Information Sciences 381 (2017) 33–45

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Lossless compression for aurora spectral images using fast online bi-dimensional decorrelation method

Wanqiu Kong[a], Jiaji Wu[a,*], Zejun Hu[b], Marco Anisetti[c], Ernesto Damiani[c,d], Gwanggil Jeon[a,e,*]

[a] *School of Electronic Engineering, Xidian University, Xi'an, Shaanxi, China*
[b] *SOA Key Laboratory for Polar Science, Polar Research Institute of China, Shanghai, China*
[c] *Dipartimento di Informatica, Università degli Studi di Milano, Via Bramante 65, 26013 Crema (CR), Italy*
[d] *Department of Electrical and Computer Engineering, Khalifa University, P.O.Box 127788, Abu Dhabi, UAE*
[e] *Department of Embedded Systems Engineering, Incheon National University, Incheon 22012, Korea*

ARTICLE INFO

Article history:
Received 26 April 2016
Revised 15 November 2016
Accepted 16 November 2016
Available online 19 November 2016

Keywords:
Lossless compression
Aurora spectral images
Online
Bi-dimensional decorrelation
Parallel

ABSTRACT

In this paper, we propose a lossless compression method to resolve the limitations in the real-time transmission of aurora spectral images. This method bi-dimensionally decorrelates the spatial and spectral domains and effectively removes side information of recursively computed coefficients to achieve high quality rapid compression. Experiments on data sets captured from the Antarctic Zhongshan Station show that the proposed algorithm can meet real-time requirements by using parallel processing to achieve outstanding compression ratio performance with low computational complexity.

1. Introduction

1.1. Importance of lossless compression for aurora spectral images

When electron plasma is blown by solar wind and enters into the Earth's atmosphere, a large number of neutral particles are excited which causes a significant event known as the aurora phenomenon. To record its visual morphology and spectral information, a specialized type of spectrometer captures the instant aurora spectra as electronic spectral images. Such aurora spectral images are helpful in the investigation of the conditions and environment in which this phenomenon occurs and the actual interaction between high-energy particles and the Earth's atmosphere. The electron energy and flux, species concentrations and electron temperature in the auroral ionosphere, and the distribution of deposited particles can be derived. Because spectroscopic emission features have a relationship with earth-solar activity, this suggests a response by Earth to the sun. Thus, there is an abundance of information to be investigated.

Since the exact arrival time of the solar wind is unknown and the aurora rapidly changes and moves due to the Earth's geomagnetic activity, a spectrometer is required to produce continuous images at a relatively high frequency until the aurora appears. As such, the offline storage cost and transmission load is substantial. For example, the auroral spectrometer at the

* Corresponding authors.
E-mail addresses: wanqiukong@stu.xidian.edu.cn (W. Kong), wujj@mail.xidian.edu.cn (J. Wu), huzejun@pric.org.cn (Z. Hu), marco.anisetti@unimi.it (M. Anisetti), ernesto.damiani@kustar.ac.ae (E. Damiani), ggjeon@gmail.com (G. Jeon).

http://dx.doi.org/10.1016/j.ins.2016.11.008

Antarctic Zhongshan Station[1] produces an aurora spectral image every 20 s. Each image is stored in a 16-bit Flexible Image Transport System (FITS) format [1] with 1024 × 1024 in size. Although data from all the Chinese stations at the Antarctic Pole are transferred via satellite, these stations all share the same communication channel, which has limited bandwidth. To solve this problem, lossless compression must be utilized to compress the aurora spectral images prior to their real-time transmission.

Early work in image compression used lossy compression algorithms which conceded a certain loss in the visual observation data to obtain a decreased file size. However, for some special images that require high resolution and high accuracy, this loss is noticeable. In response to this problem, lossless algorithms were developed. Due to the fact that aurora spectral images contain rich information for research, lossless compression is preferred for its completely reversible image reconstruction. However, due to increasing amounts of data, real-time transmission of compressed data is necessary even if the compression scheme only partially alleviates the local storage problem.

1.2. Overview of lossless compression algorithms

All lossless compression algorithms have been developed on the bases of Shannon's information entropy theory and source coding theorem [2], which identifies the theoretical lower bound when minimizing the expected length of coded symbols. Minimization is typically performed in two steps: decorrelation and entropy coding. In the compression process, the original image is reconstructed with a specified form, and entropy coding is the key aspect of reversible reconstruction. Decorrelation techniques remove redundancy as much as possible to improve compression performance. The prediction-, transform-, and dictionary-based methods are the most representative examples.

In prediction-based methods, each sample value to be predicted is a weighted linear combination of several adjacent sample values, and weighted coefficients are fixed or computed using a specific optimization strategy. The use of these methods is widespread since they are effective for decorrelating samples. Actually, linear prediction is usually considered to be differential pulse code modulation (DPCM), which was first developed to enhance audio quality based on the principle of pulse coding modulation. Later, it was applied to image processing. The derived algorithms minimize the sum of the squared error between the predicted and original values, and the error, called the residual, must be entropy encoded. JPEG–LS is a typical lossless compression standard for still images. As the core method of JPEG–LS, the low complexity lossless compression algorithm for images (LOCO-I) [3] utilizes a median edge-detection predictor with relatively low computation complexity. Context-based, adaptive, lossless image coding (CALIC) [4] involves image data modeling and provides continuous-tone and binary modes to exploit the image properties of all types. It uses various contexts to initialize a nonlinear predictor and adaptively updates it according to the source statistics. Prediction by partial matching (PPM) [5] also employs a context model, and several Markov models of different prediction orders are maintained. A modified algorithm [6] uses the *escape/switch* symbol in an extended alphabet, and the alphabet can be applied to any algorithm in the PPM family.

In lossy compression, transform coding has demonstrated excellent performance, but most transforms are not directly applied to lossless compression systems. For example, the inverse discrete cosine transform (DCT) and inverse discrete Fourier transform (DFT) both lead to certain accuracy loss in transformation since they use real and complex numbers as transform coefficients, respectively. In this case, reversible transforms are preferable. The widespread lossless image format JPEG2000 [7] utilizes the 5/3 wavelet [8] to remove high-frequency information. Similarly, a lossless implementation of set partitioning in hierarchical trees (SPIHT) [9] uses the same wavelet in embedded zero tree wavelet encoding [10]. In addition, the wavelet transform is replaced by a reversible S+P transform [11]in a variant lossless SPIHT. Specifically, there are some transforms that are not very effective for decorrelation but help to increase compression efficiency in entropy encoding. For instance, the Burrows–Wheeler transform (BWT) [12] can maximize the number at which identical symbols continuously appear, which facilitates the compression of a string that, by other methods such as run-length encoding (RLE), has multiples of repeated symbols [13]. The well-known software for file compression Bzip2 simply combines BWT with RLE.

The family of dictionary-based coding algorithms maintains a data structure called the dictionary, which stores a series of strings that are searched for the matching string of the text to be compressed. LZ77 [14] and LZ78 [15] are two original dictionary coders. Their dictionaries consist of strings matched to other strings to be encoded and they are usually dynamically updated with the encoded data. The DEFLATE algorithm [16] was subsequently developed, which combines LZ77 with Huffman encoding [17] and is utilized for the Portable Network Graphic (PNG) format or the Multiple-image Network Graphics (MNG) image format. LZWelch (LZW) [18] improves LZ78 in terms of computation efficiency while maintaining a comparable compression performance and is applied in the Graphics Interchange Format (GIF) or Tagged Image File Format (TIFF) image formats based on LZ77.

Generally used in the last processing step, entropy encoders compress both the residuals and side information. Based on the characteristics of the random process, there is some statistical redundancy between symbols. The symbol codes to be encoded are obtained according to each symbol's occurrence frequency. If the shortest codes are assigned to the most common symbols and symbols that rarely appear are assigned the longest codes, the total length of all the codes will decrease. The above-mentioned Huffman encoding [17] uses this variable-length coding scheme to achieve a minimal

[1] Located at 76° 23 min of east longitude and 69° 22 min of south latitude.