



ISRN TELCORDIA--2012-08+PR-0GARAU
Prepared for: Office of Naval Research

Fast Multiscale Algorithms for Information Representation and Fusion

Technical Progress Report No. 8

Devasis Bassu, Principal Investigator

Contract: N00014-10-C-0176

Applied Communication Sciences

One Telcordia Drive

Piscataway, NJ 08854-4157

July 2012

Approved for public release; distribution is unlimited.

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 JUL 2012		2. REPORT TYPE		3. DATES COVERED 00-00-2012 to 00-00-2012	
4. TITLE AND SUBTITLE Fast Multiscale Algorithms for Information Representation and Fusion				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) Applied Communication Sciences,One Telcordia Drive,Piscataway,NJ,08854				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 Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 13	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

1 Abstract

In the eighth quarter of the work effort, we focused on a) conducting experiments on real-world data sets using the developed algorithms, b) continued design/implementation of the Multiscale Singular Value Decomposition (SVD) algorithm and c) packaging for releasing the software as open source. This report documents experimental results with the Multiscale SVD algorithms.

The project is currently on track – in the upcoming quarter, we will continue applying the developed algorithms to various data sets and the design/implementation of the multiscale heat kernel coordinates algorithms. No problems are currently anticipated.

Table of Contents

1	ABSTRACT	II
2	SUMMARY	1
3	INTRODUCTION	2
4	METHODS, ASSUMPTIONS AND PROCEDURES	3
4.1	Multiscale Singular Value Decomposition	3
4.2	Deliverables / Milestones.....	3
5	RESULTS AND DISCUSSION	4
5.1	Experiment: LIDAR Dataset (MSVD using nearest neighbors)	4
6	CONCLUSIONS	8
7	REFERENCES.....	9

2 Summary

In this quarter, we continued design and implementation of the new multiscale SVD (MSVD) algorithms. We applied the MSVD to a publicly available LIDAR dataset for the purposes of distinguishing between vegetation and the forest floor. The final results are presented in this report (initial results were reported in the previous quarterly report).

The project is currently on track – in the upcoming quarters, we will continue applying the developed algorithms to various data sets and focus on the design and development of the multiscale heat kernel coordinates algorithms. No problems are currently anticipated.

3 Introduction

The primary project effort over the last quarter focused on completing design/development of the multiscale SVD algorithms [1]. Results from experiments conducted on a publicly available LIDAR dataset [5] are provided in Section 5.

4 Methods, Assumptions and Procedures

4.1 Multiscale Singular Value Decomposition

The Multiscale Singular Value Decomposition (MSVD) was introduced in the earlier technical reports [6][7]. The MSVD provides a spectral readout of the dataset at all scales.

We applied the MSVD algorithm to a real-world LIDAR data set. Experimental setup and initial results were reports section 5.2 of the previous quarterly report [7]. We present the full results of that experiment in this report.

4.2 Deliverables / Milestones

Date	Deliverables / Milestones	Status
Oct 2010	Progress report for period 1, 1 st quarter	✓
Jan 2011	Progress report for period 1, 2 nd quarter / complete randomized matrix decompositions task	✓
Apr 2011	Progress report for period 1, 3 rd quarter / complete approximate nearest neighbors task	✓
Jul 2011	Progress report for period 1, 4 th quarter / complete experiments – part 1	✓
Oct 2011	Progress report for period 2, 1 st quarter	✓
Jan 2012	Progress report for period 2, 2 nd quarter / complete multiscale SVD task	✓
Apr 2012	Progress report for period 2, 3 rd quarter	✓
Jul 2012	Progress report for period 2, 4 th quarter / complete experiments – part 2	✓
Oct 2012	Progress report for period 3, 1 st quarter	
Jan 2013	Progress report for period 3, 2 nd quarter / complete multiscale Heat Kernel task	
Apr 2013	Progress report for period 3, 3 rd quarter	
Jul 2013	Final project report + software + documentation on CDROM / complete experiments – part 3	

5 Results and Discussion

We present experimental results for the LIDAR dataset below.

5.1 Experiment: LIDAR Dataset (MSVD using nearest neighbors)

This publicly available dataset [5] contains 3-dimensional LIDAR point data clouds representing ten sections of riparian floor and vegetation. An analysis of the dataset for classification purposes is presented in [4]. The dataset comprises 639,520 data points, each categorized as floor or vegetation. The dataset is depicted in Figure 1.

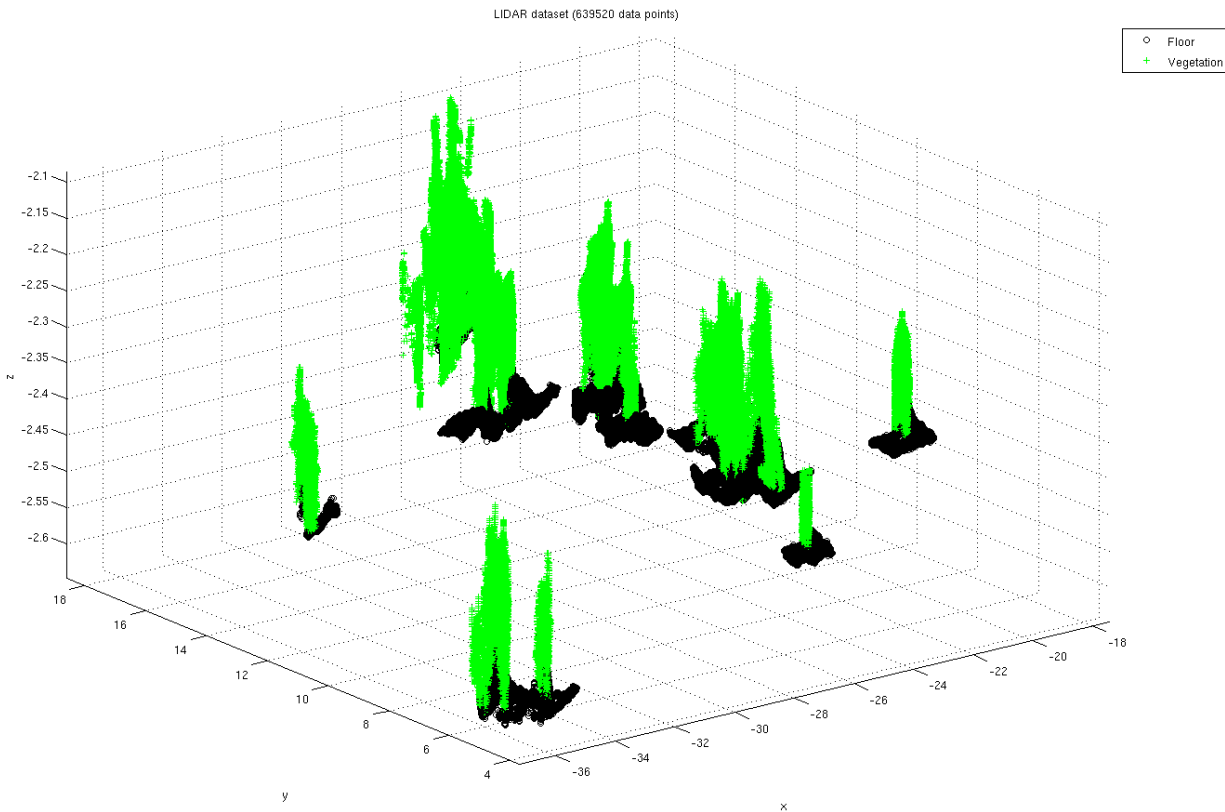


Figure 1. Example 2: LIDAR dataset

Sensitivity and specificity measures are used to provide metrics for the classification task. These are statistical measures used for measuring performance of binary classification tests and are akin to Type I and Type II errors. Sensitivity measures the proportion of actual positives which are correctly identified as such. Specificity measures the proportion of negatives which are correctly identified. The classification accuracy reported in the paper [4] is 95% as $\min\{\text{sensitivity}, \text{specificity}\}$.

The MSVD algorithms were used for the classification task of binning each point as floor or vegetation. For comparison purposes, the same training and test data sets used in reference paper. A SVM classifier (also used in reference paper) was used to classify the MSVD features. Continuous and 10-discretization pre-processing options for the SVM were considered. For 10-

discretization, all coordinates were discretized by pushing data to centers of 10 equal-integral areas of normal (0,1) distribution in addition to usual coordinate scaling (see Figure 2). The exact NN algorithm was used to compute neighbors at scales 9, 10 and 11. Scale 9 provided roughly around 100 to 200 points in each ball. It should be pointed out that the reference paper takes deep specifics of the problem (relative dimensions of leaves, branches and soil) into account. In contrast, our approach using the MSVD employs a generic methodology for data analysis.

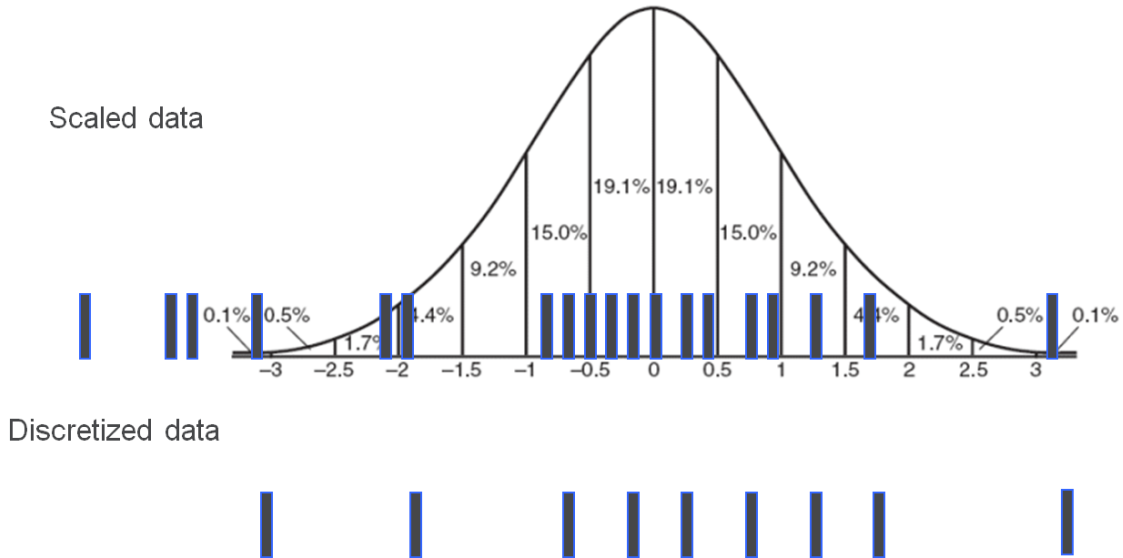


Figure 2. 10-discretization of scaled features

The original features from the dataset are simply the 3-dimensional spatial coordinates. Applying the MSVD algorithm to the dataset, we obtain the following derived features corresponding to the singular values and vectors at scales 9, 10 and 11. The dimension of our derived feature vector is 36 (3-singular values and 3 3-dimensional singular vectors at each scale).

σ^9	v_1^9	v_2^9	v_3^9	σ^{10}	v_1^{10}	v_2^{10}	v_3^{10}	σ^{11}	v_1^{11}	v_2^{11}	v_3^{11}
------------	---------	---------	---------	---------------	------------	------------	------------	---------------	------------	------------	------------

We defined an experimental design to test our intuitive understanding that multiple scales most likely will provide better results than any single scale. The experimental design is shown in Figure 3.

Coordinate (x,y,z)		Scales			
included	excluded	9	10	11	9,10,11
Y		Y			
Y		Y			
	Y	Y			
	Y	Y			
Y			Y		
Y			Y		
	Y		Y		
	Y		Y		
Y				Y	
Y				Y	
	Y			Y	
	Y			Y	
Y					Y
Y					Y
	Y				Y
	Y				Y

Figure 3. Experimental design for LIDAR dataset

The final results in terms of specificity and sensitivity for this above design are tabulated in Figure 4. Using the MSVD features reduces classification error to 2% (from 5% in paper). More importantly, as expected the combination of localized scales works better than any single scale. Further, scale 11 has lots of “empty” balls which explain the poor accuracy.

SVM Preprocessing		Coordinate (x,y,z)		Scales				Classification Metrics		
Continuous	10-discretization	included	excluded	9	10	11	9,10,11	Sensitivity %	Specificity %	Error %
Y		Y		Y				98	93	7
	Y	Y		Y				98	96	4
Y			Y	Y				95	95	5
	Y		Y	Y				95	95	5
Y		Y			Y			98	93	7
	Y	Y			Y			95	98	5
Y			Y		Y			95	98	5
	Y		Y		Y			92	97	8
Y		Y				Y		93	90	10
	Y	Y				Y		86	96	14
Y			Y			Y		97	80	20
	Y		Y			Y		84	77	23
Y		Y					Y	97	97	3
	Y	Y					Y	98	98	2
Y			Y				Y	97	98	3
	Y		Y				Y	97	97	3

Figure 4. Sensitivity/specificity results for the LIDAR dataset

6 Conclusions

The project is on track with wrapping up the multiscale SVD algorithms. A comprehensive paper on the MSVD is forthcoming. We presented the results of using these algorithms on the LIDAR dataset. We will focus on design/implementation of the new multiscale heat kernel coordinates algorithms and continue with algorithmic improvements and experimentation using the developed algorithms in the next quarter.

No problems are currently anticipated.

7 References

- [1] G. Lerman, *Quantifying curvelike structures of measures by using Jones quantities*, C.P.A.M., vol. 56, issue 8, pages 1294-1365.
- [2] G. H. Golub, W. Kahan, *Calculating the singular values and pseudo-inverse of a matrix*, Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis, vol. 2 (2), pages 205–224, 1965.
- [3] G.H. Golub, C.F. Van Loan, *Matrix Computations (3rd edition.)*, Johns Hopkins University Press, Baltimore, 1996.
- [4] N. Brodu, D. Lague, *3D Terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: applications in geomorphology*, version 3, 2012. URL: <http://arxiv.org/abs/1107.0550>
- [5] LIDAR dataset URL:
<http://nicolas.brodu.numerimoire.net/common/recherche/canupo/benchmark.tar.gz>
- [6] D.Bassu, *Fast Multiscale Algorithms for Information Representation and Fusion*, Technical Report No. 6, ISRN TELCORDIA--2011-06+PR-0GARAU, 2011.
- [7] D.Bassu, *Fast Multiscale Algorithms for Information Representation and Fusion*, Technical Report No. 7, ISRN TELCORDIA--2012-07+PR-0GARAU, 2012.