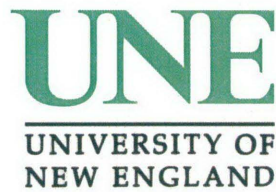


# **DIMENSIONALITY REDUCTION FOR NON-VECTORIAL DATA**

By

Yi Guo

B. Eng., M. Eng.



School of Science and Technology  
University of New England

A thesis submitted for the degree of

**PhD of Computer Science**

of University of New England

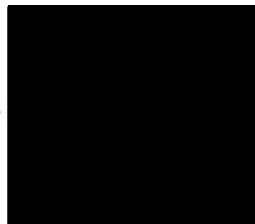
*January 2008*

## DECLARATION

---

*I certify that the substance of this thesis has not already been submitted for any degree and is not currently being submitted for any other degree.*

*I certify that to the best of my knowledge, any help received in preparing this thesis, and all sources used, have been acknowledged in this thesis.*



# Acknowledgements

I would like to thank my principal supervisor Associate Professor Junbin Gao without whom the work of the thesis would have been impossible. Though he left University of New England in July 2005, his support and supervision stayed with me over the past three years. We had thousands of emails and telephone calls with ideas, discussions, encouragement and so on. I really appreciate his insights into the area of machine learning, especially in statistical machine learning, sharing with me his abundant knowledge of mathematics and computer science. I am grateful for his generosity and his passion for education. Most of our research arose from his brilliant ideas. He brought me into this wonderful field and, with his help, my understanding and perception of dimensionality reduction problems have been gradually formed.

I would also like to thank my co-supervisor Dr Paul Kwan. I have benefited greatly from his instruction in the methodology of research, from his assistance in writing papers and for his generous support of my research and everyday living. He always encouraged me to present our latest results and created great opportunities for me to introduce our algorithms, so that I can now stand before a platform with confidence. Sincere thanks to Dr Xiaodi Huang and the mathematics people in our school, Professor Yihong Du, Dr. Shusen Yan and Dr. Adam Harris. The discussions on mathematical issues with them helped me understand the existing methods more clearly. Many thanks to Professor A. S. M. Sajeev, Mr Norman Gaywood and Mr Saint Ashley. They provided me with extra computational support so that the large numbers of experiments could be performed in parallel. I would like to thank Dr Neil D. Lawrence for sharing his GPLVM code, which was partly used in our algorithms.

The research was funded by the International Postgraduate Research Scholarship (IPRS) and

the University of New England Research Assistance (UNERA). I acknowledge the support from the research office. They made the research possible.

Finally, I want to thank sincerely my clear-sighted and strong-minded wife, Huiling Yang, who never took me away from my work and supported me all along. I am also very grateful to our parents and sisters. Their love always support me in completing my study.

# Publications

## Journal Papers

Yi Guo, Junbin Gao, and Paul W. Kwan. Visualization of protein structure relationships using constrained twin kernel embedding. *International Journal of Biomedical Science and Engineering*, submitted, 2008.

Yi Guo, Junbin Gao, and Paul W. Kwan. Twin kernel embedding. *IEEE Transaction of Pattern Analysis and Machine Intelligence*, to appear, 2008.

## Referenced conference proceedings

Yi Guo, Junbin Gao, and Paul W. Kwan. Twin Kernel Embedding with back constraints. In *HPDM in ICDM*, 2007.

Yi Guo, Junbin Gao, and Paul W. Kwan. Learning optimal kernel from distance metric in twin kernel embedding for dimensionality reduction and visualization of fingerprints. In *International Conference on Advanced Data Mining and Applications*, 2007.

Yi Guo, Junbin Gao, Paul W. Kwan, and Kevin X. Hou. Visualization of protein structure relationships using twin kernel embedding. In *International Conference on Bioinformatics and Biomedical Engineering*, pages 1–4.

Yi Guo, Junbin Gao, and Paul W. Kwan. Learning out-of-sample mapping in non-vectorial data dimensionality reduction using constrained twin kernel embedding. In *International Conference on Machine Learning and Cybernetics*, 2007.

Yi Guo, Junbin Gao, and Paul W. Kwan. Visualization of non-vectorial data using twin kernel embedding. In *First International Workshop on Integrating AI and Data Mining*, volume 0,

pages 11–17, 2006.

Yi Guo, Junbin Gao, and Paul W. Kwan. Kernel Laplacian eigenmaps for visualization of non-vectorial data. In *Lecture Notes on Artificial Intelligence*, volume 4304, pages 1179–1183, 2006.

Yi Guo and Junbin Gao. An integration of shape context and semigroup kernel in image classification. In *International Conference on Machine Learning and Cybernetics*, 2007.

Yi Guo and Junbin Gao. Manifolds of bag of pixels: A better representation for image recognition? In *2006 IEEE Conference on Systems, Man, and Cybernetics*, pages 3618–3622, Taipei, October 8-11 2006. IEEE, SMC society, IEEE.

# Abstract

Dimensionality Reduction (DR) is an important step in many advanced applications such as exploratory data analysis and manifold learning. Its main goal is to discover the mappings of the input data in a much lower dimensional space or the so-called latent space without incurring unnecessary information loss.

In most existing DR algorithms, the main objective is to preserve relational structure among objects of the input space in the latent space by minimizing the inconsistency between two similarity/dissimilarity measures, one for the input data and the other for the embedded data, via a separate matching objective function. Based on this observation, a new dimensionality reduction method called Twin Kernel Embedding (TKE) is proposed. TKE addresses the problem of embedding non-vectorial data that is difficult for conventional methods in practice due to the lack of efficient vectorial representation. TKE solves this problem by minimizing the inconsistency between the similarity measures captured respectively by their kernel Gram matrices in the two spaces. This algorithm is proven to be effective on some real world data sets and has been successfully applied to protein visualization, kernel learning, fingerprint classification etc. TKE is further extended to novel samples by introducing the backward mapping which is incorporated into the objective function as either substitution of all embeddings or regularization terms which generate BCTKE and RCTKE algorithms respectively. Intuitively, the mapping function can be integrated into any other host DR algorithms as a solution to the so-called out-of-sample problem.

This thesis starts with the analysis of the existing DR methods. Based on the understanding of their common features, we will show the development of the TKE algorithms and the details

on their behaviors at length. We present not only a series of new algorithms, but also the aspects of the design including the origin of the ideas, observations and implementation. This research provides a stepping stone for new algorithmic design in Dimensionality Reduction.



# Contents

<b>Acknowledgements</b>	<b>iii</b>
<b>Publications</b>	<b>v</b>
<b>Abstract</b>	<b>vii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Literature Review</b>	<b>6</b>
2.1 Overview . . . . .	6
2.2 Taxonomy of DR . . . . .	7
2.3 Principal Component Analysis . . . . .	9
2.4 Multi-Dimensional Scaling . . . . .	12
2.5 Laplacian Eigenmaps . . . . .	15
2.6 Isomap . . . . .	17
2.7 Locally Linear Embedding . . . . .	18
2.8 Hessian Locally Linear Embedding . . . . .	20
2.9 Local Tangent Space Alignment . . . . .	21
2.10 Gaussian Processes Latent Variable Model . . . . .	22
2.11 Kernel Principal Component Analysis . . . . .	23
2.12 Summary . . . . .	25

<b>3</b>	<b>Shape Context based IGV Kernel</b>	<b>27</b>
3.1	Introduction . . . . .	27
3.2	A brief review of shape context and semigroup kernel . . . . .	29
3.2.1	Shape distance based on shape context . . . . .	29
3.2.2	Semigroup kernels . . . . .	30
3.3	Integration . . . . .	31
3.4	Experimental Evaluation . . . . .	35
3.5	Summary . . . . .	38
<b>4</b>	<b>Kernel Laplacian Eigenmaps</b>	<b>41</b>
4.1	Introduction . . . . .	41
4.2	Kernel Laplacian Eigenmaps . . . . .	43
4.3	Experimental Evaluation . . . . .	46
4.4	Summary . . . . .	47
<b>5</b>	<b>Twin Kernel Embedding</b>	<b>51</b>
5.1	Introduction . . . . .	51
5.2	Measurement Matching Framework . . . . .	52
5.2.1	A Unified Framework . . . . .	52
5.2.2	The Matching Objective Functional . . . . .	54
5.3	Reinterpreting Several DR Methods . . . . .	55
5.3.1	MDS . . . . .	56
5.3.2	PCA and KPCA . . . . .	56
5.3.3	Laplacian Eigenmaps and KLE . . . . .	56
5.3.4	Isomap . . . . .	57
5.3.5	Locally Linear Embedding . . . . .	57
5.3.6	Gaussian Processes Latent Variable Model . . . . .	58
5.3.7	Brief Summary . . . . .	58
5.4	Twin Kernel Embedding . . . . .	59

5.4.1	Kernel on Embedded Data and Similarity Matching . . . . .	59
5.4.2	Optimization . . . . .	61
5.4.3	Implementation . . . . .	64
5.4.4	Some Remarks on TKE . . . . .	65
5.5	Experimental Results . . . . .	67
5.5.1	Parameters Setting . . . . .	67
5.5.2	Manifold learning with TKE . . . . .	67
5.5.3	Handwritten Digits Visualization . . . . .	68
5.5.4	Visualizing Texts . . . . .	70
5.5.5	Visualizing Proteins . . . . .	77
5.6	An Empirical Analysis on TKE Algorithm . . . . .	83
5.6.1	Model Selection . . . . .	83
5.6.2	Computational Cost . . . . .	86
5.6.3	The Influence of the Initialization . . . . .	89
5.6.4	Extensions of TKE . . . . .	90
5.7	Summary . . . . .	93
<b>6</b>	<b>Learning Optimal Kernel with TKE</b>	<b>95</b>
6.1	Biometric data and High Dimensionality Challenge . . . . .	95
6.2	Related Work . . . . .	97
6.3	From Distance to Kernel . . . . .	99
6.4	Experimental Results . . . . .	101
6.5	Summary . . . . .	105
<b>7</b>	<b>Twin Kernel Embedding with Back Constraints</b>	<b>107</b>
7.1	Out-of-sample Extension, Constraints and Locality Preserving . . . . .	107
7.2	Brief Review of Constraints in DR Algorithms . . . . .	110
7.2.1	Locality Preserving Projections . . . . .	110
7.2.2	Neighborhood Preserving Embedding . . . . .	110

7.2.3	GPLVM with Back Constraints . . . . .	111
7.3	Twin Kernel Embedding with Back Constraints . . . . .	113
7.4	Experimental Results . . . . .	116
7.4.1	MNIST Handwritten Digits . . . . .	116
7.4.2	Text Classification Comparison . . . . .	121
7.4.3	Image Manifold Learning . . . . .	122
7.4.4	Refinement Strategy for BCTKE . . . . .	123
7.5	Summary . . . . .	130
<b>8</b>	<b>Twin Kernel Embedding with Relaxed Constraint</b>	<b>131</b>
8.1	TKE with relaxed constraints . . . . .	132
8.1.1	Approach 1: Sum-of-square Error . . . . .	132
8.1.2	Approach 2: Incorporate LS-SVM objective function . . . . .	133
8.2	Experimental Results . . . . .	136
8.2.1	Parameters Setting . . . . .	136
8.2.2	Reuters Texts . . . . .	136
8.2.3	Proteins . . . . .	137
8.3	Summary . . . . .	138
<b>9</b>	<b>Examiner-Centric Learning Framework</b>	<b>140</b>
9.1	Introduction . . . . .	140
9.2	Overview Of Learning Framework . . . . .	142
9.3	Extraction Of Spectral Features . . . . .	143
9.3.1	Locate Reference Point . . . . .	145
9.3.2	Extract and Tessellate Region of Interest . . . . .	147
9.3.3	Spectral Decomposition using Gabor Filters . . . . .	147
9.3.4	Compute the FingerCode . . . . .	148
9.4	Learning Examiner-Centric Semantic Space . . . . .	148
9.4.1	Relevance Feedback . . . . .	149

9.4.2	Dimensionality Reduction . . . . .	149
9.5	Experimental Evaluation . . . . .	152
9.6	Summary . . . . .	154
<b>10</b>	<b>Conclusion</b>	<b>155</b>
	<b>Bibliography</b>	<b>176</b>
<b>A</b>	<b>Derivatives</b>	<b>177</b>
A.1	Derivatives in TKE . . . . .	177
A.2	Derivatives in BCTKE . . . . .	179
<b>B</b>	<b>Matlab Code for TKE and BCTKE</b>	<b>182</b>
B.1	Notes for the code . . . . .	182
B.2	tke.m . . . . .	182
B.3	FilteringKern.m . . . . .	185
B.4	adjgraph.m . . . . .	185
B.5	TKEOptions.m . . . . .	187
B.6	TKECreate.m . . . . .	188
B.7	TKEOptimise.m . . . . .	190
B.8	TKEExtractParam.m . . . . .	191
B.9	TKEExpandParam.m . . . . .	191
B.10	GenLabels.m . . . . .	192
B.11	knnrelabel.m . . . . .	193

# List of Tables

2.1	Notation conventions used in this thesis . . . . .	7
3.1	Comparison of error rate of different methods. SCIGV-K is the SCIGV kernel and IGV-K the IGV kernel. . . . .	38
5.1	Methods re-interpreted by the measurement matching framework. $\mathbf{D}_y = \{\ y_i - y_j\ ^2\}$ , $\mathbf{D}_x = \{\ x_i - x_j\ ^2\}$ . We use more general forms of PCA and LE where the kernel Gram matrix replaces the inner product matrix and weight matrix in PCA and LE respectively. . . . .	59
5.2	Comparison of 1NN classification errors of different algorithms. TKE with linear kernel outperforms others from this point of view. . . . .	70
7.1	Comparison of 1NN classification errors of different algorithms. BCTKE outperforms others from this point of view. . . . .	117
7.2	Comparison of 1NN classification errors of different algorithms in prediction of the embeddings for new data. BCTKE is the best and BCGP has comparable performance. . . . .	119
9.1	Pseudo code of relevance feedback procedure. . . . .	150
9.2	Comparison of He et al. (2004) [55]’s dimensionality reduction procedure and the one adopted in this research. . . . .	151
9.3	Comparing classification quality between baseline, semantic space in 2-dim before and after relevance feedback. . . . .	154

# List of Figures

1.1	An illustration of dimensionality reduction. 100 1's and 0's are plotted here (50 each) to show the relational structure of these digits. . . . .	2
3.1	The weighted list of components. . . . .	32
3.2	The performance of the IGV kernel on translation of images . . . . .	33
3.3	Part of training data . . . . .	36
3.4	The error rate of different $\alpha_1$ and $\alpha_2$ in $[0, 2] \times [0, 2]$ . . . . .	37
3.5	The Gram matrices. . . . .	39
4.1	The comparison of the performance of the KLE using SCIGV kernel and original LE. . . . .	48
4.2	The effects of the value of $n$ in KLE. . . . .	49
5.1	The mechanism behind measurement matching. $M_y$ and $M_x$ are similarity/dissimilarity measures for the input and the embeddings respectively. They are matched by the matching objective functional $C_o$ to ensure that the similarity/dissimilarity structure in $\mathcal{Y}$ is optimally preserved in $\mathcal{X}$ . . . . .	53
5.2	Comparing TKE with other algorithms on synthetic data (swiss roll) . . . . .	69
5.3	The result of different algorithms on MNIST handwritten digits database (part 1). . . . .	71
5.4	The result of different algorithms on MNIST handwritten digits database (part 2). . . . .	72
5.5	The comparison of the purity of different algorithms on different data sets. Horizontal axis is the size of the neighborhood and vertical axis is purity. . . . .	73
5.6	The results on text data. . . . .	76
5.7	Visualization results of protein structures . . . . .	79

5.8	The results of other methods using the decomposition of the MAMMOTH kernel	81
5.9	The result of TKE with different kernels . . . . .	82
5.10	The 1NN error rate against different regularization parameters. . . . .	85
5.11	The effect of different $n$ to TKE and KLE. . . . .	86
5.12	The convergence of the objective function. . . . .	88
5.13	The convergence of the 1NN error rate. . . . .	88
5.14	The effect of different initialization (part 1). . . . .	91
5.15	The effect of different initialization (part 2). . . . .	92
6.1	Fingerprint with minutiae marked. . . . .	97
6.2	Fingerprints of 10 different subjects . . . . .	101
6.3	Visualization of fingerprints using TKE and KLE. . . . .	103
6.4	Kernel Gram matrices . . . . .	104
6.5	SVM and Kmeans on the embedded data . . . . .	106
7.1	The best results of different DR algorithms (in the sense of minimum 1NN classification errors) on handwritten digits database. All the figures use the same symbols as shown in the legend of figure (b). Visually, the BCTKE, TKE and BCGP have the best cluster structure. . . . .	118
7.2	BCTKE in prediction and comparison with other methods. The results of different algorithms are organized in rows. From top to bottom, they are BCTKE, BCGP, LPP (with $n = 4$ ) and NPE (with $n = 5$ ). The columns are the results of different algorithms on different sets. They are training set, test set and the union of training and test set from left to right. All the figures share the same legend as shown in the figure at the top left corner. The results presented here are also optimized with respect to the 1NN classification errors on the union of training set and test set as discussed in the last sub-section. The 1NN classification errors are compared in Table 7.2. . . . .	120
7.3	The performance comparison of BCTKE and LPP on Reuters-21578 text classification. The mean of different $k$ NN classification error rates is plotted as lines and standard deviation as little bars. . . . .	122



7.4	The embeddings of faces in 2D latent space. . . . .	124
7.5	The results of the refined BCTKE on the first data set. All the figures use the same symbols as shown in the legend of figure (a). . . . .	126
7.6	The results of the refined BCTKE on the second data set with circles for <i>coffee</i> , pluses for <i>interest</i> and arrows for <i>corn</i> . . . . .	127
7.7	The results of other algorithms. (a), (b) are the TKE and KPCA on the first data set (treating training set and test set altogether, see (a) in figure 7.5 for symbols); (c), (d) are the TKE and KLE on the entire second data set using same symbols as figure 7.6. . . . .	128
8.1	The results on Reuters texts. . . . .	137
8.2	The result of TKE and KPCA with MAMOTH kernel . . . . .	138
8.3	The result of RCTKE with MAMOTH kernel . . . . .	139
9.1	The Learning Framework showing the major components, <i>Features Extraction</i> and <i>Semantic Space Learning</i> . Semantic Space Learning is in turn comprised by Relevance feedback and Dimensionality reduction modules. . . . .	144
9.2	Extraction of Spectral Features by multispectral decomposition using a bank of Gabor filters aligned in eight different directions including $0^\circ$ , $22.5^\circ$ , $45^\circ$ , $67.5^\circ$ , $90^\circ$ , $112.5^\circ$ , $135^\circ$ , $157.5^\circ$ . All images are shown in half their original sizes. . . . .	146
9.3	Semantic space embedded in two dimensions by TKE before and after 50, 100, 150 and 200 relevance feedback iterations. . . . .	153
A.1	The organization of the partial derivatives in $\mathbf{K}$ wrt. $\mathbf{X}$ . . . . .	179