# Compressing Sets of Similar Images

Samy Ait-Aoudia[1], Abdelhalim Gabis[1], Amina Naimi[2]
[1]INI – Institut National d'Informatique, BP 69M, Oued-Smar 16270, Algiers, Algeria
s_ait_aoudia@ini.dz, a_gabis@ini.dz
[2]INRIA Rocquencourt – Institut National de Recherche en Informatique et Automatique, 78153 Le Chesnay, France
amina.naimi@inria.fr

*Abstract* - **Databases applications are widely used nowadays. Among the "visual" data widely manipulated we find medical and satellite images. Applications using these types of data, produce a large amount of similar images. Thus a compression technique is useful to reduce transmission time and space storage. Lossless compression methods are necessary in such critical applications. Set Redundancy Compression (SRC) methods exploit the inter-image redundancy and achieve better results than individual image compression techniques when applied to sets of similar images. In this paper, we make a comparative study of SRC methods using standard archivers. We also propose a new SRC method based on Jiang Predictor.**

*Keywords* - **Set Redundancy Compression, Similar Images, Compression, Min-Max, Centroid.**

## I. INTRODUCTION

Databases applications are widely used nowadays. Among the data widely used we find medical and satellite images. Applications using these types of data, produce a large amount of similar images. Thus a compression technique is useful to reduce transmission time and space storage. In addition, medical or satellite images must be stored without any loss of information since the fidelity of images is critical in diagnosis. This requires lossless compression techniques.

Image compression techniques (see [1,2]) concentrate on how to reduce the redundancies presented in an individual image. This model of compression ignores an additional type of redundancy that exists in sets of similar images, the "Set redundancy".

The term "Set Redundancy", was introduced for the first time by Karadimitriou [3] and defined as follows: "*Set redundancy is the inter-image redundancy that exists in a set of similar images, and refers to the common information found in more than one image in the set*". The compression techniques based on set redundancy follows the model presented in the figure 1.

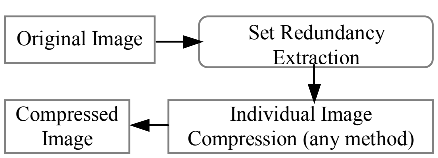


Fig. 1: SRC methods improvement on similar images.

These methods are referred to as SRC (for Set Redundancy Compression) methods. After Set Redundancy extraction, any compression algorithm can be applied to achieve higher compression ratios.

In this paper, we test Set Redundancy Compression (SRC) methods combined with standard archivers. The SRC methods tested are : the Min-Max Differential method (MMD), the Min-Max Predictive (MMP) method, and the Centroid method. The archivers used for individual compression are : Rar compressor which is based on [4,5,6] the Zip archiver and the Huffman Encoder [4].

This paper is organized as follows. The different SRC methods are explained in section 2. We present in section 3, a new predicting scheme for the Min-Max Predictive method. Experimental results on CT (Computed Tomography) and satellite images are given in section 4. Section 5 gives conclusions.

## II. SET REDUNDANCY METHODS

In this section we present the different types of SRC methods : the Min-Max Differential method [7], the Min-Max Predictive method [8], and the Centroid [9]. These methods are fast, lossless and easy to implement.

### *II.1. Min-Max Differential*

MMD uses for extracting the "Set Redundancy" a "maximum image" and a "minimum image". To create the minimum (Min) image, the pixel values across all the images in the set are compared, and for each pixel position the smallest value is chosen. Similarly, the maximum (Max) image is created by selecting the largest pixel value for each pixel position. Then, the set redundancy can be reduced by replacing every image in the set by its differences from the Min or the Max image such that for every pixel position MMD fins and stores the smallest difference value (see figure 2).

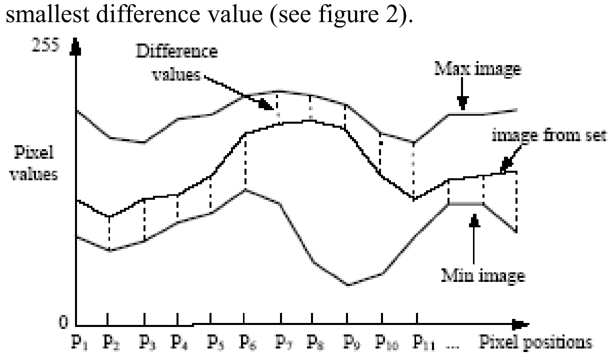


Fig. 2: Min-Max Differential method.

To synchronize encoding and decoding, the encoder uses consistently Min or Max curves until it finds a difference value larger than (max-min)/2. In that case, it encodes this value and switches to the other curve. The decoder follows the same rule; when it finds a difference larger than (max-min)/2 it also switches to the other curve.

*II.2. Min-Max Predictive*

The MMP method uses also the Min and Max images. For each position $i$, the Min image provides the minimal value $min_i$ of all the images, likewise the image Max provides the maximum value $max_i$. These two values are the limits of the range of the possible values that a pixel $i$ can have. By dividing the interval into $N$ levels, each pixel position $i$ can be represented as level $L_i$ between its corresponding minimum and maximum values. The level $L_i$ is given by the equation:

$$L_i = N\left(\frac{value(P_i) - \min_i}{\max_i - \min_i}\right) \quad 1$$

From equation (1), a prediction scheme for the value of pixel $P_i$ can be defined as :

$$value\_predicted\ (P_i) = \min_i + \frac{L'_i}{N}(\max_i - \min_i) \quad 2$$

Where:

$L'_i$ : the level predicted for a pixel at position $i$.

$N$ : number of levels.

The prediction concerns only the element $L'i$ in the preceding formula. The MMP method predicts the value of a pixel $P_i$ by using the level information from treated neighboring pixels. Three predictors were given in [8]:

$L_i = L_{i-1}$

$L_i = (L_{upper} + L_{left}) / 2$

$L_i = L_{upper} + L_{left} - L_{upperleft}$

These predictors determine three variations of Min-Max Predictive methods referred to as MMP1, MMP2, and MMP3. The encoding process consists of storing the differences between the predicted values and the original values. These differences replace the original values. To restore the original image from the differences stored, the decoding process calculate the predicted values, and then adds the corresponding differences values.

*II.3. The Centroid Method*

The "Centroid" method [9], uses the average image of a set of similar images to predict the values of the difference image. If the prediction is efficient enough, the difference image will contain small values having a Laplacian distribution with most of values very close to zero.

A simple scheme for predicting the pixel value at position $i$ in image $j$ is :

$$F_{i,j} = m_i \quad 3$$

Where $mi$ is the average value at position $i$,. and $Fi$ is the predicted value. This scheme is not very efficient. A more sophisticated scheme [9] can be expressed by:

$$F_{i+1,j} = m_{i+1} + x_{i,j} - m_i \quad 4$$

$$D_{i+1,j} = x_{i+1,j} - F_{i+1,j} \quad 5$$

Where $F_{i+1,\,j}$ is the predicted value at position $i+1$, $X_{i,j}$ is the pixel value at position $i$, $m_i$ is the average value of position $i$ across all images, and $D_{i+1,j}$ is the difference value of position $i+1$ in image $j$ between the original and the predicted values. The detailed demonstration of equations 4 and 5 can be found in [3]. The equation 4 is so called the centroid method.

## III. PROPOSED METHOD

The predicting scheme extends the scheme used in the MMP3 method. It's based on the Jiang Predictor proposed in [10] to improve the JPEG-LS Algorithm for compressing individual images.

From the Jiang predictor we derive a predicting scheme that will be used in the MMP method instead of the schemes defined in II.2. the derived scheme is as follows:

**if** ($L_{upperleft} \geq \max(L_{left}, L_{upper})$)
    **if** ( $L_{upperleft} - \max(L_{left}, L_{upper}) > 10$
        and $L_{upperright} < L_{upper}$ and $L_{left} - L_{upper} \geq 5$ )
    $Li = (L_{upperright} + \min(L_{left}, L_{upper})) / 2$
    **else**
        $Li = \min(L_{left}, L_{upper})$
**else**
    **if** ( $L_{upperleft} \leq \min(L_{left}, L_{upper})$ )
    **if** ($10 \leq L_{upperright} - L_{upper} \leq 50$ and $|L_{upper} - L_{left}| \leq 10$
        and $\min(L_{left}, L_{upper}) - L_{upperleft} \geq 5$)
    $Li = (L_{upperright} + \max(L_{left}, L_{upper})) / 2$
      **else**
        $Li = \max(L_{left}, L_{upper})$
    **else** $Li = L_{upper} + L_{left} - L_{upperLeft}$;

$L_{left}$, $L_{upper}$, $L_{right}$, $L_{upperright}$, $L_{upperleft}$ are the levels of the neighboring pixels of the pixel being treated.

We call the new method resulting from this predicting scheme: MMPJ for MMP Jiang.

## IV. EXPERIMENTAL RESULTS

The evaluation of set redundancy methods is made on sample sets of medical and satellite images. The first set is constituted by 10 512×512 CT medical images taken from "M.D. Anderson Cancer Center in Houston, Texas". This set is the same used in [3,7,8,9]. The second set is composed of a 10 640×512 satellite images that show meteorological images of Australia. The images were collected from Australian Bureau of Meteorology. All images were gray-level, and were scaled to 8 bits/pixel. All experiments were performed under Windows XP operating system.

To make the evaluation of the SRC methods, we have used the standard compression algorithms RAR, ZIP, HUFFMAN. The images are compressed by these algorithms with and without using the set redundancy extraction. Each algorithm is tested separately and the attained compression ratios are compared. The compression ratio is given by :

$$R = \frac{Size(original\_image)}{Size(compressed\_image)} \quad 6$$

The improvement against standard compression method is also needed in the evaluation. It shows if the use of SRC methods is really effective. The improvement in compression is defined by:

$$A = \frac{R_{SRC} - R}{R} \quad 7$$

where $R$ is the compression ratio achieved when using a regular compression method only, and $R_{SRC}$ is the compression ratio achieved when combining SRC with that regular compression method.

### *IV.1. Medical images experiments*

The set of CT images used in the experiments is shown in figure 3. The set contains axial CT brain scans where horizontal slices of the brain at the eye-level are depicted. The images were selected from patients of both sexes, various ages, and a variety of pathological conditions. Compression ratios and improvement in compression when using SRC methods are presented in table 1. The histograms representing improvements and compression ratios are shown in figure 4 and figure 5 respectively.

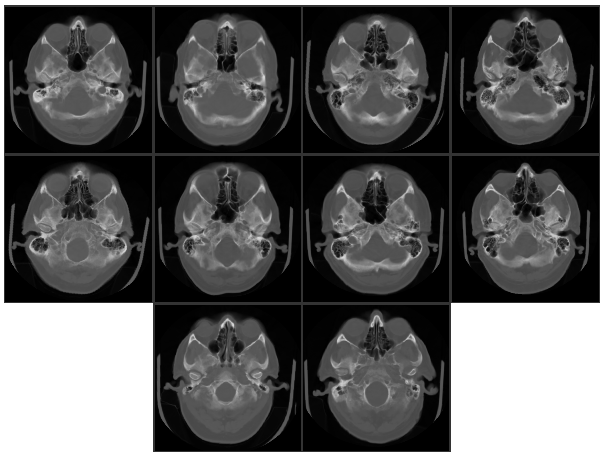

Fig. 3: Medical test images.

TABLE 1: EXPERIMENTAL RESULTS ON MEDICAL IMAGES.

| Compression Technique | Average Size (KO) | Average compression ratio | Improvement |
|---|---|---|---|
| Original Image | 256 | - | |
| Huffman | 193.45 | 1.32 :1 | |
| Centroid + Huffman | 98.41 | 2.60 :1 | 96 |
| MMD + Huffman | 125.93 | 2.03 :1 | 54 |
| MMP1 + Huffman | 84.08 | 3.04 :1 | 130 |
| MMP2 + Huffman | 75.35 | 3.39 :1 | 156 |
| MMP3 + Huffman | 69.15 | 3.70 :1 | 180 |
| MMPJ + Huffman | 67.41 | 3.79 :1 | 187 |
| RAR | 76.09 | 3.36 :1 | |
| Centroid + RAR | 72.60 | 3.52 :1 | 4 |
| MMD + RAR | 82.52 | 3.10 :1 | -7 |
| MMP1 + RAR | 67.37 | 3.80 :1 | 13 |
| MMP2 + RAR | 62.73 | 4.08 :1 | 21 |
| MMP3 + RAR | 57.37 | 4.46 :1 | 32 |
| MMPJ + RAR | 55.68 | 4.60 :1 | 37 |
| Zip | 99.94 | 2.56 :1 | |
| Centroid + Zip | 80.47 | 3.18 :1 | 24 |
| MMD + Zip | 87.35 | 2.93 :1 | 14 |
| MMP1 + Zip | 75.94 | 3.37 :1 | 31 |
| MMP2 + Zip | 68.36 | 3.74 :1 | 46 |
| MMP3 + Zip | 66.36 | 3.85 :1 | 50 |
| MMPJ + Zip | 64.55 | 3.97 :1 | 55 |

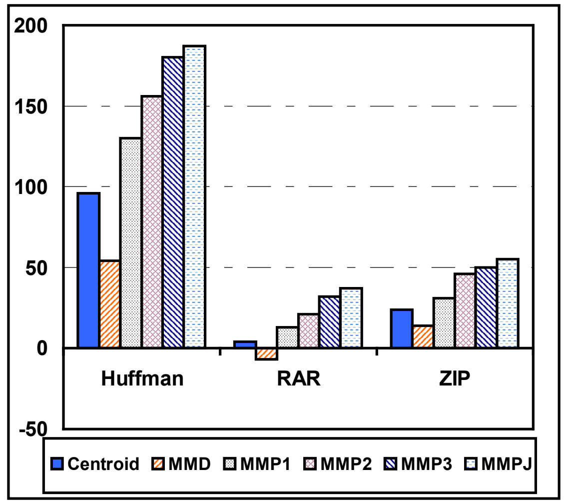


Fig. 4: SRC methods improvement on medical images.

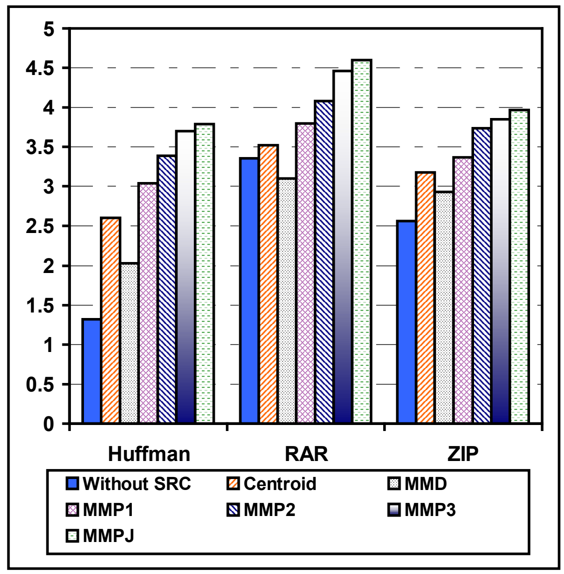


Fig. 5: Average compression ratios on medical images.

### *IV.2. Satellite images experiments*

Results of tests on satellite images (Figure 6) are presented in table 2. The histograms representing improvements and compression ratios using SRC methods are shown in figure 7 and figure 8 respectively.

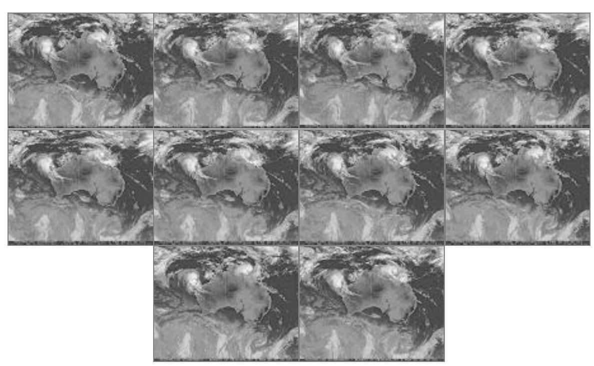

Fig. 6: Satellite test images.

TABLE 2: EXPERIMENTAL RESULTS ON SATELLITE IMAGES.

| Compression Technique | Average Size (KO) | Average compression ratio | Improvement |
|---|---|---|---|
| Original Image | 320 | - | |
| Huffman | 284.71 | 1.12 :1 | |
| Centroid + Huffman | 228.59 | 1.40 :1 | 25 |
| MMD + Huffman | 250.28 | 1.28 :1 | 14 |
| MMP1 + Huffman | 189.06 | 1.69 :1 | 51 |
| MMP2 + Huffman | 183.39 | 1.74 :1 | 55 |
| MMP3 + Huffman | 203.25 | 1.57 :1 | 40 |
| MMPJ + Huffman | 187.97 | 1.70 :1 | 51 |
| RAR | 229.51 | 1.39 :1 | |
| Centroid + RAR | 223.13 | 1.43 :1 | 3 |
| MMD + RAR | 226.34 | 1.41 :1 | 1 |
| MMP1 + RAR | 191.35 | 1.67 :1 | 20 |
| MMP2 + RAR | 183.30 | 1.74 :1 | 25 |
| MMP3 + RAR | 202.10 | 1.58 :1 | 13 |
| MMPJ + RAR | 189.67 | 1.68 :1 | 21 |
| Zip | 233.24 | 1.37 :1 | |
| Centroid + Zip | 223.46 | 1.43 :1 | 3 |
| MMD + Zip | 230.86 | 1.39 :1 | 1 |
| MMP1 + Zip | 193.56 | 1.65 :1 | 20 |
| MMP2 + Zip | 188.25 | 1.70 :1 | 24 |
| MMP3 + Zip | 204.01 | 1.57 :1 | 16 |
| MMPJ + Zip | 192.69 | 1.65 :1 | 18 |

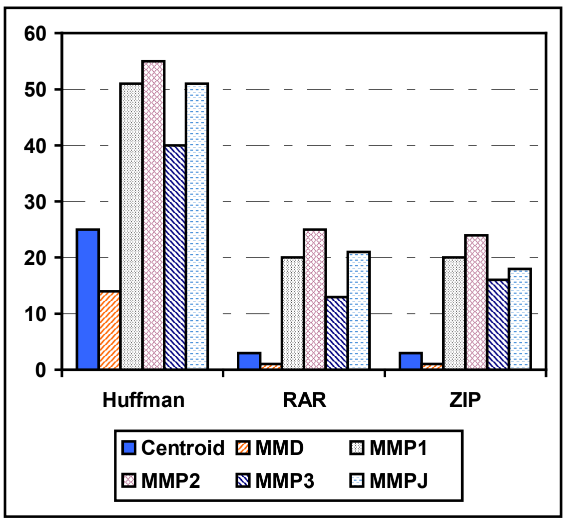


Fig. 7: SRC methods improvement on satellite images.

*IV.3. Discussion*

The best improvement on CT images is carried out by the couple “MMPJ + Huffman” with 187%. On satellite images, “MMP2 + Huffman” carry out the best improvement with 55%. For the compression ratios, “MMP2+Rar” and “MMPJ+Rar” are the best compression couples on satellite images set and CT images set with respectively “1.74:1” and “4.60:1” compression ratios.

As shown in figure 8, the improvement in compression is smaller for the satellite images. This can be attributed to the fact that satellite images are "noisy". The similarity at pixel level is limited on these images.

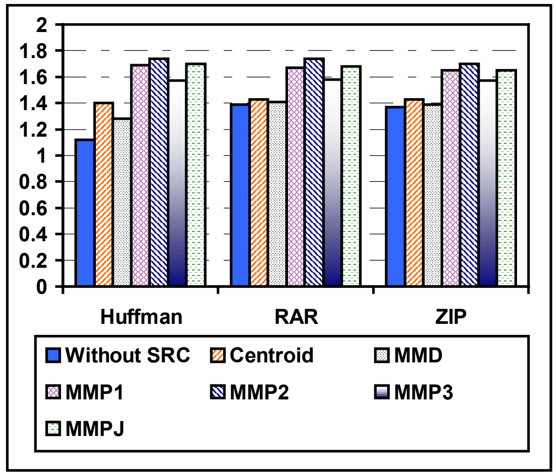


Fig. 8: Average compression ratios on satellite images.

## V. CONCLUSION

From the results shown in previous tables, we see that the majority of SRC methods carry out an improvement compared to standard compression. This is a good indicator for the effectiveness of using SRC techniques on similar images datasets. The results also show that the MMP methods perform better than the other SRC techniques. Further tests must be conducted on large datasets to show the efficacy of using SRC techniques.

## VI. REFERENCES


[1] D. Clunie, Lossless compression of grayscale medical images and effectiveness of traditional and state of the art approaches, in Proc SPIE-Medical Imaging, vol.3980, 2000.

[2] Celik M.U., Sharma G., and Tekalp A.M., Gray-level embedded lossless image compression, ICASSP 2003, 6-10 April 2003, Honk-Kong, pp. III_245-248.

[3] Kosmas Karadimitriou : Ph.D Thesis, Set redundancy, the enhanced compression Model, and methods for compressing Sets of similar images, Louisiana State University Baton Rouge, LA 70803, USA August 1996.

[4] D.A. Huffman, “A method for the construction of minimum redundancy codes”, Proc. IRE, n.40, pp. 1098-1101, Sep. 1952.

[5] Shkarin, D., Improving the efficiency of PPM algorithm, Problems of Information Transmission, 37(3), July 2001, pp. 226-235.

[6] Ziv, and A. Lempel, “A universal algorithm for sequential data compression”, IEEE Trans. Inf. Theory, vol. IT-23, no. 3, pp. 337-343, May 1977.

[7] K. Karadimitriou and J.M. Tyler, The Min-Max Differential Method for Large-Scale Storage and Compression of Medical Images, in Proc of Annual Molecular Biology and Biotechnology Conference, Baton Rouge, Feb. 1996.

[8] Karadimitriou K. and Tyler J.M., Min-max compression methods for medical image databases, ACM SIGMOD Record Volume 26 , Issue 1, March 1997, pp. 47-52.

[9] Karadimitriou K. and Tyler J.M., The Centroid method for compressing sets of similar images, Pattern Recognition Letters vol. 9, n° 7 (1998), pp. 585-593.

[10] J.Jiang, B.Guo and S.Y.Yang. Revisiting the JPEG-LS prediction scheme. IEEE Proc-Vis. Image Signal Processing, El. 147, No. 6, December 2000.