Optimizing Multitemporal Data Regression for Minimizing Residual Entropy

AL MAMUN, XIUPING JIA and MICHAEL RYAN
School of Engineering and Information Technology
University College, University of New South Wales
Canberra, ACT2600
AUSTRALIA
M.Mamun@student.adfa.edu.au, {X.Jia, M.Ryan}@adfa.edu.au

Abstract: - Comparison of two images of the same object or phenomenon taken some time apart is a bit complicated due to fact that, there may be sensor variation, illumination variation, non-uniform attenuation, atmospheric absorption or any other environmental effects which may render some changes. That eventually makes it quite difficult to achieve a good compression rate in terms of any prediction based approach. That's why optimizing the functional relation between two distinct time satellite image is the basic and prerequisite for many studies in the sequential transmission of remote sensed data. Shannon's entropy represents an absolute limit on the best possible lossless compression of any communication. In the sequential transmission of the remote sensed data the images will be available frequently, so we need to achieve the lowest entropy of the data to be sent. We want to have an optimization process that will minimize the residual entropy of any regression based prediction approach. Minimizing the entropy of the residual for any prediction will in fact make the residual symbol's probability as independent as possible (decor related). In this paper we have used simple linear model to be optimized by the modified objective function of entropy minimization rather that sum square based least square approach. We have used brute force search or exhaustive search for the optimization process. Considering our objective function, optimization search techniques like Simulated Annealing, Tabu search, Genetic algorithm, Cross Entropy method are the best candidates for this procedure. From the experimental results we can easily differentiate our procedure from the SSE based regression analysis. In fact in most of the cases of the image pairs our optimization method achieves lower entropy than that of the least square regression analysis.

Key-Words: - Entropy, Regression, Residual, Temporal, REM and Optimization.

1 Introduction

Frequently acquired satellite images of the same scene exhibit high degree of temporal correlation. This correlation together with the spectral and spatial correlation is the source of data compression in the field of image transmission and for the archival purpose. We are considering the case of sequential image data transmission [1] [2] where images will be available every month, week or perhaps day. A better way to transmit a current image data set is to predict that from the previous image which are avaiable at both transmission and receiver ends.

The key issue is to quantify the relationship between the two images of a same scene some time apart by selecting a suitable regression model. The image which has already been transmitted is used for prediction in this case. With a given regressive model, we need to estimate the model parameters. The most common regression analysis finds the parameters to predict the data that fits the current data best in a least-squares sense. This optimization is based on minimizing the sum squared error (SSE). The efficient transmission of the residual or the error of the prediction has also drawn high attention of many researchers in the past few years. Context modelling has been used for the adaptive correction to remove any correlation that remains in the residuals [3] [4]. In [5] [6] the authors come up with FFT based regression method to increase the coefficient of determination of linear fitting. Unfortunately, it increases the size of the residual unless the phase of the current data is added back. This phase locking requirement is not suitable for sequential data transmission since the receiver has no phase information on the current data.

When coding techniques, such as Huffman coding, are applied to the residual data, the maximum compression which can be achieved is determined by residual's entropy. In this paper, we introduce an optimization process that will minimize the number of individual symbols in the residuals, ie., minimizing the entropy. Considering our objective function, optimization search techniques like Simulated Annealing, Tabu search, Genetic algorithm, Cross Entropy method are the candidates for this procedure. We have used brute force search or exhaustive search for the optimization process. Experiments were conducted to illustrate the proposed optimization process using a subset of Landsat ETM data.

2 Temporal Prediction Model

Assume that the receiver end has a copy of the previous image in the sequence, $\boldsymbol{X_1}$. The current data $\boldsymbol{X_2}$ will be predicted using $\boldsymbol{X_1}$. The first issue in temporal prediction is which model to adopt. While the same scene taken some short time apart often shows little change in terms of the ground cover types in the sequence of images, it is not realistic to expect that the non-changed areas would have identical brightness values, even though they may be only one day apart. Noticeably differences are often induced by sensor noise, illumination variation, or atmospheric effect [7]. Since those factors are often random, a linear regression is generally selected to model the relationship between the two bands recorded at two dates. Higher ordered regression is not considered in this study, since there is no strong evidence of nonlinear relationship present in the data and nonlinear models introduce more model parameters to transmit. Based on this understanding, the first degree polynomial is used in this study.

$$\boldsymbol{X}_2' = a \times \boldsymbol{X}_1 + b \quad (1)$$

where a and b are the parameters of the linear model. The errors of this temporal prediction form the residual image, r,

$$\mathbf{r} = \boldsymbol{X}_2 - \boldsymbol{X}_2' \quad (2)$$

3 Entropy Minimization Approach

The best model parameters, a and b, can be derived by an optimasation process. The results are data dependent, and vary with the selection of object function. A common objective function is the Sum of Squared Error of the regression.

$$\min\{\mathbf{r}_{SSE}(a,b)\} = \min\left\{\sum_{i,j}\left[x_2(i,j) - (ax_1(i,j)+b)\right]^2\right\}$$
$$i = 1,\ldots,\ M,\ \ j = 1,\ldots,\ N \quad (3)$$

M and N are the number of rows and collumns of the image.

To improve the model fitness, R^2, Fourier transform has been introduced [3, 4], the objective function becomes

$$\min\{\mathbf{Fr}_{SSE}(a,b)\} = \min\left\{\sum_{i,j}\left[\left|FFT\{x_2(i,j)\}\right| - \left|FFT\{ax_1(i,j)+b\}\right|\right]^2\right\}$$
(4)

The problem for the above object functins is that it may not minimize the entropy of the residual.

Entropy is the measure of randomness of any data indicating the minimum amount of information contained in that data. If the residual derived from the regression based prediction has a flat histogram then the entropy will be the highest and it will take large bandwidth to send that data. From Shannon information theory the entropy of the residual image, **r**, is defined as:

$$E_r = -\sum_r P_r \log P_r$$
$$r = 1, \ldots L_r \quad (5)$$

where P_r is the probability of r. L_r is the number of brightness values in the residual image.

The entropy indicates the minimum bits required to represent the data. For lossless compression, the highest compression rate relies on minimizing the residual entropy. Therefore, the object function (3) is proposed to be replaced by

$$\min\{\mathbf{r}_E(a,b)\} = \min\left\{-\sum_r P_r \log(P_r)\right\} \quad (6)$$

4 Discrete Space Optimization and Entropy Minimization

The issues associated with optimization are the speed and reliablity. It can be a computational costly and may run into a risk of local minimum. The brute force search or exhaustive search is adopted in this study since it is straightforward and easy to implement.

The temporal correlation between two images of the same scene taken at distinct time can be of strongly positive or strongly negetive. If we consider the strong positive case then the search space varied along the positive $\boldsymbol{X_1}$ axis to positive $\boldsymbol{X_2}$ axis in the first quadrant. Considering them as the different states of correlation, the optimization procedure probabilistically decides between moving the system to the next state or staying in the current state. The probabilities are chosen so that the system ultimately tends to move to a state of lower energy. Typically this step is repeated until the system reaches a state that is good enough for the application to

optimize the entropy of the residual for the prediction or until a given computation budget has been exhausted. So it is like a global optimization problem that locates a good approximation to the global optimum of our given objective function in a large discrete search space.
Brute-force search or exhaustive search examines all possible candidates for the solution and checking whether each candidate satisfies the problem statement. Let $a = \tan(\alpha)$. For the linear predictive model the search space, in general, for the parameters α and b are,

$$0 < \alpha < \pi/2$$
$$-L_x < b < L_x$$

where L_x is the maximum value of the image. For example, $L_x = 255$ if the image is in 8bits.

The step size of the search is critical for obtaining a good trade off between search speed and search quanlity. A small step should be applied whenever the number of iterations is managble.

The proposed entropy minimisation has an advantage of independent of model offset, b. Therefore, only one parameter, a, is required to determine. This saves the computational load greatly.

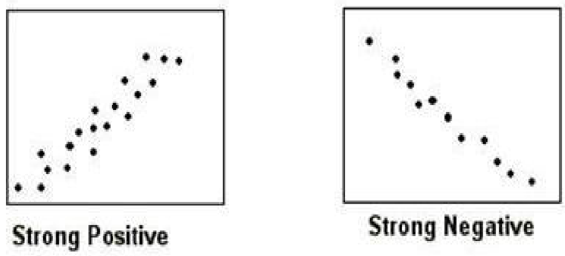


Fig. 1 Possible positive or negative correlation between the current data and previous data.

5 Experiments

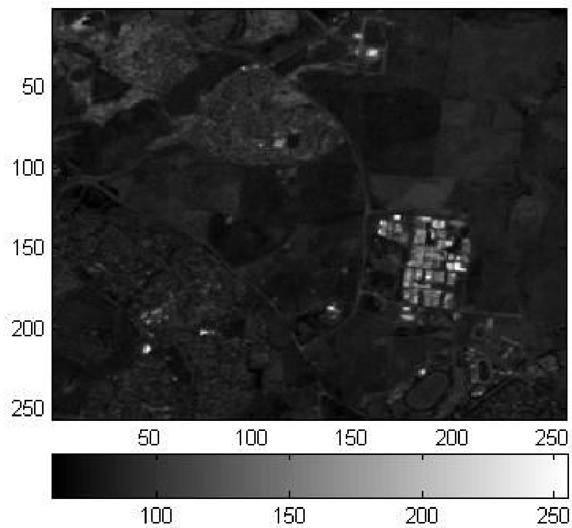
The data used was a subset of Landsat ETM data recorded over Canberra, Australia, in 2000 and 2001 shown in Fig. 2. From the contrast bar shown below of each image it is quite easy to observe that they are taken in two different environmental conditions, yet the temporal correlation is very high that is reflected by the 3D scatter plot shown in Fig. 3.

The residual entropy minimization (REM) was implemented and compared with SSE minimization on the original data and transformed data. The step sizes of 1 and 0.017 were used for the parameters a and b, respectively.

Figure 4 (a) and (b) show the impact of the offset parameter, b, for the SSE and entropy cases, respectively. For a fixed value of a, we have varied b for the REM and SSE optimization. As we can see in Fig. 4, REM minimization doesn't depend on the value of b. So the model $\boldsymbol{X}_2 = a\boldsymbol{X}_1$ is enough for the minimization of REM. Then we can make sure all the residual values positive by adjusting the value of b.

In Fig. 5 and Fig. 6 the entropy based and SSE based optimization results at each iteration are shown. Reducing the step size we can achieve slightly improved results but with the cost of increased time complexity.

Table 1 shows the model parameters achieved through the optimization processes and the resulted entropy of the residual with SSE. In Table 2 the overall results for the remaining bands of the image are given.



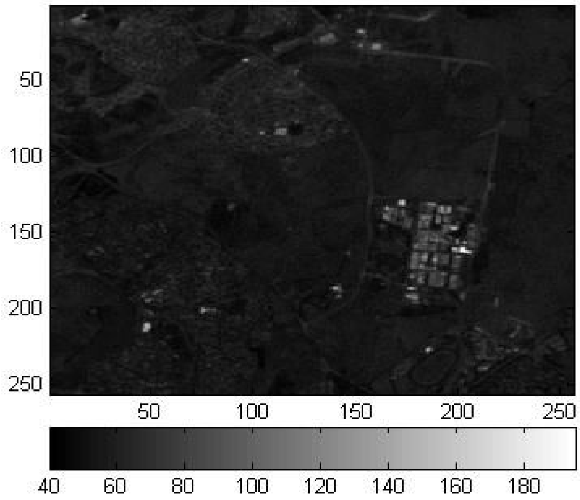


Fig. 2 Images captured in the year 2000 (above) and 2001 (below) used for the experiments (band1).

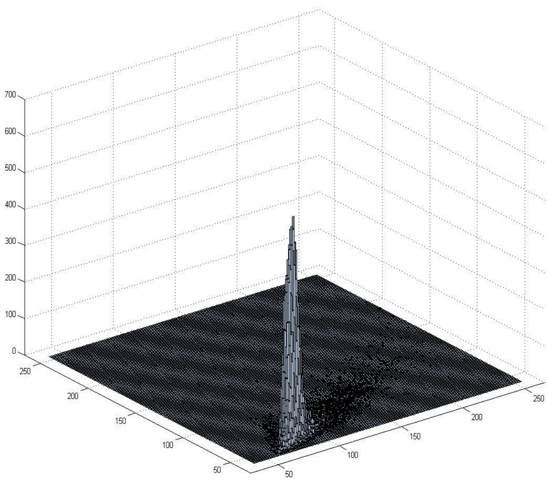
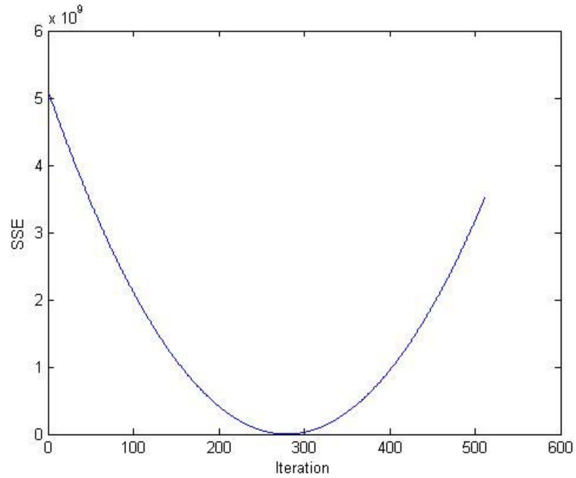
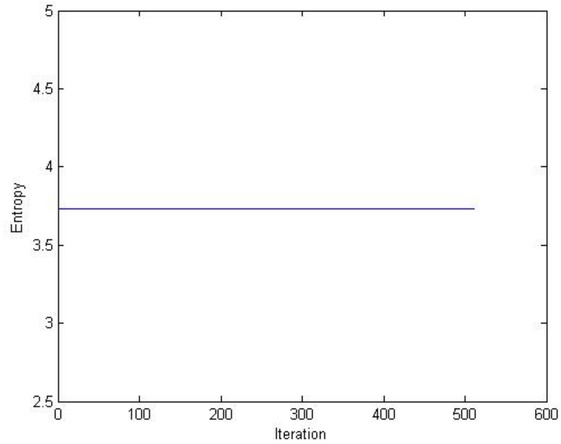


Fig. 3 3D scatter plot between the images given in Fig. 2



(a)



(b)

Fig. 4 Response of the system to varied b (-255 to 255) with a fixed a .

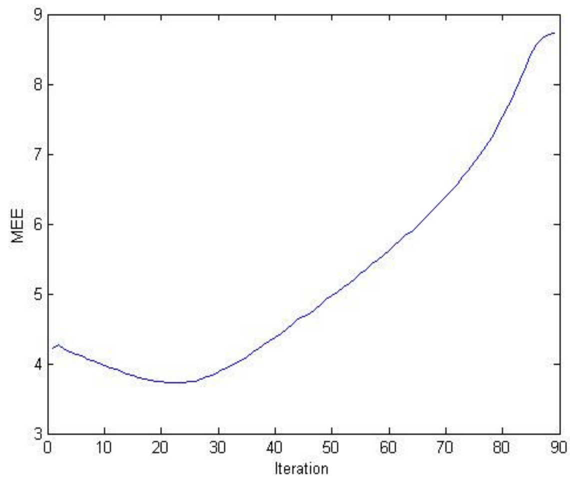


Fig. 5 REM optimization.

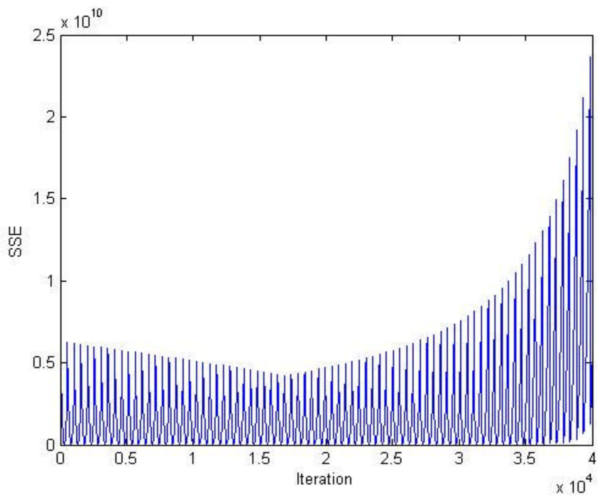


Fig. 6 SSE optimization.

Table 1: Entropy Variation for the Different Approaches (Band1).

Methods	SSE	Residual Entropy
Entropy minimization $a = 0.4040$	32043699	3.7251
Least Square minimization $a = 0.4877, b = 15$	882801	3.7446
FFT regression $a = 0.4951, b = 14.1062$	894690	3.7690

Table 2: Entropy Variation for all other bands with model parameters.

	Least Square minimization		Entropy minimization		FFT regression	
	Model Parameters	Residual Entropy	Model Parameters	Residual Entropy	Model Parameters	Residual Entropy
Band2	$a = 0.4245$ $b = 15$	3.9985	$a = 0.4040$	3.9984	$a = 0.5048$ $b = 9.1941$	4.0421
Band3	$a = 0.4245$ $b = 15$	4.7047	$a = 0.3839$	4.6990	$a = 0.4827$ $b = 10.4717$	4.7434
Band4	$a = 0.8098$ $b = 7$	4.8185	$a = 0.7813$	4.8163	$a = 0.9732$ $b = -5.1422$	4.8949
Band5	$a = 0.5543$ $b = 7$	5.4740	$a = 0.5774$	5.4734	$a = 0.6542$ $a = -4.8098$	5.5034
Band7	$a = 0.4452$ $a = 15$	5.1828	$a = 0.4877$	5.1818	$a = 0.6097$ $b = 2.4895$	5.2433

6 Discussions and Conclusion

The method proposed here regarding the regression analysis between two sequential multispectral images is directly from the data compression point of view. Whatever processing we do in data transmission we have to keep in mind about the entropy since low entropy is a mandatory success factor in any prediction based approaches. Any compression approaches can adopt our way to design their prediction and coding part. While there is no significant improvement in entropy with respect to SSE based regression in the tests presented in this paper, this method reduces the model complexity, and narrows search space and therefore the computational cost is largely saved. Experiments were performed with the images having only 256 gray level values. Improvement in entropy with the images of higher resolution will be very interesting to see.

Most of the coding techniques in modern time are based on the probability theory. So our future plan is to design alternative objective functions, such as mutual information and joined entropy, and optimize them with more sophisticated advanced approaches.

References:

[1] M. A. Mamun, X. Jia and M. Ryan, Adaptive Data Compression for Efficient Sequential Transmission and Change Updating of Remote Sensing Images, *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2009, Cap Town, South Africa.

[2] M. A. Mamun, X. Jia and M. Ryan, Combined Time Domain and Spectral Domain Data Compression for Fast Multispectral Imagery Updating, *Digital Image Computing: Techniques and Applications (DICTA)*, Melbourne, Australia, 2009.

[3] L. Bai, M. He and Y. Dai, Lossless compression of hyperspectral images based on 3D context prediction, *3rd IEEE Conference on Industrial Electronics and Applications(ICIEA 2008)*, No: 3-5, pp: 1845-1848, June 2008.

[4] M. Weinberger, G. Seroussi and G. Sapiro, The LOCO-I Lossless Image Compression Algorithm: Principles and Standardization into JPEG-LS, *IEEE Trans. Image Processing*, No.8, pp.1309–1324, 2000.

[5] K. Peleg, FFT based calibration in remote sensing, *International Journal of Remote Sensing*, No. 19:12, Vol.2301–2315, 1998.

[6] K. Peleg and G. L. Anderson, FFT regression and cross-noise reduction for comparing images in remote sensing, *International Journal of Remote Sensing*, No. 23.10, pp.2097 — 2124, 2002.

[7] J.A. Richards, X. Jia, *Remote Sensing and Digital Analysis*, Springer Verlag, 4th ed, 2005.

[8] http://en.wikipedia.org/wiki

[9] D. Salomon, *Data Compression: The Complete Reference*, Springer, 4th Edition, Dec 2006.