www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issues 6 June 2016, Page No. 16782-16784

Concurrent Data Compression for Cost Effective Applications

M. S. Narayanan
PG Professor, ECE Department Rajalakshmi Engineering College
Thandalam-602105 Tamilnadu
narayanan.ms@rajalakshmi.edu.in

Abstract

This work presents a novel idea of compression of multi- parameter data encountered in various fields. The advantages of a time domain approach are emphasised. A general mathematical model is proposed that can be adopted for any 3D object. The ease with which the method can be adopted for on line data capture in remote and rural areas is the key aspect of this work.

Keywords: Image processing, Data compression

Introduction

Image compression evolved over many decades and the application of transform methods particularly DCT [1] followed by WT [2] proved to be a game changer. The Image compression methods followed the path of hardware development. For example, MPEG 21[3] was possible due to the extraordinary developments in network components. However due to royalty fees, the transform based image processing software is very expensive and beyond the reach of ordinary people. On the other hand another technology has moved computing closer to the masses. Mobile phones have brought technology to the poor and rich alike. The applications of image processing are pervading all aspects of our life. It is interesting to note that a lot of researchers in the field of image processing are so excited about transform methods; they even tend to forget history. When the fax machines appeared in the market it was the royalty free differential and run length encoding algorithms [4] that made the products successful.

Currently there are many disciplines that depend heavily on image compression. The first and foremost discipline is medical informatics. The increased use of mobile phones is a real challenge. While medical data can be processed in the diagnostic centres due
to the availability of equipment, it is not the case with field workers and ordinary citizens who cannot afford the cost of transform based image processing software.

Most time domain image compression methods can be classified under lossless compression, which helps us to recover the original image from the compressed image. Let us look at the need for lossless and time domain compression methods. In order to understand and appreciate this requirement we have to take a closer look at the application areas. In medical informatics, a doctor or health worker at a remote location may have to depend solely on a mobile phone to process a medical image and also the *related data* [5]. Another example is the requirement of a technical specialist in a remote oil field, which may require the data of a reservoir model for his work. This person also may have to depend on a mobile phone.

Another interesting feature of present day technology is the additional data associated with an image. Until recently image data was looked in isolation. Now we have to look at integrated data for any application. Consider for example the requirements of a neurologist, who not only needs the MRI scan data, but also the clinical and lab data associated with the same *object*. Hence we need to look at a technique for concurrent compression of *all data* associated with an object. For this purpose we need to define a mathematical model so that a generalized method can be developed. This will enable us to ensure wider application.

Mathematical model

Image processing for scientific and technical applications span a very wide area. However, to a very large extent we deal with the geometry of the object as well as *multiple* parameters that vary across the *geometrical* space of the object. The extensive work on image processing, done in the past and present has its focus on one parameter, *pixel* value. Hence in the literature, mostly we are looking at 2D and 3D image data. Suppose we take a modest approach of looking at simple objects such as a tomato, there is no doubt that a transform based lossy compression will give an excellent compression ratio. But today a user needs much more than *mere image data*. For example, if we are looking at the research in farm products, various users need the *3D image and other data*. For example the chemical composition, fertilizer residue etc. may be needed along with pixel values. Suppose we look at objects that can be approximated by geometry, in general we can represent such objects in three dimensional geometrical spaces. So based on resolution, an infinitesimal mass within the object can be approximated as a point. For this mass the reflectance value is just one *attribute,* which we usually call pixel value. For the same mass there can be many attributes such as moisture content, chemical composition etc. We can assign a dimension to each attribute, so that we can *represent* all the attribute values at any point in the object in an N dimensional hyperspace. Hence for each sample point of the object in the geometrical space each parameter has a value. *Pixel value* is just one parameter. At the data generation point we may end up with a huge volume of data. When we look at the data, one parameter may show minor deviation in value across the entire geometrical space while another may show major variations. Our goal is to achieve optimal compression.

We assume the parameter time to be fixed. In general considering the Cartesian, cylindrical and polar coordinate systems we can represent the values of the parameters as

Cartesian: D_{xyz} (P) = d(x, y, z: p1, p_2, p_3, …………p_n) (1)

Cylindrical: $D_{\rho\phi z}$ (P) = d(ρ, ϕ, z: p_1, p_2, p_3, ……………………p_n) (2)

Polar: $D_{r\theta\phi}$ (P) = d(r, θ, ϕ: p_1, p_2, p_3,…………p_n) (3)

Where, D (P) indicates the data of all parameters at that point. The basic nature of the data corresponding to one parameter may vary from another. For example p_1 represent pixel value, a scalar, while p_2 may represent field intensity, a vector.

Another parameter may represent chemical composition that may have sub-data elements. Most parameters may have high degree of correlation in a volume of mass in geometrical space. Consider the data obtained from two points adjacent to each other represented by D_{x1y1z1} (P), D_{x2y2z2} (P). Hence by applying differential encoding we register the data D_{x1y1z1} (P), and for the next one D_{x2y2z2}(P), the difference D_{x1y1z1}(P) - D_{x2y2z2}(P) is registered. Most of the parameter values will be zero when the correlation is high. It is unlikely that all parameters will vary in the immediate neighbourhood throughout the geometrical space.

If the geometry of the object under investigation can be approximated to a three dimensional object that can be represented by a mathematical equation, then if there is homogeneity for a given parameter, for all data points the *values of data elements* are identical. In such a case a high degree os compression is achieved. Moreover, since the object can be represented by a mathematical equation, the memory wasted otherwise for all sample points are replaced by small number of bytes needed to store the equation parameters. This is particularly helpful when using portable devices for field studies and on site measurements.

We have certain objects that can be approximated by the revolution of a two dimensional curve about an axis. For example certain fruits and vegetables can be approximated by the revolution of a cardioid about a vertical axis. In such cases the object can be approximated by the sections obtained about the vertical axis. This approach is eminently suitable for CT scans. We have then,

$D(P) \approx \sum D_n(P_{sn})$ (4)

Where, P_{sn} denotes the parameter values of section n.

Real world objects are far from objects defined by geometry. But experience tells us that a large portion of a real world object may follow a known geometrical shape with some deviations. For example some fruits are approximately spherical. The data for such a fruit can be represented as

$$D(P) = F[D_1(P), D_2(P)] \qquad (5)$$

For example in `the case of an object that is approximately spherical, but having a bulge near the areas close to the equatorial plane D_1 can be sample points of a perfect sphere while D_2 represents the sample points of the g*eometrical* shape that represents the deviations.

Data Representation

The multi-parameter data should be represented in a manner suitable for data and mathematical manipulation. One common parameter is the reflected or transmitted irradiance in visible or other spectrum normally specified as pixel value. This parameter has at least two sub-parameters *brightness* and *colour* the other parameters are more complicated. A physical object, under investigation yields multi-parameter data some parameters are measured at a point, while others in a volume. This is due to the fact that some parameters need finite volume or mass for measurement. Hence we need to define a framework for integration of multi-parameter data. A physical object has three dimensions consisting of infinite points. The sampling points are finite. Hence all measured data are finite. The physical object has a finite volume. For example all medical investigations on a human is confined to the entire volume of the patient. Similarly a part of the body under investigation represents the volume that includes all sampling points. Each parameter measurement uses a sample volume according to the *resolution.* The parameter with the highest resolution will generate higher quantity of raw data. We have to think in terms of compressing data corresponding to all parameters as soon as it is generated. If the parameters are measured at different times at different locations, we need to process them with a holistic approach so that compression and integration can be carried out easily. In some cases small deviation in one or more *parts* of the object may be the key factor.

Conclusions

Concurrent data compression helps us to save time and money. This method offers high level of compression. Using multiresolution images it is possible to adopt different levels of compression. The simplified approach of differential encoding enables on site data capture and processing using simple hardware. The most attractive aspect of this scheme is the use of this method in remote areas in online or offline mode. The author hopes that this paper will stimulate extensions of this work from eminent research workers all over the world. Since the work presents a general framework, the scope for implementation styles is enormous.

References

[1] http://www.lokminglui.com/dct.pdf

[2] http://cdn.intechopen.com/pdfs/34963.pdf

[3] http://mpeg.chiariglione.org/standards/ Mpeg-21

[4] Frank J Merkl- Binary Image Compression using Run Length Encoding and Multiple scanning Techniques. [Thesis], Thesis/Dissertation Collections, Rochester Institute of Technnology New York 1988

[5] M. S. Narayanan, N. Gokul – Computer and Information Technology Initiatives for National Development, Proceedings of NCNAT-2015 Navi Mumbai – 22nd August 2015 Pp 29-32

M. S. Narayanan earned his B. E. degree in electronics and communication engineering from National Institute of Engineering (Mysore University) in 1975. He got his M.Tech .degree from IIT Bombay in 1977. He has 39 years of professional experience equally spanned between the academia and the industry. Presently he is a PG professor in the ECE department at Rajalakshmi Engineering College. His research interests include Image Processing, Data Compression, Information Theory, Rural Development, E-Governance and energy efficient portable devices. He is a senior member of IEEE(USA), professional member of ACM(USA) and a life member of ISTE (India) .