Paper 57

Denoising Aerial Gamma-Ray Survey Data with Non-Linear Dimensionality Reduction

Ramos, F. [1], Dickson, B. [2], Kumar, S. [3]

1. ARC Centre of Excellence for Autonomous Systems, Mechanical and Mechatronic Engineering, The University of Sydney
2. Dickson Research Pty Ltd, 47 Amiens St, Gladesville NSW 2111, Australia
3. 22 Bowness Street, Kellyville, NSW 2155, Australia

ABSTRACT

Denoising aerial gamma-ray surveying makes possible the extraction of previously hidden detail. Conventional methods for denoising spectral data make strong assumptions about the levels and type of noise which reduces their efficiency. The proposed methodology cast the problem as manifold learning followed by non-linear regression. The model makes no assumptions about the level and type of noise and performs significantly better than previous techniques on both synthetic and real data.

INTRODUCTION

Aerial surveying is an important source of information for geological mapping for mineral exploration, soil analysis for agriculture and pollution studies. In nature, the three major gamma-ray emitters are: K (Potassium), U (Uranium) and Th (Thorium). Their radiation can be measured using gamma detectors mounted in a small aircraft flying at low altitudes. This process enables the surveying of extensive areas making information extraction and analysis of the data an economically important procedure. The data so obtained has many uses ranging from geological and soil mapping, exploration for a variety of minerals, environmental studies and land use planning.

The major limitation concerning this technique is noise. Noise has many sources, the dominant one being due to the Poisson nature of radioactive decay. Each area of a survey acts as a separate source with different amounts of source activity. In addition to this, there are variations in the amounts of the vegetation, soil moisture and differences in altitude as well as high frequency variations in the ground concentrations of the three elements. Figure 1 (points) shows a typical gamma spectrum as obtained by the aircraft. The spectrum contains a significant amount of noise which makes the identification of patterns in the data difficult. After denoising, (Figure 1 line) the spectrum reveals clearly the gamma-ray lines associated with the concentration of the three radioactive elements.

Over the past decade a number of noise reducing methods have been introduced for application to aerial gamma-ray survey data. In all these methods the full or substantial sections of all spectra are processed by a statistical procedure so as to produce cleaned spectra before subsequent conventional processes are applied to extract window data. This method of full survey processing was first proposed by Hovgaard and Grasty (1997) who used Singular Value Decomposition (SVD) after first applying a weighting of the count data derived from the sum of all spectra. This method, known as noise adjusted SVD

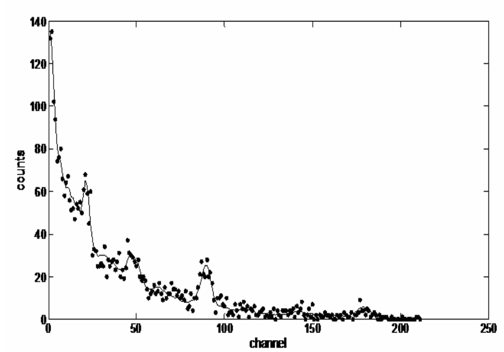

Figure 1: A noisy gamma-ray spectrum (dots) overlain with its noise-reduced daughter after application of NLDR.

(NASVD), assumes the noise in a survey has a Poisson distribution which may be represented by scaling of the average spectrum. It recognizes no other noise sources and is currently the method most commonly used world-wide for noise reduction as well as being recommended by the IAEA (2003).

A method of noise whitening known as Minimum Noise Fraction (MNF, Green et al. 1988) yields noise cleaning that is similar to NASVD (Minty and McFadden, 1998). This method requires that the noise covariance matrix be derived from the

In "Proceedings of Exploration 07: Fifth Decennial International Conference on Mineral Exploration" edited by B. Milkereit, 2007, p. 847-851

data which may be done using the Maximum Autocorrelation Factors (MAF) method of Switzer and Green (1984). Carey and Craig (2004), recognizing that the MAF method may see any major shifts in radiation levels between adjacent areas as noise, presented a method of Least-Square Differences (LSD) to determine the noise covariance matrix prior to the MNF transformation.

Dickson and Taylor (1998) found that an empirical variant of MNF (eMNF) could give much better noise cleaning and edge enhancement in most cases but, with some data sets, undesirable distortions occurred to the data. However, this method showed that there was potential to clean aerial survey data and reveal much more information that was being achieved by either the NASVD or MNF methods. In 2002, Dickson and Taylor introduced a method based on Kohonen Self-Organizing Maps (Kohonen, 2001) that gave results slightly better than eMNF although at the expense of a much greater computing times. This method also suffered from the inherent problems of the self-organizing map method in that there was no formal stopping point for processing.

This paper proposes a denoising algorithm for gamma-ray survey data based on manifold learning. Non-linear dimensionality reduction (NLDR) is employed to compute the underlying structure of the data. By calculating the intrinsic dimensionality of the spectra, the algorithm selects dimensions that are more representative of the data while eliminating dimensions with noise. The most representative dimensions are employed to learn a mixture of linear models through Expectation Maximization (Dempster et al., 1977). Non-linear regression is performed using these mixtures to recover the denoised spectra from the low dimensional representation. The algorithm has been tested on both synthetic and real data and shows significant improvements over all the previous approaches. In particular, it can reveal interesting features in Th/U ratio images that are not seen using other noise-reduction technique.

DENOISING ALGORITHM

Most of the previous methods, except for the SOM method, make strong assumptions over the type of noise. Rather than trying to model the noise directly from the data, the proposed algorithm casts the denoising problem as a manifold learning task. Spectra points in a high-dimensional space (up to 256 dimensions) are assumed to lie in a non-linear manifold (a topological space that is locally Euclidean). Learning the structure of this manifold allows mapping of points to a low-dimensional space while preserving most of the information content. This unsupervised learning problem is also known as dimensionality reduction and has recently received a lot of attention in the machine learning community with several new methods being proposed (Tenenbaum et al. 2000; Roweis and Saul, 2000; Belkin and Niyogi, 2003; Weinberger et al. 2005). Only some of these methods can handle large non-linear manifolds in a computationally efficient manner, a problem expected with the large data sets obtained during aerial surveying. Our approach uses Isomap (Tenenbaum et al. 2000) due to its convenient convergence and isometric properties.

Dimensionality reduction can be interpreted as a form of information compression. Once the structure of the manifold has been recovered, the resulting low-dimensional embeddings contain most of the information while eliminating noise and preserving the main properties of the data. However, the opposite operation, i.e. convert points in the low-dimensional space back to their original high-dimensional space, is not an easy task. To address this problem, we learn a non-linear function representing the Isomap mapping through Expectation Maximization (EM). This function has the form of a mixture of linear models where each component of the mixture is a linear regression function with Gaussian noise. Note, combining multiple linear models for different regions of the input space, results in an overall non-linear function that maps from the high-dimensional inputs to the low-dimensional outputs.

A regression function is applied to convert the embeddings computed from Isomap back to their high-dimensional representation, eliminating noise. This operation is performed by computing probabilistic inferences on the mixture of linear models and results in cleaned spectra. The diagram of Figure 2 depicts the main steps of the algorithm, briefly summarized below:

1. Isomap is applied to the spectra to yield low-dimensional embeddings for the data;
2. The original spectra plus the low-dimensional embeddings are used to train a mixture of linear models through Expectation Maximization;
3. A cleaned spectrum is obtained by computing probabilistic inferences on the mixture of linear models given the embeddings produced by Isomap.

Further detail of the methods employed is given in Kumar et al., 2005.

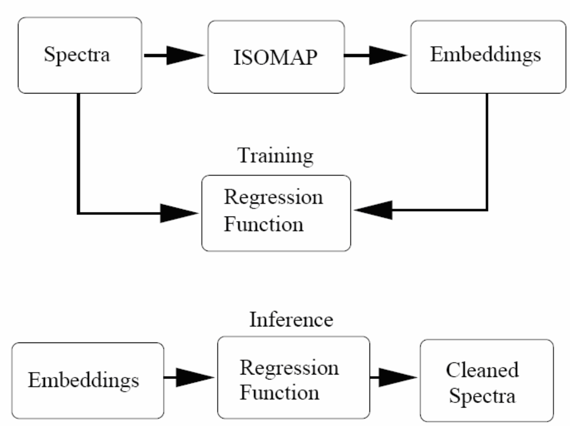

Figure 2: Denoising algorithm. First, Isomap is applied to obtain low-dimensional embeddings for the spectra. The original data in conjunction with low-dimensional points are used to train a non-linear regression model from high to low dimensional spaces. The regression model then converts the embeddings to denoised spectra.

RESULTS

Evaluation of new methods of noise reduction requires studies on both synthetic surveys where the correct answer is known and on well-characterized real surveys. To evaluate the synthetic surveys, statistical tests are needed to compare the known and processed data. In this study the method of Universal Image Quality Index (UIQ, Wang and Bovik, 2002) is used in preference to the more commonly applied root mean square error (RMSE). As illustrated by Wang and Bovik (2002), many common problems in image quality, with quite different effects on image quality, are not recognised by use of RMSE.

Synthetic data

The synthetic survey for this study was prepared using nearly pure spectra of K, U and Th measured within three 200-litre calibration sources (Dickson and Taylor, 2003). A colored image was used to represent the K, U and Th distributions in the survey with the red-green-blue colours taken represent the different amounts. A stretch was applied so that Th/U ratios were in the range 3.5 to 5.5. For this study the image of Lena, which is widely used in the image-processing community, was chosen. Noise-free spectra were then calculated corresponding to the concentrations of K, U and Th for each pixel in the image. Finally, the value of each channel in each of these spectra was taken as the mean of a Poisson distribution and a single value was sampled from that distribution to replace the original value. In this way Poisson noise is added to each channel, simulating the dominant noise observed gamma-ray spectra. The spectra can then be processed to obtain the K, U and Th concentrations from the spectra using standard three-window processing. In this case no corrections are required for background, cosmic radiation, airborne radon or height variations, all additional sources to noise in real surveys. In applying the NLDR method, 211 channels from 0.30 to 3.0 MeV were used. The NLDR treatment used 3 dimensions and 16 mixture components

The original image and that derived from processing of the noisy data are shown in Figure 3. Values of UIQ for images calculated from the noisy data relative to the original, noise-free data, are listed in Table 1. (We choose to compare our results to those obtained using NASVD as this method is currently the most widespread for noise reduction and has been endorsed in an IAEA Technical Report (IAEA, 2003).

Table 1: UIQ values for Lena test images

	Noisy Image	NASVD	NLDR
K	0.52	0.60	0.62
U	0.23	0.62	0.78
Th	0.62	0.85	0.88
Th/U	0.00	0.03	0.16
K/Th	0.26	0.40	0.41

The ternary rgb images obtained by either NASVD or NLDR appear very similar which is expected. The two dominant colors in these rgb images are red and green (K and Th) for which the UIQ values show are similar by either method. The major difference is seen with U (blue) but the eye is less sensitive to blue and so the changes in noise level are not recognized. (The choice of blue for U was not accidental!).

Ratio images are used in exploration and geological mapping to enhance the differences between the three radioelements. Images of Th/U and K/Th ratios are also shown in Figure 3. As is expected there are only small improvements with NLDR over NASVD for images of the K/Th ratio. However, marked differences are seen in the Th/U ratio image. This ratio (or its inverse) is particularly useful for uranium exploration. The Th/U image for noisy data and standard processing shows no discernible changes (Figure 3) which is reflected in the UIQ for this image (0.0). There is no correlation with the original ratio image. Processing of the noisy data by NASVD and NLDR gave marked decreases in the level of noise as indicated in the UIQ vales but the NSAVD image also shows no discernible variation. This feature of NASVD has been seen with other test data sets (Dickson and Taylor, 2003). On the other hand the NLDR ratio image reveals the major changes in the ratio, albeit at a much reduced contrast scale that in the original data. Similar improvements for the Th/U ratio have been previously reported for the eMNF (Dickson and Taylor, 1998) and self-organizing maps (Dickson and Taylor, 2003) methods.

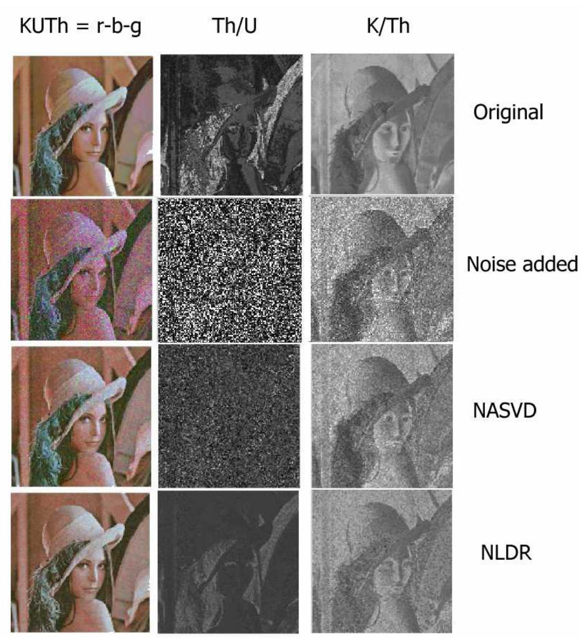

Figure 3: Effect of NASVD and NLDR denoising on the synthetic data derived from the Lena image.

REAL DATA EXAMPLE

An aerial survey conducted south of Sydney, NSW, over the Dendrobium colliery was reprocessed to evaluate the effectiveness of the NLDR method on real data. The survey used 16 liter crystal volume mounted in a helicopter. The survey was flown with 25 m flight line spacing at a nominal ground clearance of 40 m. One second counting intervals were used which is equivalent to 34 m of forward travel (122 km/h). Total coverage of the survey area amounted to 4,462 line kilometers. The survey was designed to optimize the measurement of the magnetic response from igneous intrusions and structures that could influence coal future mining.

The area is comprised of sandstone of the Sydney Basin. The terrain over the area is moderately steep, with some heavily forested areas with tall trees. A feature of the survey area is deeply incised gorges, one of which is flooded as part of a water storage dam, a track along a pipeline and a row of large power lines which traversed the area, all of which are oriented approximately north-south.

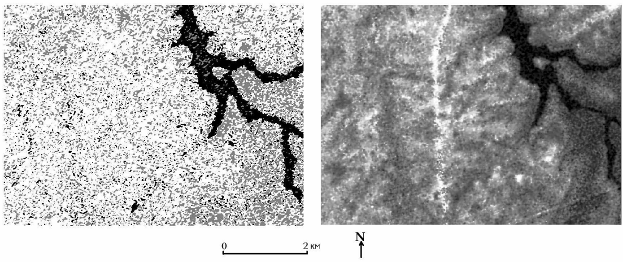

Figure 4: Images of Th/U ratio after denoising with NASVD (left) and NLDR (right).

For the reprocessing exercise, 211 channels of data (from 0.40 to 2.90 MeV) were extracted. After NLDR treatment using 3 dimensions and 16 mixture components, the spectra were processed, using conventional methods (IAEA, 2003) and calibration data supplied by the contractor, to extract the corrected window counts for K, U and Th. These counts were then gridded using a minimum curvature procedure with a grid cell size of 15m.

The most interesting result was found with the Th/U ratio image (Figure 4). Mirroring the result obtained with the Lena synthetic survey, the Th/U ratio calculated after NASVD processing shows no structure, apart from the lake which appears black, whereas that obtained after NLDR processing shows many variations including a north-south line of higher ratios. This line is coincident with the power-line that traverses the area and appears to result from the helicopter rising to a height of 80 m from an average elevation of 40 m to clear the line. This change in the Th/U ratio indicates a subtle change in U and Th as no such trend is seen in the individual U or Th images or in the ternary images of the area. Possibly the air contains a low concentration of aerial radon which contributed more to the signal at 80 m than at 40 m.

CONCLUSIONS

This study has demonstrated that NLDR processing can generated better data for the U and Th channels producing, in particular, more accurate Th/U ratio images. This has been demonstrated with both synthetic and real data. Although in the Dendrobium data shown, the feature revealed is of no exploration interest, the use of Th/U images is a major tool in the exploration for uranium and gold and for revealing fine detail in mapping geology, particularly around igneous bodies. The NLDR method thus should be of major benefit to users of aerial gamma-ray survey data.

ACKNOWLEDGEMENTS

This work is partially supported by the ARC Centres of Excellence programme funded by the Australian Research Council (ARC) and the New South Wales State Government. The authors would like to thank BHPB Illawarra Coal Holdings for providing the gamma-ray survey of the Dendrobium area.

REFERENCES

Belkin, M. and P. Niyogi, 2003.Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation, 15, 1373–1396.

Carey, R. and M. Craig, 2004. Cleaning scattered multi-channel data. Exploration Geophysics, 35, 131–136.

Dempster, A. P., Laird, N. M. and D. B. Rubin, 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society B. 39, 1-37.

Dickson, B. L. and G.M. Taylor, 1998. Noise reduction of aerial gamma-ray surveys. Exploration Geophysics, 29, 324–329.

Dickson, B. L. and G. M. Taylor, 2003. Quietening the noise: an evaluation of noise reduction methods applied to aerial gamma-ray survey data. Exploration Geophysics, 34, 97–102.

Green, A. A., Berman, M., Switzer, P. and M. D. Craig, 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. IEEE Transactions on Geoscience and Remote Sensing, 26, 65–74.

Hovgaard J. and R.L. Grasty, 1997. Reducing statistical noise in airborne gamma-ray data through spectral component analysis. in Gubins, A. G., ed., Proceedings of Exploration 97, Fourth Decennial International Conference on Mineral Exploration, 753–764.

IAEA, 2003. Guidelines for radioelement mapping using gamma ray spectrometry data. Technical Report IAEA TECDOC-1363, International Atomic Energy Agency, Vienna, Austria,.

Kohonen T., 2001. Self-organizing maps. Series in Information Science, 30. Springer-Verlag New York, Inc

Kumar, S., Ramos, F. T., Upcroft, B. and H. Durrant-Whyte. 2005. A Statistical Framework for Natural Features Representation. in Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Canada,

Minty, B. and P. McFadden, 1998. Improved NASVD smoothing of airborne gamma-ray spectra. Exploration Geophysics, 29, 516–523.

Roweis, S. T. and . L. K. Saul, 2000. Nonlinear dimensionality reduction by locally linear embedding. Science, 290, 2323–2326.

Switzer, P. and A. A. Green, 1984. Min/max autocorrelation factors for multivariate spatial imagery. Technical Report Technical Report 6, Department of Statistics, Stanford University, Stanford, USA,.

Tenenbaum, J., DeSilva, V. and J. C. Langford. 2000. A global geometric framework for nonlinear dimensionality reduction. Science, 290, 2319–2323.

Wang, Z. and A. C. Bovik, 2002. A universal image quality index. IEEE Signal Processing Letters, 9, 81–84.

Weinberger,K. Q., Packer, B. D. and L. K. Saul, 2005. Nonlinear dimensionality reduction by semidefinite programming and kernel matrix. in Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AISTATS), 381–388, Barbados, West Indies.