A MULTILEVEL MEMORY-ASSISTED LOSSLESS COMPRESSION ALGORITHM FOR MEDICAL IMAGES

Zhinoos Razavi Hesabi, Borhan Kazimipour*, Mohamed Deriche†, Antonio Navarro‡*

* School of Computer Science and IT, RMIT University, Melbourne, Australia
† Department of Electrical Engineering, King Fahd University of Petroleum & Minerals, Saudi Arabia
‡Institute of Telecommunication- University of Aveiro, Aveiro, Portugal
zhinoos.razavi@rmit.edu.au, mderiche@kfupm.edu.sa

ABSTRACT

As medical imaging facilities move towards film-less imaging technology, robust image compression systems are starting to play a key role. Conventional storage and transmission of large-scale raw medical image datasets can be very expensive and time-consuming. Recently, we proposed a memory-assisted lossless image compression algorithm based on Principal Component Analysis(PCA). In this paper, we further improve the performance of the algorithm in two different directions: Firstly, we replace PCA with NMF (Non Negative Matrix Factorization). NMF has several advantages in representing images with an image-like basis, results in sparse factors, and provides better user control over iterations. Secondly, we expand the single-level model with a new multilevel decomposition/projection framework to further reduce entropy of residual images. Our experimental results on X-ray images confirm that both modifications provide significant improvements over the single level PCA based algorithm as well as existing non-memory based techniques.

Index Terms— Lossless Compression, Medical Imaging, Non-negative Matrix Factorization, Unsupervised Learning.

1. INTRODUCTION

Every day, a huge volume of digital images is produced by medical applications such as telemedicine and tele-radiology. There are many challenges associated with the storage, retrieval and transmission of this amount of data due to the practical limitations in communication bandwidth and constraints on time and space. To tackle these issues, many compression techniques have been proposed [1]. These techniques aim at effectively reducing the cost of storage and enhancing the speed of transmission. Since lossy compression techniques may adversely affect the accuracy of diagnosis systems, robust lossless compression algorithms are highly sought in different medical imaging applications [2].

Most of the literature on medical image compression focuses on reducing redundancy within a single image [3]. In contrary, only a limited body of research considered the possibility of extracting commonalities (similarities and correlation) among a set of images [4]. Indeed, some studies have shown that extracting cross-image redundancy can significantly improve performance of traditional lossless compression techniques [5].

The compression algorithms that exploit correlation within a set of similar images can be categorized into two categories: Set Redundancy Compression (SRC) algorithms and Memory-Assisted Compression (MAC) techniques. The main difference between these categories is that SRC algorithms extract "interimage redundancy" [6] from a set of images, then compress the residues from the same set of images. MAC techniques, in contrast, learn the common pattern (a.k.a. prototype or model images) from the training set, then use these to compress unseen images from the testing set. Therefore, the ability to memorize the commonalities and exploit these to compress unseen images is a unique feature of MAC techniques [3].

In this paper, we extend our previous MAC framework [3] by improving its template extraction algorithm and adding an extra learning level. More specifically, our main contributions in this paper are as follows: Firstly, we substitute Principal Component Analysis (PCA) by Non-negative Matrix Factorization (NMF). Although there are more constraints on NMF, the algorithm captures better the similarities and hence enhances compression rate. Secondly, we extend our previous MAC framework to a multilevel framework. With the expense of a linear increase in computational complexity, the multilevel framework outperforms substantially its single level counterpart. Finally, we study the sensitivity of the proposed technique to the sizes of the training and testing datasets.

The rest of the paper is organized as follows. Section 2 introduces the basis extraction algorithm using NMF. Section 3 discusses the proposed multilevel framework. Then, Section 4 discusses our experimental results. Finally, Section 5 draws some concluding remarks.

2. LEARNING SIMILARITIES USING NMF

In our previous MAC framework [3], we used a simple Principal Component Analysis (PCA) algorithm [7] for determin-

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

ing the projection basis. Technically, PCA transforms the data from the original coordinate system into a new one such that the first coordinate (a.k.a. first component) contains the largest variance of the data, the second coordinate contains the second highest variation, etc.... [7]

Despite its simplicity and effectiveness, PCA application is sometimes limited by a number of assumptions, including the linearity in the combination of basis vectors, importance of directions with the largest variance, and orthogonality of principal components. In addition, the eigenvalue decomposition of the covariance (or correlation) matrix for large size datasets (specially with large size images) is usually a tedious task. Moreover, the performance of the decomposition may drop when the images are noisy or sparse (as is in different biomedical imaging modalities).

To lessen some of the aforementioned difficulties, we propose a new MAC framework in which Non-negative Matrix Factorization (NMF) [4] is used for the template extraction phase. Technically, NMF is an unsupervised dimensionality reduction technique which identifies a set of non-negative components of an object and converts a data matrix into the product of two smaller matrices. A bold advantage of NMF over PCA is that it can be simply implemented using iterative algorithms. This means a balance (or a compromise) between accuracy of the approximation and speed of the algorithm can easily be achieved. Another advantage that makes NMF more applicable to our scenarios is the fact that NMF can be seen as parts based representation. Unlike PCA, LDA, and ICA, NMF representation is based on a positive-based combination of the basis images which can be seen as "true" template images.

In its simplest form, NMF works as follows: Assuming that an image database is represented by matrix $\mathbf{V}_{n\times m}$, where each column is a vectorised image containing n non-negative elements (i.e., pixel values), and m is the number of images in the set. NMF factorizes $\mathbf{V}$ into two matrices $\mathbf{W}_{n\times r}$ and $\mathbf{H}_{r\times m}$, where r is usually smaller than both n and m. More formally:

$$\mathbf{V}_{ij} \approx (\mathbf{WH})_{ij} = \sum_{k=1}^{r} \mathbf{W}_{ik}\mathbf{H}_{kj}, \text{ subject to } \mathbf{W}, \mathbf{H} \geq 0 \quad (1)$$

where the columns of $\mathbf{W}$ are the basis images of size n, and each column of $\mathbf{H}$ is a coefficient vector representing one of the m images.

To calculate the basis and coefficient matrices, both $\mathbf{W}$ and $\mathbf{H}$ are traditionally initialized by random positive numbers. Then, their elements are iteratively fine-tuned according to the following assignments:

$$\mathbf{W}_{ik} \leftarrow \mathbf{W}_{ik} \frac{(\mathbf{VH}^T)_{ik}}{(\mathbf{VHH}^T)_{ik}} \quad (2)$$

$$\mathbf{H}_{kj} \leftarrow \mathbf{H}_{kj} \frac{(\mathbf{W}^T\mathbf{V})_{kj}}{(\mathbf{W}^T\mathbf{WH})_{kj}} \quad (3)$$

Note that any column of $\mathbf{W}$ can serve as a basis image. In this study, we only use the first column as the template image reflecting the similarities of all images in the training set. Actually, similarly to PCA and LDA, the energy contained in the first basis image represents a large percent of the total energy. Assuming that the images of training and testing sets are strongly similar,as Figures 1.a and 1.b, we expect that this first basis image can also capture most of the redundancy in the testing set (see Figures 1.e and 1.f). The next section describes how this extracted redundancy can be used with the proposed multilevel image compression framework.

3. MULTILEVEL MEMORY-ASSISTED COMPRESSION

This section discusses the new framework which introduces the concept of multilevel feature extraction to the MAC framework. Before explaining the details of the proposed multilevel framework, let us briefly review the existing (i.e., single-level) MAC algorithm, originally proposed in [3].

In a basic sender/receiver scenario, we assume that a set of "similar"images $\mathcal{T} = \{\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3, \cdots, \mathbf{T}_{|\mathcal{T}|}\}$ is available at the sender side. This set of images can be used as the training set. The similarities within the training set are captured using an eigenvalue eigenvector decomposition of the estimated covariance matrix. Either the eigen-images of PCA (see [3]) or basis images of NMF (see Section 2) can be used to represent the whole training datasets.

Before any compression, the reconstructed image from PCA or NMF is subtracted from the images that need to be compressed (so-called testing images). More formally:

$$\mathbf{R}_i = \mathbf{I}_i - \mathbf{M} \quad (4)$$

where $\mathbf{R}_i$, $\mathbf{I}_i$ and $\mathbf{M}$ are the i-th residue image, the i-th image to be compressed, and the reconstructed template image, respectively. Now, the sender compresses the residue $\mathbf{R}_i$ (instead of the original image $\mathbf{I}_i$), using any arbitrary lossless compression tool. Then, it sends the compressed residue $\acute{\mathbf{R}}_i$ to the receiver. At the other side, the receiver needs to decompress each $\acute{\mathbf{R}}_i$ to retrieve corresponding $\mathbf{R}_i$, using the same lossless algorithm. Then, the template image $\mathbf{M}$ is added to each $\mathbf{R}_i$ to reconstruct all the original images $\mathbf{I}_i$'s. Clearly, the more images to compress, the higher the improvement this method can achieve.

The main steps of multilevel MAC framework are very similar to the components of the single-level MAC framework. The key difference between these two models is that under the multilevel framework, the residue images from the previous levels are treated as the next level "'original" images. In other words, assume that $\mathbf{M}^0$ is the reconstructed image learned from $\mathcal{T}$ (similar to $\mathbf{M}$ in single-level MAC). This means $\mathbf{V}$ in Eq. (1) should be constructed based on $\mathcal{T}$. Then, the first set of residue images is defined as follows:

$$\mathbf{R}_i^1 = \mathbf{I}_i - \mathbf{M}^0 \quad (5)$$

At every level, the basis image $\mathbf{M}^j$ ($\forall j > 0$) is obtained from $\mathcal{R}^j = \{\mathbf{R}_1^j, \mathbf{R}_2^j, \cdots, \mathbf{R}_{|\mathcal{I}|}^j\}$, which is simply the set of all residue images computed at that level. This means for the calculation of $\mathbf{M}^j$, the $\mathcal{R}^j$ should be used to form $\mathbf{V}$ in Eq. (1). The residue images at the j-th level ($j > 0$) can simply be computed as:

$$\mathbf{R}_i^j = \mathbf{R}_i^{j-1} - \mathbf{M}^{j-1} \tag{6}$$

In the multilevel MAC framework, the sender only compress the residue images from the last level (i.e., $\mathbf{R}^l$). Then, it sends the compressed residues along with all template images (i.e., $\mathbf{M}^j$s). Figure 2 illustrates the different steps of the proposed multilevel MAC framework.

Note that adding any extra level yields to two contradicting objectives: On one hand, each extra level reduces the energy of the last residue images and improves the compression ratio. On the other hand, each level produces one more template image which should be transferred (once per set). Consequently, to achieve the best result, the number of levels should be carefully chosen. In brief, the proposed NMF-based multilevel MAC encoder works as follows:

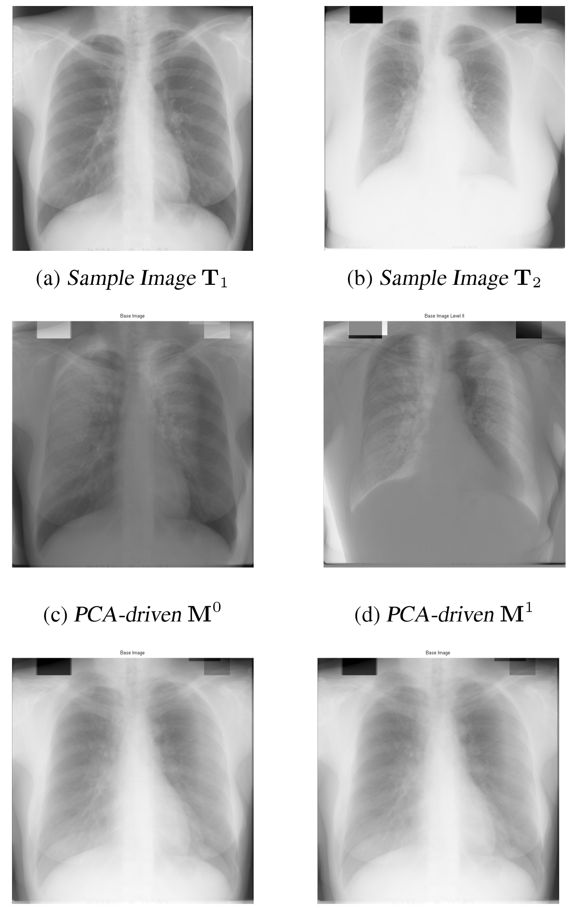
1. Learn similarities across a set of train images $\mathcal{T}$ by applying NMF. Store the resulting template into $\mathbf{M}^0$ (see Section 2).
2. Get the input test image set $\mathcal{I} = \{\mathbf{I}_1, \mathbf{I}_2, \mathbf{I}_3, \cdots, \mathbf{I}_{|\mathcal{I}|}\}$.
3. Obtain residue-set $\mathcal{R}^1$ by subtracting $\mathbf{M}^1$ from $\mathcal{I}$ (see Eq. (5)).
4. Apply NMF on the residue-set from the previous level ($\mathcal{R}^{j-1}$) to obtain $\mathbf{M}^j$.
5. Calculate $\mathcal{R}^j$ according to Eq. (6).
6. Repeat step 4 and 5 until final residue-set ($\mathcal{R}^l$) is computed. Then go to the next step.
7. Store $\mathcal{M} = \{\mathbf{M}^1, \mathbf{M}^2, \mathbf{M}^3, \cdots, \mathbf{M}^l\}$.
8. Apply any lossless compression algorithm on $\mathcal{R}^l$ to obtain $\acute{\mathcal{R}}$.

In order to reconstruct the original image set ($\mathcal{I}$), the following steps must be taken:

1. Apply a decompression algorithm, which matches the compression technique that is applied in step (8), on $\acute{\mathcal{R}}$ to obtain $\mathcal{R}^l$.
2. Add $\mathbf{M}^j$ to each image of $\mathcal{R}^j$ to yield $\mathcal{R}^{j-1}$.
3. Repeat step 2 until $\mathcal{R}^1$ is calculated. Then go to the next step.
4. Add $\mathbf{M}^0$ to each image of $\mathcal{R}^1$ to obtain $\mathcal{I}$.

4. EXPERIMENTS

This section outlines our experimental setup and datasets used. It also presents our initial results and discusses the findings.



(a) *Sample Image* $\mathbf{T}_1$ (b) *Sample Image* $\mathbf{T}_2$

(c) *PCA-driven* $\mathbf{M}^0$ (d) *PCA-driven* $\mathbf{M}^1$

(e) *NMF-driven* $\mathbf{M}^0$ (f) *NMF-driven* $\mathbf{M}^1$

Fig. 1: *Random sample X-ray images (a,b), PCA-driven template images (c,d) and NMF-driven (e,f) template images.*

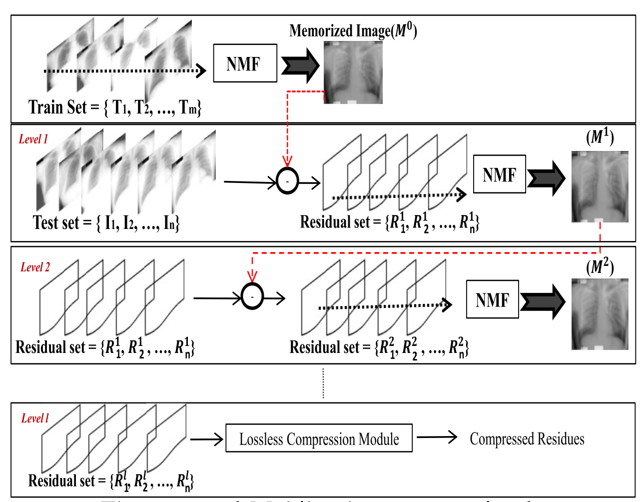


Fig. 2: *The proposed Multilevel memory-assisted compression algorithm.*

4.1. Experiment Setup

As mentioned briefly in Section 1, this paper targets three main goals. Firstly, we wanted to extend our PCA-based work to NMF. Secondly, we further wanted to enhance the single-level MAC framework to a multilevel approach. Finally, we wanted to study the effect of the training and testing sets sizes on the performance of the algorithms.

In order to achieve all aforementioned goals, we examined the performance of single and multilevel PCA-based and NMF-based MAC methods. We also included the well-known Context Tree Weighting (CTW) [8] technique as the non-memory-assisted baseline. Since the superiority of MAC technique over traditional lossless compression algorithms has been shown in [3], there was no need to compare the new MAC framework to the traditional methods. Note that in all MAC techniques, the same implementation of CTW is used to provide a fair comparison. Because of space limitation, the comparison between MAC algorithms with different lossless encoders is left for the future studies.

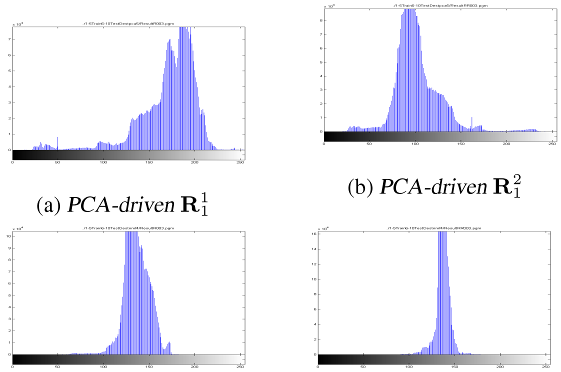
To study the sensitivity of the algorithms to the training and testing dataset sizes, three different sizes (i.e., 5, 10, 15) were examined for both training and testing sets. Therefore, each method is run nine times, in total. Note that for all cases, training and testing sets are completely exclusive.

In this study, all the experiments were carried out on the lung images from the famous Japanese Society of Radiological Technology dataset [9]. All raw images were of size 2048×2048 pixels, encoded in 8-bit gray-scale format. Figures 1.a and 1.b display two randomly chosen samples from the dataset. In future work, we plan to include additional medical image datasets.

4.2. Experiment Results and Discussion

Figure 3 illustrates the histograms of the residue images of a randomly selected X-ray image. The residues in figures 3.a and 3.b are computed using PCA, whilst figures 3.c and 3.d are produced using NMF-driven templates. A comparison between the histograms reveals that both modifications (NMF and multilevel) narrow the distribution of pixel values and increase the number of pixels with zero value. Furthermore, table 1 shows that the entropies of raw image, and image after PCA and NMF reconstruction with different levels. The table demonstrates a decreasing entropy trend of raw image after decorrelation by PCA, NMF for different levels. It can be clearly seen that the entropy of raw image is substantially decreased by NMF compared to PCA. This means the resulting residue images can be compressed much more efficiently.

The experimental results of all algorithms with nine different training/testing set sizes are depicted in Figure 4. The bar charts compare the algorithms according to the compression ratio improvements over the memory-less CTW [8]. Each sub-figure depicts the relative improvement of the algorithms when applied on a fixed number of training images.



(a) *PCA-driven* $\mathbf{R}_1^1$ (b) *PCA-driven* $\mathbf{R}_1^2$

(c) *NMF-driven* $\mathbf{R}_1^1$ (d) *NMF-driven* $\mathbf{R}_1^2$

Fig. 3: *Histograms of residue images of a randomly chosen X-ray image. Template images are computed either by single-level/two-level PCA (a,b) or by single-level/two-level NMF-based template extraction (c,d).*

In contrast, different groups of bars within a sub-figure show the performance of the algorithms when the number of test images varies. Note that in all figures, larger values for compression ratio improvement indicate better performance.

Table 1: *Entropy Results*

Entropy	PCA	NMF
Raw image	6.7569	6.7569
$\mathbf{M}^1$	6.2984	7.1897
$\mathbf{M}^2$	6.1086	6.2888
$\mathcal{R}^l$	6.1118	5.0064
$\mathcal{R}^2$	5.6630	5.0050

As Figure 4 confirms, all variants of MAC framework improve the no-memory-assisted baseline with at least 39%. Among all the variations, single-level PCA and multi-layer NMF (i.e., NMF 2) demonstrate the least and the most improvements, respectively. Indeed, the best performance among all variations is the multi-layer NMF-based algorithm with 15 train and 15 test images (62.25% improvement).

By comparing NMF variations with similar PCA ones, we see that NMF-based algorithms are the preferred algorithms. On the other hand, comparison between single and two-level methods reveals that the multi-level learning significantly improves its single-level parent.

Another observation worth mentioning here is the fact that is when the number of training images is fixed, any increment in the number of testing images results in better improvement without any exception. This effect is more visible when the number of training images is moderate (see figure 4.b). For small and large training sets (e.g.,figures 4.a and 4.c), the effect of testing set size is superficial. In general, growth in

(a) $|\mathcal{T}| = 5$ (b) $|\mathcal{T}| = 10$ (c) $|\mathcal{T}| = 15$

Fig. 4: *Comparing PCA, NMF, two-level PCA, and two-level NMF MACs to CTW.*

the size of training set enhances compression ratio. However, there is one exception; Comparing PCA-based methods (both one and two-level) in Sub-figures 4.a and 4.b shows that when the number of images in testing set is not large enough, increasing the number of training samples from 5 to 10 has an adverse effect on the compression ratio. This phenomenon is unique to PCA-based MAC and is not seen in NMF-based techniques. Comparing figure 4.b with figure 4.c confirms that this surprising phenomenon is just an exception and does not occur again when the number of training samples increases from 10 to 15 images.

5. CONCLUSION

In this study, we improved our previously developed memory-assisted lossless compression algorithm by using a more relevant application-inspired template extractor and by increasing the levels of set similarity learning. Our experimental results on the JRST medical images confirm that both modifications successfully improve the compression ratio. Furthermore, we showed that larger training and testing sets enhance the performance of all variants of memory-assisted techniques. Since the proposed framework and its parent are still very recent, more future work can be carried to study and enhance these algorithms. For example, we are interested to find a rule-of-thumb for choosing the number of learning levels. We also aim at investigating the effect of different lossless encoders (e.g., CALIC and JPEG-LS) on the performance of the proposed techniques. Finally, we plan to include more medical image datasets to draw stronger conclusions.

6. ACKNOWLEDGMENTS

This work was sponsored by RIMT (Australia), University of Aveiro (Portugal), and the National Plan for Science, Technology and Innovation (MAARIFAH)-KACST-KFUPM S&T Unit (Saudi Arabia), Project 12-BIO2366-04.

REFERENCES

[1] B Alagendran and S Manimurugan, “A survey on various medical image compression techniques,” *International Journal of Soft Computing and Engineering (IJSCE) ISSN*, pp. 2231–2307, 2012.

[2] M Ferni Ukrit, A Umamageswari, and Dr GR Suresh, “A survey on lossless compression for medical images,” *International Journal of Computer Applications (0975–8887) Volume*, 2011.

[3] 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*. IEEE, 2014, pp. 2030–2034.

[4] Yasser El-Sonbaty, Meer Hamza, and George Basily, “Compressing sets of similar medical images using multilevel centroid technique,” in *In Proceedings of Digital Image Computing: Techniques and Applications*. Citeseer, 2003.

[5] Samy Ait-Aoudia and Abdelhalim Gabis, “A comparison of set redundancy compression techniques,” *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 216–216, 2006.

[6] Kosmas Karadimitriou, *Set redundancy, the enhanced compression model, and methods for compressing sets of similar images*, Ph.D. thesis, Citeseer, 1996.

[7] Ian Jolliffe, *Principal component analysis*, Wiley Online Library, 2002.

[8] Nicklas Ekstrand, “Lossless compression of grayscale images via context tree weighting,” in *Data Compression Conference, 1996. DCC’96. Proceedings*. IEEE, 1996, pp. 132–139.

[9] Japanese Society of Radiological Technology, “JSRT Database,” `http://www.jsrt.or.jp`.

[10] Daniel D Lee and H Sebastian Seung, “Algorithms for non-negative matrix factorization,” .