RGC Ref. No.: UGC/FDS14/E03/14
(please insert ref. above)

RESEARCH GRANTS COUNCIL
COMPETITIVE RESEARCH FUNDING SCHEMES FOR THE LOCAL SELF-FINANCING DEGREE SECTOR

FACULTY DEVELOPMENT SCHEME (FDS)

Completion Report
(for completed projects only)

Submission Deadlines:
1. *Auditor's report with unspent balance, if any: within **six** months of the approved project completion date.*
2. *Completion report: within **12** months of the approved project completion date.*

Part A: The Project and Investigator(s)

1. Project Title

Data Functional Modelling with Outliers

2. Investigator(s) And Academic Department(s) / Unit(s) Involved

Research Team	Name / Post	Unit / Department / Institution
Principal Investigator	Dr. LAM Shu Yan/ Assistant Professor	Department of Mathematics and Statistics/ The Hang Seng University of Hong Kong
Co-Investigator(s)	Dr. CHOY Siu Kai/ Associate Professor	Department of Mathematics and Statistics/ The Hang Seng University of Hong Kong
Others		

3. Project Duration

	Original	Revised	Date of RGC / Institution Approval *(must be quoted)*
Project Start Date	01/01/2016		
Project Completion Date	31/12/2017	30/6/2018	3/10/2017
Duration *(in month)*	24	30	3/10/2017
Deadline for Submission of Completion Report	31/12/2018	30/6/2019	3/10/2017

Part B: The Final Report

5. Project Objectives

5.1 Objectives as per original application

1. Develop a new theoretical framework by incorporating the theories of clustering into PCA.
2. Develop efficient implementations for the new theoretical framework.
3. Conduct extensive experimental testing of the proposed framework for various applications.

5.2 Revised objectives

Date of approval from the RGC: N/A

Reasons for the change:

5.3 Realisation of the objectives

(Maximum 1 page; please state how and to what extent the project objectives have been achieved; give reasons for under-achievements and outline attempts to overcome problems, if any)

Objective 1. Develop a new theoretical framework by incorporating the theories of clustering into PCA.

Research Activities: We have developed a new theoretical framework that can perform PCA via clustering. We have proved that the well-known l_1 −PCA method can be expressed as a special formulation of a two-group clustering problem. Based on this new discovery, we have developed two different algorithms to perform PCA. The first one is to change the two-group clustering problem to be a trimmed based clustering problem. This method has been applied to face recognition and object recognition. A paper entitled "A Trimmed Clustering-Based l1-Principal Component Analysis Model for Image Classification and Clustering Problems with Outliers" has been published [See Section C for details]. Another formulation is to change the two-group clustering problem to be a statistical mode-based clustering. The statistical mode is widely known as the item that appears most in a dataset. We have developed a mathematical relaxation and successfully applied to video surveillance problem. This work is now under review entitled "Statistical Bootstrap-based Principal Mode Component Analysis for Dynamic Background Subtraction".

Objective 2. Develop efficient implementations for the new theoretical framework.

Research Activities: For the principal mode component analysis, we have developed a novel method that can quickly obtain the global optimal solution and independent of the video size. That means, if the computation time of a video sequence with small frames such as 100 x 100 is almost the same as a video sequence with large frames such as 1024 x 1024.

Objective 3. Conduct extensive experimental testing of the proposed framework for various applications.

Research Activities: We compared the two proposed algorithms with several state-of-the-art methods in face recognition, object recognition and video surveillance problems.

5.4 Summary of objectives addressed to date

Objectives *(as per 5.1/5.2 above)*	**Addressed** *(please tick)*	**Percentage Achieved** *(please estimate)*
1. Develop a new theoretical framework by incorporating the theories of clustering into PCA.	√	100%
2. Develop efficient implementations for the new theoretical framework.	√	100%
3. Conduct extensive experimental testing of the proposed framework for various applications.	√	100%

6. Research Outcome

6.1 Major findings and research outcome
(Maximum 1 page; please make reference to Part C where necessary)

The key objective of this project is to propose new algorithms that can handle outlier problems in different imaging data, namely, face recognition, object recognition and video surveillance data. The major findings of this project are

a. Face and Object Recognition: Although the widely used l_1 −PCA is robust to many face and object recognition problems, they are sensitive to the presence of outliers. The correct classification rate can be significantly dropped down if outliers are present. In this project, we found that the l_1 −PCA can be expressed as a two-group clustering problem. Based on this, we have developed a trimmed-clustering based PCA method. We mathematically proved that this new method is insensitive to outliers. Experimental results show that this new method not only outperforms l_1 −PCA but also state-of-the-art methods for data with outliers. This work has been published with title "A Trimmed Clustering-Based l1-Principal Component Analysis Model for Image Classification and Clustering Problems with Outliers" in Applied Sciences with impact factor 1.689.
b. Video Surveillance: The purpose of video surveillance problem is to detect moving foreground from a static or even dynamic background. Owing to the nearly static nature of the background, the foreground can be treated as outliers. Many methods have been developed and perform very well. However, detecting foreground from dynamic background such as a campus with waving tree branches is still challenging. Many existing methods can wrongly recognize the waving tree as foreground instead of the moving persons in the campus. We have developed a statistical mode-based PCA method that can find the mostly appeared video frames. This method outperforms many state-of-the-art methods. This is owing to the fact that although the background is dynamic, there are many repeatedly appeared patterns such as waving tree branches. The proposed mode-based method can capture the mostly repeated patterns. This work entitled "Statistical Bootstrap-based Principal Mode Component Analysis for Dynamic Background Subtraction" is now under review.

6.2 Potential for further development of the research and the proposed course of action *(Maximum half a page)*

We discovered that that there is a close connection between the proposed framework and the binary quadratic programming problems. Because of this, we have

developed a novel stochastic based optimization that can find nearly the global optimal solution of several types of binary quadratic programming problems quickly. This discovery is ground-breaking because the binary quadratic programming problems are always known as NP-hard problem. That means, the global optimal solution cannot be found in real-time. They usually need several hours or months to obtain a good but no global optimal solution. We are now preparing the paper and planning to submit this to a higher rank journal. The tentative title of the paper is "Stochastic Neighborhood Formulation for Binary Quadratic Programming".

7. Layman's Summary

(Describe in layman's language the nature, significance and value of the research project, in no more than 200 words)

Extracting useful information is one of the most important tasks in many pattern recognition and machine learning problems. However, outliers that are usually known as unusual observations appear in many real-world imaging problems and confuse existing methods. In face recognition problems, the outliers can be the sunglasses or scarves of facial images. Existing methods can wrongly treat the sunglass or scarves as important information that can recognize a human face. In video surveillance problems, the outliers can be the moving persons in a garden with big tree moving its branches. Existing methods can wrongly treat the waving tree branches as a moving foreground instead of the moving persons. In this project, we develop two different novel methods based on principal component analysis and clustering. In the past, these two techniques were developed separately. Seldom work have been reported and discussed their connection. We have found that these two techniques are closely related and successfully combined these two methods and proposed two different algorithms to handle the above problems. We have conducted extensive experiments and compared our proposed algorithms with different methods. Experimental results show that our proposed methods can not only handle the above problems but also outperform state-of-the-art methods.

Part C: Research Output

8. Peer-Reviewed Journal Publication(s) Arising Directly From This Research Project

(Please attach a copy of the publication and/or the letter of acceptance if not yet submitted in the previous progress report(s). All listed publications must acknowledge RGC's funding support by quoting the specific grant reference.)

The Latest Status of Publications				Author(s) *(denote the corresponding author with an asterisk*)*	Title and Journal / Book *(with the volume, pages and other necessary publishing details specified)*	Submitted to RGC *(indicate the year ending of the relevant progress report)*	Attached to this Report *(Yes or No)*	Acknowledged the Support of RGC *(Yes or No)*	Accessible from the institutional repository *(Yes or No)*
Year of Publication	**Year of Acceptance** *(For paper accepted but not yet published)*	**Under Review**	**Under Preparation** *(optional)*						
2019				**Benson S. Y. Lam* and S. K. Choy**	**A Trimmed Clustering-Based l_1-Principal Component Analysis Model for Image Classification and Clustering Problems with Outliers/** ***Applied Sciences 2019, 9, 1562.***		**Yes**	**Yes**	**Yes**
		2019		**Benson S. Y. Lam*, Amanda Chu and H. Yan**	**Statistical Bootstrap-based Principal Mode Component Analysis for Dynamic Background Subtraction/Pattern Recognition**		**No**	**Yes**	**No**
			2019	**Benson S. Y. Lam* & Alan W. C. Liew**	**Stochastic Neighborhood Formulation for Binary Quadratic Programming**		**No**	**Yes**	**No**

9. Recognized International Conference(s) In Which Paper(s) Related To This Research Project Was / Were Delivered

(Please attach a copy of each conference abstract)

Month / Year / Place	Title	Conference Name	Submitted to RGC *(indicate the year ending of the relevant progress report)*	Attached to this Report *(Yes or No)*	Acknowledged the Support of RGC *(Yes or No)*	Accessible from the institutional repository *(Yes or No)*

N/A						

10. Whether Research Experience And New Knowledge Has Been Transferred / Has Contributed To Teaching And Learning

(Please elaborate)

We have developed a new module AMS 4660 Machine Learning. This course was launched in Jan, 2018. We introduced some updated problems to students including face recognition, object recognition and video surveillance. We also introduced techniques on how the outliers are behaved in imaging data and also how they can be handled.

11. Student(s) Trained

(Please attach a copy of the title page of the thesis)

Name	Degree Registered for	Date of Registration	Date of Thesis Submission / Graduation
N/A			

12. Other Impact

(e.g. award of patents or prizes, collaboration with other research institutions, technology transfer, teaching enhancement, etc.)

N/A

13. Public Access Of Completion Report

(Please specify the information, if any, that cannot be provided for public access and give the reasons.)

Information that Cannot Be Provided for Public Access	Reasons
N/A	

RESEARCH GRANTS COUNCIL
COMPETITIVE RESEARCH FUNDING SCHEMES FOR
THE LOCAL SELF-FINANCING DEGREE SECTOR

FACULTY DEVELOPMENT SCHEME (FDS)

Completion Report - Attachment
(for completed projects only)

RGC Ref. No.:	**UGC/FDS14/E03/14**
Principal Investigator:	Dr. LAM, Shu Yan
Project Title:	Data Functional Modelling with Outliers

Statistics on Research Outputs

	Peer-reviewed Journal Publications	Conference Papers	Scholarly Books, Monographs and Chapters	Patents Awarded	Other Research Outputs (Please specify)
No. of outputs arising directly from this research project [or conference]	1 (1 under review, 1 under preparation)				