algorithms specialized in a single data type, obtains good results on many types of data with this single algorithm. So our goal is not to compare the compression ratio obtained, but to demonstrate significant results on many kinds of data without making modifications to the algorithm and avoiding to stack many layers of processing.

II. Description of the Technique

A. Data Pre-Treatment

In this step we create the vector that will be coded by the transform. For this, a subdivision of the input data in blocs is done, followed by a reordering if necessary. The interest of this reordering is to optimize the regularity of the bloc by increasing the correlation between each element of input data and its following. The processing to be applied will vary according to the nature of the input data, so for text data Burrows & Wheeler transform [4] which sorts blocs of text lexicographically increasing rehearsals will be applied. For image inputs, the matrix will be divided and each block obtained will be sorted according to the flow geometry inside the block [5],Or simply by choosing among the horizontal, vertical and diagonal direction to minimize the entropy resulting. The same operation is applied to video on the images sequences. The goal of those pre-treatment is to minimize high frequencies (defined here by a large differential between a coefficient and its following) for images, sounds and videos, and to reduce the disorder of characters in textual data.

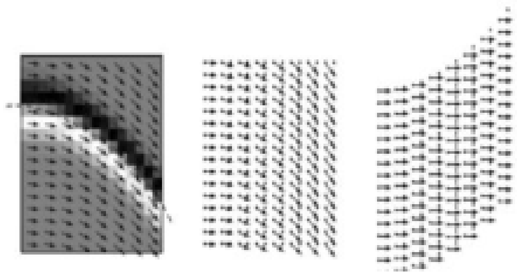

Figure 1. Example of resolution of the flow geometry in a piece of image