| RGC Ref.: N_HKU 729/13 |
| NSFC Ref. : 61361166006 |
| *(please insert ref. above)* |

The Research Grants Council of Hong Kong
NSFC/RGC Joint Research Scheme
__Joint Completion Report__

(Please attach a copy of the completion report submitted to the NSFC by the Mainland researcher)

Part A: The Project and Investigator(s)

1.   Project Title

Research problems on carving and tamper detection of fragmented multimedia evidence for forensic investigation

2.   Investigator(s) and Academic Department/Units Involved

	Hong Kong Team	Mainland Team
Name of Principal Investigator *(with title)*	Siu Ming Yiu (Dr)	Xiamu Niu (Dr)
Post	Associate Professor	Professor
Unit / Department / Institution	Computer Science/The University of Hong Kong	School of Computer Science & Technology/ Harbin Institute of Technology
Contact Information	smyiu@cs.hku.hk	qiongli@hit.edu.cn
Co-investigator(s) *(with title and institution)*	Lucas CK Hui (Dr, The University of Hong Kong, now at Astri)	J. Fang (Dr, Jinan University) S. Wang (Dr, Harbin Institute of Technology) L. Jiang (Dr, Harbin Institute of Technology) L. Li (Dr, Harbin Institute of Technology)

3.   Project Duration

	Original	Revised	Date of RGC/ Institution Approval *(must be quoted)*
Project Start date	01/01/2014		
Project Completion date	31/12/2017		
Duration *(in month)*	48		
Deadline for Submission of Completion Report	31/12/2018		

Part B: The Completion Report

5.   Project Objectives

5.1   Objectives as per original application

1. Propose and design robust algorithms for carving multimedia (image and video) data fragments of mixed files and/or from different distributions on the consecutiveness and an evaluation measure on how accurate the carved file is.
   2. Propose and design highly accurate algorithms for tamper detection of multimedia (image and video) data based on the fully or partially carved images/videos resulting from (1) together with a confidence measure.
   3. Propose and develop an overall accuracy (scientific) measure to give a confidence level on how likely the carved images/video have been tampered or modified.

5.2　Revised Objectives

Date of approval from the RGC: ＿＿＿＿＿＿＿＿＿＿＿

Reasons for the change: ＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

1.
2.
3. ....

6. Research Outcome

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

Regarding the carving of multimedia files, we have the following major findings. For JPEG files, we have developed effective algorithms for reconstructing the JPEG files based on the following scenarios: (a) with and without partial information damaged or permanently erased; (b) the deleted JPEG files are heavily fragmented (i.e., the same picture has been chopped into many fragments); (c) with multiple files mixed in the pool; and (d) real data from a SD card.

In particular, we have derived a new algorithm to deal with the case when the header of a JPEG file was permanently erased, a common situation for file carving which was not addressed adequately by previous research results [6]. The effectiveness of our algorithm has been demonstrated on a real dataset collected from a SD card with many JPEG files mixed together. Most of the existing carving tools are only effective if the target file is not heavily fragmented. We developed a new measure and a new algorithm to handle heavily fragmented files [3]. Our experiments show that our algorithm is almost 20% better than the most popular commercially available tool. In [1], we also derived an effective method to estimate the width of a JPEG file in order to increase the accuracy of file carving results.

For audio files, we proposed a semi-automatic framework to reconstruct related voice chat files for forensic investigation [7]. For video files, we leverage the techniques in perceptual hash and optical flow to develop a 2-stage framework for carving video files [5]. There were very little work in carving audio and video files, our results provide important insights for follow-up research.

For detecting tampered images and videos, we have developed several techniques to identify and assess the tampered components. In particular, in [4], we derived a detection algorithm based on DCT coefficient analysis and the evaluation results show that our approach is promising. We also made use of the idea of checking whether two consecutive frames are smooth or not (abrupt changes) to detect if a frame has been deleted or not [2]. Regarding the overall assessment of a carved tampered file, we found that the assessment of the carving result and the assessment of the tampered component are mostly independent. The measure for the accuracy of the carving and the measure for the tampered component can be considered separately to provide an overall assessment result.

The above summary mainly focuses on the findings based on the joint effort of both teams (resulting with joint papers).

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

(1) In this current project, we have an assumption that the files or more precisely the fragments to be carved are already classified according to their file types and the accuracy is high. However, as more and more file types appear, this assumption may not be always valid. That is, existing classification algorithms for file fragments may not be as accurate as we expect as the number of file types increases. We have conducted a preliminary study on this problem (see [8]). For future development, we should revisit the file fragment classification problem.

[8] Qian Chen, Qing Liao, Zoe L. Jiang, Jun-bin Fang, Siu-Ming Yiu, Guikai Xi, Rong Li, Zhengzhong Yi, Xuan Wang, Lucas C.K. Hui, Dong Liu, En Zhang, "File fragment classification using grayscale image conversion and deep learning in digital forensics", IEEE Symposium on Security and Privacy workshops 2018, p.140-147, 2018.

(2) The research results of this project are important for forensic investigation, in particular, for law enforcement units. More efforts should be put on transferring the research results into industrial/practical software tools for the community.

7. **The Layman's Summary**
   (describe <u>in layman's language</u> the nature, significance and value of the research project, in no more than 200 words)

File carving and tamper detection are two major problems in forensic investigation for multimedia files. From the research perspective, both problems are difficult and challenging. From a practical point of view, different scenarios, for instances, having many JPEG fragments mixed together or with the JPEG headers permanently erased, complicate the already difficult research problems a lot. On the other hand, solving these problems have a big impact on helping the law enforcement units to identify critical and relevant digital evidence as well as provide effective approaches for people to quickly identify fake videos or images, which are getting common in the Internet.

The outcomes of this project provide a number of effective algorithms and frameworks for solving some of the open research problems in these two areas. Transferring these research results into practical software could result in useful tools for practitioners in these areas. The results, not only advance the effectiveness of carving and tamper detection algorithms for multimedia files, but also provide important insights for the community to continue exploring better methodologies for solving related problems.

Part C: Research Output

8. Peer-reviewed journal publication(s) arising <u>directly</u> from this research project
(Please attach a copy of each 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) *(**bold** the authors belonging to the project teams and 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 this Joint Research Scheme *(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)*						
[1] 2018				Xianyan Wu*, Qi Han, **Xiamu Niu**, Hongli Zhang, **Siu-Ming Yiu**, **Jun-bin Fang**	JPEG image width estimation for file carving IET Image Processing, 12(7), p.1245-1252, 2018.	No	Yes	Yes	No
[2] 2016				Liyang Yu, Huanran Wang, Qi Han, **Xiamu Niu***, **Siu-Ming Yiu**, **Jun-bin Fang**, Zhifang Wang	Exposing frame deletion by detecting abrupt changes in video streams Neurocomputing 205, p.84-91, 2016.	No	Yes	Yes	No

[3] 2016				Yanbin Tang, **Junbin Fang**, K.P. Chow, **S.M. Yiu***, Jun Xu, Bo Feng, Qiong Li, Qi Han	Recovery of heavily fragmented JPEG files Digital Investigation, 18, S108-S117, 2016.	No	Yes	Yes	No
[4] 2016				Liyang Yu, Qi Han, Xiamu Niu*, S.M. Yiu, Junbin Fang, Ye Zhang	An improved parameter estimation scheme for image modification detection based on DCT coefficient analysis Forensic Science International, 259, p.200-209, 2016.	Yes (Feb 2016)	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 delivered paper. All listed papers must acknowledge RGC's funding support by quoting the specific grant reference.)*

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 this Joint Research Scheme *(Yes or No)*	Accessible from the institutional repository *(Yes or No)*
[5] 02/2017/ Florida, US	Semantic video carving using perceptual hashing and optical flow	IFIP Int. Conf. Digital Forensics 2017	No	Yes	Yes	No
[6] 11/2015/ Florida, US	Fragmented JPEG file recovery using pseudo headers	IFIP Int. Conf. Digital Forensics 2015	Yes (Feb 2016)	No	Yes	No

[7] 10/2015/ Málaga, Spain	An integrated audio forensic framework for instant message investigation	The 10th International Conference on Systematic Approaches to Digial Forensic Engineering (SADFE 2015)	Yes (Feb 2016)	No	No	No

10. 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
Yanbin Tang	PhD	09/2011	01/2016

11. Other impact *(e.g. award of patents or prizes, collaboration with other research institutions, technology transfer, etc.)*