

A Memory-Assisted Lossless Compression using Combination Methods

Uday Joshi
Associate Professor
University of Mumbai
K.J. Somaiya College of
Engineering

Bhakti Palkar
Associate Professor
University of Mumbai
K.J. Somaiya College of
Engineering

Khushi Khanchandani
PG student,
University of Mumbai
K.J. Somaiya College of
Engineering

ABSTRACT

Image compression can be done in two ways namely lossy or lossless. Lossy methods are used for expected results like photographs where some negligible loss of data is tolerable. Rapid growth of medical science such as ehealth and telemedicine requires lossless Image compression. The correlation and redundancy which exists across different medical images are considered to achieve better compression ratio. . Better compression results in less storage and less overhead while transmission on the network. A memory-assisted hybrid compression technique is proposed. The approach is motivated by using a combination of methods, lossy and lossless. The original image is compressed using PCA algorithm and then using lossless coding methods. PCA is used to find the correlation among the similar medical images. PCA algorithm is used with only some principal components.

General Terms

Lossless Image Compression.

Keywords

Lossy and Lossless Compression, PCA, Memory Assisted Compression

1. INTRODUCTION

Many hospitals have satellite centers where they make use of telecardiology application. In telecardiology applications digital images are saved and preserved into large databases. An advance in technology has given the medical science to use appropriate compression algorithm to reduce the size of the image and transmit the images on the network. The intention is to reduce the size of original image and so speed up the transmission speed for images on network and save the memory space required to store image dataset. In telecardiology these images take high bandwidth capacity to transmit. So compression is necessity for handling large image dataset.

In computer base diagnosis, loss of any part of the information can be harmful. Therefore, lossless image compression is the technique for medical images. Literature based on lossless compression for medical images considers compression algorithms within an individual image so as to achieve better compression performance. More data can be compressed using inter image redundancies.

The motive is to explore the commonalities among the same family of medical images. Algorithm is implemented which will retrieve the any random image without decompressing the entire database. Better compression can be achieved using memory assisted compression technique. A memory assisted compression can be framed in to two scenarios. In the first scenario the traditional lossless compression is applied on the test images. In the second scenario the lossy compression along with lossless compression technique is applied. In the literature LZW, arithmetic coding, RLE and Huffman lossless techniques and PCA algorithm as lossy compression technique.

In telemedicine applications, where medical images are taken every day and transmitted over the network, the images from the previous day, for example, can serve as the memory for the next day, and so on. Our focus here is mostly on the second phase, i.e., the memory-assisted compression. In the second phase, for compression of every image in the database, the common information stored in the memory is used to eliminate the inter-image redundancy and only the residual is fed to traditional lossless compression algorithms. The proposed two-phase structure enables individual access to all the images without the need to decompress the whole database and at the same time, all the dependencies present among the images are used for efficient compression.

2. BACKGROUND AND RELATED WORK

As previously described, the use of correlation information across images has been very limited in lossless image compression. Most of the previous work only focuses on the compression of a single image regardless of the redundancy that is present in the set of images to which this specific image belongs. In this section, the review of some published research on lossless image compression based on individual medical images is done. MacMahon et al.[2] proposed a form of adaptive block Cosine Transform coding, in which considerable compression of digital radiographs with minimal degradation of image quality is allowed. Their results obtained for chest radiological images showed a compression ratio as high as 25. Ekstrand [3] presented a lossless compression algorithm based on Context Tree Weighting (CTW). The algorithm performs optimally in terms of redundancy for a wide range of data sources including medical gray scale images. The results show enhanced performance compared to JPEG, JBIG, and

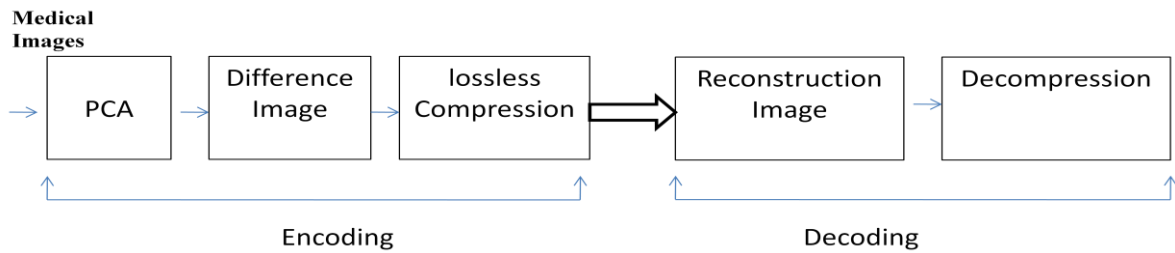


Fig1 The proposed memory-assisted compression algorithm

CALIC. Asraf et al.[4] proposed a novel hybrid lossy and lossless compression method using neural networks, vector quantization, and Huffman coding. The method was tested on CT (Computerized Tomography) images achieving a compression ratio of 5 to 10. A lossless medical image compression method was presented by Kil et al.[5]. The method was based on redundancy analysis and segmentation of image into Variable Block Size (VBS) in order to extract similarities and smoothness of blocks. It was reported that the technique out-performs Huffman, JPEG-LL and lossless JPEG2000 by 10-40%. Ghare et al.[6] introduced a lossless image coding algorithm based on pixel redundancy reduction and using 2 matrices of Gray Scale and Binary. The algorithm achieved a maximum compression ratio of 4. Miaou et al.[7] proposed an image compression technique which combines JPEG-LS and inter frame coding with motion vectors showing a better than JPEG-LS alone. They tested their algorithm with six capsule endoscope image sequences and improved the average compression by 13.3% and 26.3% over JPEG-LS and JPEG-2000, respectively. Our focus is on removing inter-image redundancy using memory-assisted compression. Again, the proposed algorithm can be used as an add-on to any existing image compression technique. In experiments, we work with LZW, arithmetic coding, RLE and Huffman lossless technique

3. PROBLEM AND PROPOSED APPROACH

The main rationale behind memory-assisted compression is learning source statistics at some intermediate entities, and then leveraging the memorized context to reduce redundancy of the universal compression of finite length sequences. To the best of our knowledge, this is the first attempt in using cross-image correlation in medical image compression. Indeed, the concept of memorization is applied with medical image sequences. The basic problem setup in a telemedicine application is displayed in Fig. 1. The source contains a set of correlated medical images (e.g. Chest X-ray images) at the server node S(e.g., the central hospital) that need to be encoded and transmitted to the destination node (e.g. remote hospital) D through the network

The assumption is made that hospital D has already memorized a database of previously transmitted images in its database which is also shared with S. In the absence of memorization, traditional compression techniques can still be applied for transmitting the sequence of images from S to D. Here, the communication overhead to send the images from S to D can substantially be reduced if memorized context is available to the encoder and the decoder. In this paper two different scenarios to compare results of traditional lossless image compression algorithms with proposed memory-assisted compression algorithms.

First, traditional state-of-the-art compression algorithms are applied on a set of medical images. In this scenario,

redundancy is considered only within a single image for encoding and decoding it. The problem is that every single image is encoded without considering other images in the same set resulting in a low compression ratio and additional overhead. Next, memory-assisted compression algorithm using the simple PCA algorithm [8] within the same set of images. In this case, the encoder and decoder have access to the memorized context for compression of new unknown images. A significant improvement in compression ratio for lossless medical images over state-of-the-art algorithms can be achieved.

4. THE PROPOSED MEMORY-ASSISTED COMPRESSION ALGORITHM

In this section, the proposed algorithm is described in detail in further details. Consider a basic scenario in which a set of X-ray images is available at node S, and node D requests one of the images, as shown in Fig. 1. This scenario can be an abstraction of a transmission problem or a storage and retrieval problem. Our benchmark is the case in which each image is compressed individually and sent to D. Then, at node D, each compressed image is decompressed independently. In the proposed method, the outcome of the learning phase, called M, of memory-assisted compression is available at both S and D. Then, using M, just the residuals of other test images are encoded at node S and decoded at node D. The memory-assisted compression method[11] consists of two main phases:

- 1) Learning (memorization)
- 2) Memory-assisted Compression (testing).

The main question now is how to consider and model the memorization concept from a set of gray-scale medical images? The simple answer comes from the Karhunen Lo-ève transform (KLT) [9]. KLT is shown to be the optimal orthogonal transform through which the energy (information) contained in the signal is compacted. Indeed, with KLT, most energy is redistributed over a small number of components or simply called eigen images. These eigen images are obtained from the decomposition of the estimated covariance matrix. For our experiments, the PCA transformation matrix is used to decorrelate the images and remove inter-image redundancy. PCA is a statistical approach used to find an orthogonal transformation to decorrelate random variables. The PCA technique has extensively been used in diverse signal and image processing applications. It has originally been introduced as a dimension reduction technique. The technique starts with a set of observation vectors of dimension N. For images, the columns are concatenated into a large vector of size N (number of pixels). Let M be the number of observations in the training set:

$$X_i = [p_1, \dots, p_N]^T, I = 1, \dots, M \quad \dots(1)$$

From these observations, the mean vector and covariance matrix are estimated:

$$m = 1/M * \sum_{i=1}^M (x_i) \quad \dots (2)$$

$$C = 1/M * \sum_{i=1}^M ((x_i - m)(x_i - m)^T) \quad \dots (3)$$

Let's represent the mean centered observations by: $w_i = x_i - m$, $i=1, 2, \dots, M$. The goal is to find a subspace whose basis vectors correspond to the maximum-variance directions in the orthogonal space. Let W represent this linear transformation that maps the original N -dimensional space onto a K -dimensional feature subspace where normally $K \ll N$. The columns of W are the eigen vectors, e_i , of the covariance matrix C .

For a given observation vector x_i , the transformation results in a new vector y_i given by: $y_i = W^T (x_i - m)$. As the first few eigen values represent most energy in the data, usually select K to be much smaller than N . The original observation vectors can then be reconstructed using the inverse transform. $X_i = W^T y_i + m$. Note that since C is an $N \times N$ matrix where N is the total number of pixels in the image, finding the eigen vectors may not be easy to do. Luckily, a number of techniques have been proposed to get around this difficulty. For our experimental setup, the PCA procedure discussed above was first applied to the training set of images. Once the training stage is completed, move to the coding stage. In this stage, test images are first projected over the PCA space. Using the reduced PCA space, the test images are re-constructed. These reconstructed images are close approximation of the original images. An error image is obtained by subtracting the reconstruction image from the original test image. It is simple to show that the pixel values of such an error image are uniformly distributed. As such, the error images can be compressed in a more optimized way using the traditional LZW, arithmetic coding, RLE and Huffman lossless techniques [12]. This means that only the feature vectors as well as the new compressed error images are required to send to the receiver. At the receiver, first reconstruct the image projection then add to it the decompressed error image. To evaluate the performance of the proposed approach, consider two scenarios and 4 generic compression techniques. The compression techniques considered include: LZW, arithmetic coding, RLE and Huffman lossless techniques. The two scenarios are explained below:

- Scenario 1(Comp): It denote the case of using the LZW, arithmetic coding, RLE and Huffman lossless techniques directly on the test set.

- Scenario 2(PCA Comp): Here, PCA is applied on a train set of images. Then, for testing, the images are first projected and reconstructed using the PCA. Second, the residual images are encoded using the LZW, arithmetic coding, RLE and Huffman lossless techniques. At the receiver, the images are reconstructed using the decoded residuals added to the PCA-reconstructed images.

5. SIMULATION RESULTS

In this section, experimental results are discussed in detail. For lossy compression, the RMSE can be a measure of performance, but for lossless coding, compression efficiency is usually measured using compression ratio. Computational complexity is another factor that determines the efficiency of the method. This can be the number of CPU cycles, number of hardware gates, or memory bandwidth, etc. These are usually application dependent. Here the compression ratio is defined as the compressed image size S_{comp} over the original image size s_o , i.e. $CR = S_{comp} / s_o$. So In other words, the CR represents

the number of bits the scheme uses to represent each bit in the uncompressed image. The image database used is the JRST database which can be downloaded from [10]. It contains 154 nodule and non-nodule 8-bit Chest X-ray images with matrix size of 2048 x 2048 pixels. A subset containing 20 Chest X-ray images are selected. 10 images are selected as our training set while the other 10 images were used for testing the algorithm. The experiments were repeated 10 times by randomly changing the 20 images. The LZW, arithmetic coding, RLE and Huffman lossless techniques were applied on the test set as our benchmark lossless image compression algorithms entropy for the original image (Fig. 2) was 7.55 and decreased to 5.54 for the image after PCA decorrelation. The histogram of the first eigen image (Fig.4), which clearly proves the maximum variance that such an image can model.

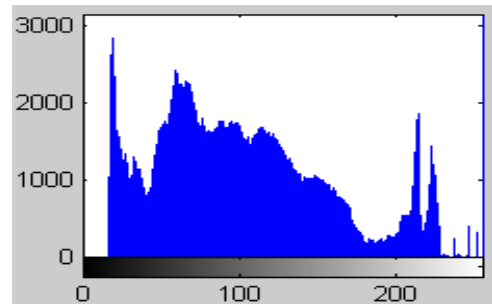


Fig 2. Histogram of a single Image

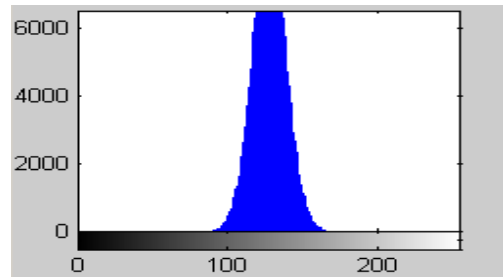


Fig 3. Histogram of a single Image After PCA

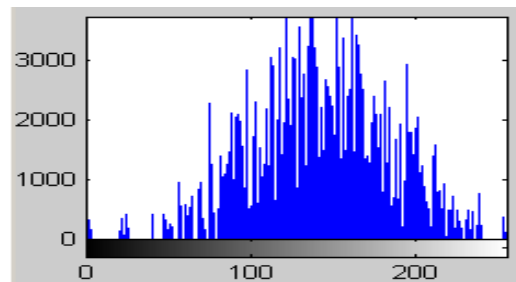


Fig 4. Histogram of a single Eigen Image

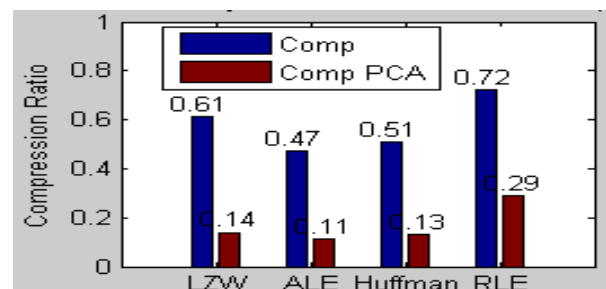


Fig 5. Compression Ratio For Different Lossless Image Compression Algorithms Using Both The Comp And The PCA Comp Scenarios.

Fig. 5 displays compression ratio for different lossless image compression algorithms using both the Comp and the PCA Comp scenarios. The compression ratios (CR) for LZW, arithmetic coding, Huffman and RLE were 0.14, 0.11, 0.13, 0.29 respectively. After applying our memory-assisted algorithm, the compression ratio was improved significantly. The ALE, in particular, achieved a slightly better CR than the others. Experiments have been performed on different sets of training images with various sizes. As can be seen, there is only a minor improvement in compression ratio when there is increase in the size of the training set (number of images) from 10 to 30 images. The highest compression efficiency is achieved for the ALE and the Huffman algorithms when used with proposed PCA Comp algorithm. By applying ALE and the Huffman algorithms, obtained gains are 76% and 75% respectively. These reported gains are for training sets of 10 images. The same trend can be observed for the image set of 20 and 30. As expected, few PCA components are indeed important in the reconstruction. Actually, our experiments showed that more than 97% of the total energy is contained in the first 5 eigen values.

6. CONCLUSION

The discussion is based on a new method for lossless compression using the concept of memory-assisted universal coding. The proposed approach is well suited to compress large datasets of medical images especially for recurrent usage. The compression ratio decreases with increase in the number of test images (i.e. better). This happens because the overhead starts paying off after a certain images are sent. Subsequently the compression ratio goes on reducing (i.e. gets better) with the increase in the no. of test images ultimately converging to a fixed value like 0.14 in case of PCs = 5. After compression, the compression ratio is always less than one. The compression ratio is minimum in case of no. of PCs = 5 (optimum value). This is obviously because the overhead of transmitting PCs increases. More than 97% of energy is contained in the first 5 eigen values[11]. The algorithm consists of a learning phase followed by a testing phase. In the learning phase, PCA is performed on training images to extract a set of eigen images which are used to reconstruct the different test images. The reconstructed images are simply represented (coded) by low dimensional feature vectors. The error (or residual) images are then compressed using traditional lossless compression algorithms such as the LZW, arithmetic coding, and Huffman and RLE algorithms. Our experimental results using the JRST database showed that the performance of traditional lossless algorithms can be improved by an average of 65% using the proposed algorithm. The proposed concept of using memory to enhance the performance of universal coders is expected to have a major impact in areas where images exhibit high correlation.

7. ACKNOWLEDGMENT

Khushi Khanchandani wish to express sincere thanks and deep sense of gratitude to respected mentor and guide Mrs. Bhakti Raul, Associate Professor and Mr. Uday Joshi, Associate Professor also Department Computer Engineering, K J Somaiya College of Engineering, for his in depth and enlightening support with all of his kindness.

Apart from technical advice I have got lots of encouragements and constructive criticism, which motivates to strive harder for excellence.

8. REFERENCES

- [1] A. Beirami, M. Sardari, and F. Fekri, "Results on the fundamental gain of memory-assisted universal source coding," in Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on, July 2012, pp. 1087 – 1091.
- [2] H. MacMahon, K. Doi, S. Sanada, S. Montner, M. Giger, C. Metz, N. Nakamori, F. Yin, X. Xu, H. Yonekawa, and et. al., "Data compression: Effect on diagnostic accuracy in digital chest radiography," vol. 178, no. 1, pp. 175–9, Jan 1991.
- [3] N. Ekstrand, "Lossless compression of grayscale images via context tree weighting," in Data Compression Conference, 1996. DCC '96. Proceedings, Mar/Apr 1996, pp. 132-139.
- [4] R. Asraf, M. Akbar, and N. Jafri, "Statistical analysis of difference image for absolutely lossless compression of medical images," in Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE, 30 2006-sept.3 2006, pp. 4767 –4770
- [5] S. Kil, J. Lee, D. Shen, J. Ryu, E. Lee, H. Min, and S. Hong, "Lossless medical image compression using redundancy analysis," IJCSNS International Journal of Computer Science and Network Security, vol. 6, no. 1, pp. 50–56, Jan 2006.
- [6] S. Ghahre, M. M. Ali, K. Jumari, and M. Ismail, "An efficient low complexity lossless coding algorithm for medical images," vol. 6, no. 8, pp. 1502–1508, 2009.
- [7] S.-G. Miaou, F.-S. Ke, and S.-C. Chen, "A lossless compression method for medical image sequences using jpeg-ls and inter frame coding," Information Technology in Biomedicine, IEEE Transactions on, vol. 13, no. 5, pp. 818 –821, Sept. 2009.
- [8] I. Jolliffe, "Principal component analysis," Springer, 1986.
- [9] J. Burl, "Estimating the basis functions of the karhunen-loeve transform," Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 37, no. 1, pp. 99 –105, Jan 1989.
- [10] [Online]. Available: www.jsrt.or.jp
- [11] Zhinoos Razavi Hesabi, Mohsen Sardari, Ahmad Beirami, Faramarz Fekri, Mohamed Deriche, and Antonio Navarro, "A Memory Assisted Lossless Compression Algorithm For Medical Images", in Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, May 2014, pp. 2030 – 2034
- [12] Anmol Jyot Maan, "Analysis and Comparison of Algorithms for Lossless Data Compression", International Journal of Information and Computation Technology. ISSN 0974-2239 Volume 3, Number 3 (2013), pp. 139-146