

# Enhancing Dust Storm Detection Using PCA based Data Fusion

Hesham El-Askary<sup>1</sup>, Abhishek Agarwal<sup>2</sup>, Tarek El-Ghazawi<sup>2</sup>, Menas Kafatos<sup>1</sup>, and Jacqueline Le-Moigne<sup>3</sup>

<sup>1</sup>Center for Earth Observing and Space Research (CEOSR), George Mason University (GMU)

<sup>2</sup>Department of Electrical and Computer Engineering, George Washington University (GWU)

<sup>3</sup>NASA Goddard Space Flight Center, Greenbelt, MD

[helaskar@gmu.edu](mailto:helaskar@gmu.edu), [agarwala@gwu.edu](mailto:agarwala@gwu.edu), [tarek@gwu.edu](mailto:tarek@gwu.edu), [mkafatos@gmu.edu](mailto:mkafatos@gmu.edu), [lemoigne@cesdis.gsfc.nasa.gov](mailto:lemoigne@cesdis.gsfc.nasa.gov)

**Abstract**— Principal Component Analysis (PCA) has been widely used as a data reduction technique to overcome the curse of dimensionality. In this research we show a different use for PCA technique as a tool for data fusion. PCA as a data fusion technique is performed over the Multiangle Imaging Spectroradiometer (MISR) data, studying dust storms to better serve their identification. The multi-angle viewing capability of MISR is used to enhance our understanding of the Earth's environment that includes climate particularly of atmosphere and of land surfaces. In this research the multi angle MISR images clearly show a dust storm over the Liaoning region of China as well as parts of northern and western Korea on April 8, 2002. PCA is used to combine the obtained information from the different angle views and frequency bands of MISR datasets. Performing K-means clustering on the original and the assimilated products apply a quantitative measure that is introduced. Upon classifying the first 4 principal components (PCs) having 95% of the information content similar results were obtained as compared to the classification using original datasets.

**Keywords**-PCA; Data Fusion; MISR; Dust Storms.

## I. INTRODUCTION

Dust storms are a significant air pollution contributor impacting urban areas as well as a health hazard for people with respiratory problems [1], [2]. Hence, timely warnings of dust storms must be initiated in populated regions for health concerns and traffic control. Storms can travel over large parts of the Earth, in Asia, Africa, affecting even North America and Europe. As an environmental related phenomenon, dust storms have increased in East Asia regions over the last decade. Such increase is attributed to the massive deforestation and increased droughts. As a climate related phenomenon, dust storms participate in modifying the energy budget through cooling and heating the atmosphere [3], [4], [5]. Dust storm detection and tracking could be difficult as they share some similar characteristics to clouds. Dust storms can vary in their shape, particle size, and distribution; hence normally show a varying behavior.

PCA is a linear transformation of a multivariate dataset into a new coordinate system. In remote sensing applications, the multiple variables are typically the different bands of a multispectral or hyperspectral data. PCA have been

widely used as a dimension reduction technique. It has the ability to reduce the dimensionality of a dataset while retaining most of the variance by concentrating the majority of the information into the first few components [6], [7], [8]. In this research we have used PCA as a data fusion tool of the multi angular based MISR observations. This technique is demonstrated on a large dust plume, which was observed over Liaoning region of China on April 8, 2002, as a case study shown in Fig. 1

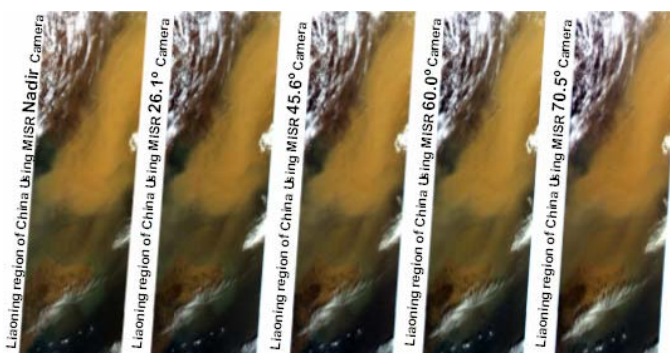


Figure 1. Different angle views of a large dust plume on April 8, 2002 over Liaoning region of China, parts of northern and western Korea

These multi-angle measurements can provide more information than traditional single angle remote sensing measurements, thus can enhance the fine discrimination between different materials.

In that respect we see our fusion method as a tool of combining two or more different images to form a new image, which aims at obtaining information of greater quality. Data fusion allows formalizing the combination of this information, as well as it monitors the quality of information in the course of the fusion process [9], [10]. Data fusion usually takes place at three different levels: pixel, feature, and decision [11], [12]. In this work, we are using the feature information revealed by the dust event for the fusion process. The feature refers to the GIS which helps in classifying multispectral images provided by several sensors, in our case cameras.

Using the MISR capability of different viewing angles, identification of dust storms can be greatly improved

# Report Documentation Page

*Form Approved*  
*OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE <b>25 JUL 2005</b>	2. REPORT TYPE <b>N/A</b>	3. DATES COVERED <b>-</b>			
4. TITLE AND SUBTITLE <b>Enhancing Dust Storm Detection Using PCA based Data Fusion</b>		5a. CONTRACT NUMBER			
		5b. GRANT NUMBER			
		5c. PROGRAM ELEMENT NUMBER			
6. AUTHOR(S)		5d. PROJECT NUMBER			
		5e. TASK NUMBER			
		5f. WORK UNIT NUMBER			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>Center for Earth Observing and Space Research (CEOSR), George Mason University (GMU)</b>		8. PERFORMING ORGANIZATION REPORT NUMBER			
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)			
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)			
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release, distribution unlimited</b>					
13. SUPPLEMENTARY NOTES <b>See also ADM001850, 2005 IEEE International Geoscience and Remote Sensing Symposium Proceedings (25th) (IGARSS 2005) Held in Seoul, Korea on 25-29 July 2005. , The original document contains color images.</b>					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>UU</b>	18. NUMBER OF PAGES <b>4</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

[13], [14]. For example, dust storm events, which are difficult to be detected by nadir viewing may be easily detected by off-nadir, angle views, because off-nadir sensors view thicker depth of the atmosphere. MISR has the potential to enhance the detection of small dust storms, thus it might be helpful in early detection of dust storms. It has been shown that MISR can be used to detect large dust storms like the one over northwestern part of India [13], [14].

In the current research we are focusing on integrating the spatial and the spectral information rendered by the different viewing angles and the different frequencies. PCA serves as the data fusion tool for such data integration. The used data is comprised of 4 different frequencies being served by 5 different angular observations. The data fusion is performed using PCA in two different ways, once by fixing the frequency component and once by fixing the angular component. Combining such information could be useful in discriminating between dust clouds and regular clouds. It could be beneficial in decreasing the background effects for desert regions by selecting suitable viewing angles as well.

## II. ANALYSIS AND DISCUSSION

Using PCA for data fusion we performed three different experiments using various angular and frequency combinations. Quantitative analysis is carried out for each experiment output to compare the obtained classes with those obtained from the original dataset. The computational complexity of these processes is not the theme of our current research. However, it will be discussed thoroughly in future publication.

### A. Experimental Layout

In the first experiment, see Fig.2, we used all the bands rendered by the different angle images in form of a multi dimensional product with 20 bands and performed the PCA. The first eigenvalue out of the 20 obtained components contain about 93.9 percent of the total data variation (information content). The classification results were produced using the first few eigenvalues for the representative set.

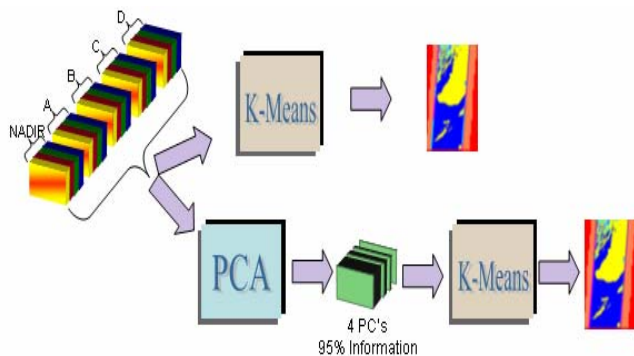


Figure 2. Experiment one layout showing the methodology used for comparing classification results from PCA based fused data to Original data

In the second experiment we have first fused information from different angles for one particular frequency. The first eigenvalue of each frequency contains approximately 94.2 percent of the total data variance. These first components with information from different cameras for each frequency are then fused as shown in Fig. 3, to produce a result comparable to first experiment.

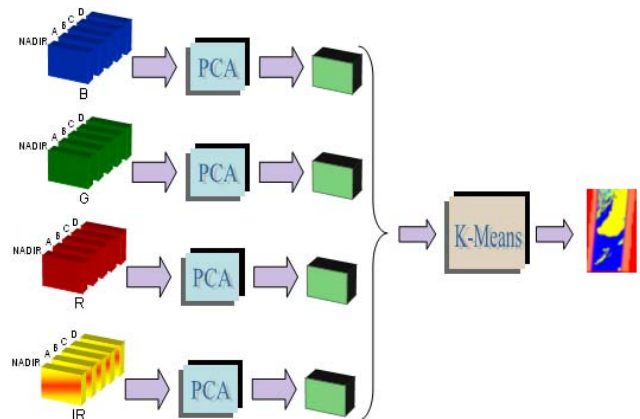


Figure 3. Experiment two layout showing the frequency based methodology used for comparing classification results from PCA based fused data to Original data.

In the third experiment we have first fused information from different frequencies for each individual camera. The first eigenvalue of each camera contains approximately 99 percent of the total data variance. These first components with information from different frequencies for each camera are then fused as shown in Fig. 4, to produce a result comparable to first and second experiments.

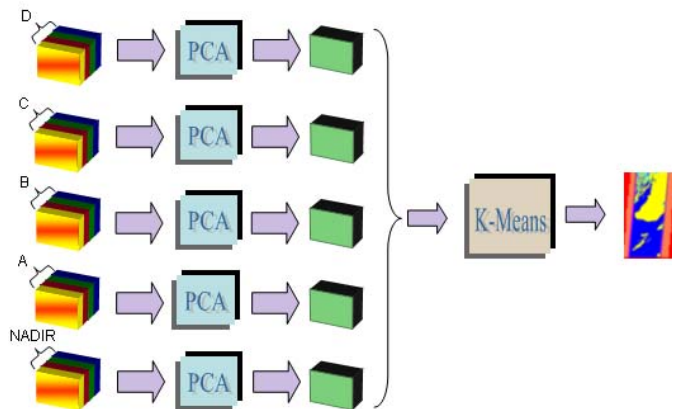


Figure 4. Experiment three layout showing the camera based methodology used for comparing classification results from PCA based fused data to Original data.

From the above three layouts, MISR served as the one sensor that can provide different spatial information content over the visible and IR spectrum. This unique property is revealed from the fact of availability of the multi-camera observations.

### B. Principal Components

The principal component computation involves computing the covariance matrix of the data, computing its eigenvalues and eigenvectors, plus additional work to form the principal components. In PCA, the components can be arranged or produced in a descending order of variance or information content. Most of the information contained by the data can be found in the first few principal components. Table I. below shows the percent information contained by the first principal components obtained from the performed experiments.

TABLE I. Comparison of the First Principal Component's Information Content obtained from the three experiments.

Experiments	Eigen Values X 10 <sup>9</sup>	Sum Eigen Values X 10 <sup>9</sup>	Percent Information Content	Percent Variation	
<b>ALL Cam+Freq</b>	10.732	11.423	0.939	<b>93.954</b>	
<b>Frequency Based</b>	<b>IR</b>	2.574	2.731	0.942	<b>94.255</b>
	<b>Red</b>	2.657	2.822	0.941	<b>94.155</b>
	<b>Green</b>	2.758	2.921	0.944	<b>94.423</b>
	<b>Blue</b>	2.789	2.948	0.946	<b>94.613</b>
<b>Camera Based</b>	<b>NADIR</b>	2.698	2.706	0.996	<b>99.694</b>
	<b>Camera A</b>	2.253	2.269	0.993	<b>99.320</b>
	<b>Camera B</b>	2.179	2.199	0.990	<b>99.092</b>
	<b>Camera C</b>	2.127	2.149	0.989	<b>98.987</b>
	<b>Camera D</b>	2.076	2.098	0.989	<b>98.914</b>

It is shown that the first eigenvalue for each experiment contains more than 90 percent of the total data variation (i.e. of the information) present in the original data set. With this, we can say that the intrinsic dimensionality of this dataset is effectively 1, given that we are only interested in 90 percent of the information content (data variation). The rest of the PCs contain total data variation of less than 10 percent of all original bands.

The percent variation in the camera based experiment is much higher in value than those obtained from the first two experiments. This can be attributed to the fact that PCA technique shows higher efficiency on spectral domain as compared to spatial domain. The first experiment showed the least percent variation as expected, since it tries fusing data from spatial and spectral domains and hence, there is no specific theme for the fusion process.

### C. K-means Clustering

To examine the three PCA based fusion layouts described above, a quantitative analysis is required. In our case, we use k-means clustering to compare the three fused outputs with the original data. K-means algorithm makes use of the spatial and the spectral information from the image under study. It creates clusters of discrete classes having sets of similar objects. A good clustering result is the one where

the objects in the same class are more or less alike, and objects in different classes are in some sense different [15].

This technique was performed four times, on the original data and with the three other experiments. Each time we run that algorithm, the same settings are used as shown in table II.

TABLE II. K-means classification settings used for the three experiments.

<b>Number of classes</b>	5
<b>Change threshold</b>	8%
<b>Maximum iteration</b>	15

On performing clustering of the original and the assimilated data products, similar classification results were observed as expected shown in Fig. 5.

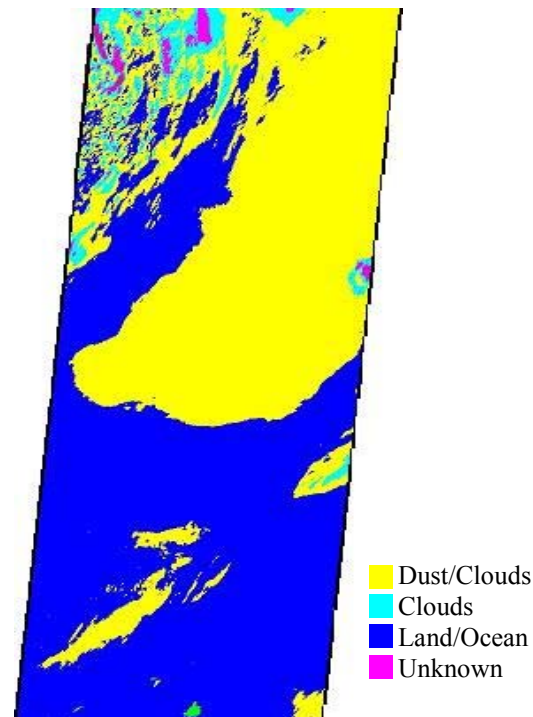


Figure 5. Kmeans clustering obtained from the original and the PCA fused data from the three experiments.

Such result coherence manifests the fact the PCA has well performed as a fusion tool. This is because of the fact that results that can be obtained from the original dataset with a certain dimensionality  $x$  can also be obtained from the assimilated out puts with dimensionality  $y$ , where  $x > y$ . The main differences in the outputs will be discussed in a future research dealing with computational and time complexity. It is also worthwhile to notice that clustering revealed some limitations. This is due to the fact it is not able to cluster well the original dataset having 100% of the information as compared to the PCA outputs having from 93-99 % of the information. However, in the latter case it is way faster than using the whole original dataset as will be discussed later.

### III. CONCLUSIONS

In this research, we introduced PCA as a tool for data fusion. It has been shown from the different layouts that PCA can fuse more information based on the way it is performed. The First PC from the camera based gave the highest information since it tries to fuse spectral information. Whereas, the first PC from the frequency based fuses spatial information and hence showed lower variance. In the case where fusion was done on all the bands, the least values were obtained. The classification results show that the three cases produce similar classification accuracy for dust storm event. Hence, PCA can be used as an effective method to fuse information. Using PCA in parts helps in better fusing the information from various data sources.

### IV. FUTURE WORK

Since PCA is a global operation it will be worthwhile to study the computational efficiency and time complexity of the above mentioned experimental layouts. It will be also good to investigate the obtained clustering outputs corresponding to each PCA experimental design. This analysis will reveal the most optimum layout for using PCA as a data fusion tool. Looking at the other obtained PCs besides the first few can help us to discriminate the dust events from the surrounding noisy elements, like clouds.

### REFERENCES

- [1] J. M. Prospero, 1999. "Long-range transport of mineral dust in the global atmosphere: Impact of African dust on the environment of the Southeastern United States," *Proceedings of the National Academy of Science* 96:3,396-3,403.
- [2] H. El-Askary, M. Kafatos, and M. Hegazy, "Environmental Monitoring Of Dust Storms Over The Nile Delta, Egypt Using MODIS Satellite Data," *Proceedings Of Third International Symposium Remote Sensing of Urban Areas, Istanbul, Turkey*, June 2002, pp. 452.
- [3] B. G. Liepert, "Observed reductions of surface solar radiation at sites in the United States and worldwide from 1961 to 1990," *Geophysical Research letter*, 29(10), May 2002.
- [4] S. Dey, S. N. Tripathi, R. P. Singh, and B. Holben, "Influence of Dust Storms on the Aerosol Optical Properties over the Indo-Gangetic basin.," *J. Geophys. Res.* 109, D20211, doi:10.1029/2004JD004924, 2004.
- [5] R.B. Husar, D.M. Tratt, B.A. Schichtel, S.R. Falke, F. Li, D. Jaffe, S. Gasso, T. Gill, N.S. Laulainen, F. Lu, M.C. Reheis, Y. Chun, D. Westphal, B.N. Holben, C. Gueymard, I. McKendry, N. Kuring, G.C. Feldman, C. McClain, R.J. Frouin, J. Merrill, D. DuBois, F. Vignola, T. Murayama, S. Nickovic, W.E. Wilson, K. Sassen, N. Sugimoto, and W.C. Malm. "Asian dust events of April 1998," *Journal of Geophysical Research*, vol 106, pp. 18,317-18,330.
- [6] I. Joliffe, *Principal Component Analysis*, Springer, Berlin, 1986
- [7] S. Kaewpijit, J. Le Moigne, and T. El-Ghazawi, "Hyperspectral Imagery Dimension Reduction Using Principal Component Analysis on the HIVE," *Science Data Processing Workshop, NASA Goddard Space Flight Center*, Feb. 2002.
- [8] L. Sunghyun, K. H. Sohn, and C. Lee, "Principal component analysis for compression of hyperspectral images," *Geoscience and Remote Sensing Symposium, IGARSS '01. IEEE 2001 International vol 1*, pp. 97 - 99, July 2001.
- [9] L. Wald, "A European proposal for terms of reference in data fusion," *International Archives of Photogrammetry and Remote Sensing*, vol. XXXII, Part 7, 651-654, 1998.
- [10] L. Wald , "Some terms of Reference in Data Fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol37, No. 3, pp 1190 -1193, May1999.
- [11] C. Pohl, and J.L. van Genderen, "Multisensor image fusion in remote sensing: Concepts, methods and applications," *Int. J. Remote Sensing*, vol. 19, no5, pp 823-854,1998.
- [12] G. Brandstatter, and A. Sharov, "Environmental monitoring in the high Arctic using different types of high-resolution satellite imagery," *Int. Arcg. Photogramm. Remote Sensing*, pt. 7, vol.32, pp201-216, 1998.
- [13] H. El-Askary, R. Gautam and M. Kafatos, "Remote Sensing of Dust Storms over the Indo-Gangetic Basin," *Indian Journal or Remote Sensing*, 32(2): 121-124, 2004.
- [14] H. El-Askary, R. Gautam, R. P. Singh, and M. Kafatos, "Dust Storms detection over the Indo-Gangetic Basin using Multi Sensor Data," *Advances in Space Research Journal*, 2005. (In press)
- [15] T. M. Lillesand, and R. W. Kiefer, "Remote Sensing and Image Interpretation," John Wiley & Sons, Inc., New York, NY, 1994.