

DOCTORAL THESIS

Towards Low-cost and Real-time Mobile Sensing

LI, Hao

Date of Award:
2021

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Abstract

With the increasing popularity of sensor-equipped smartphones, mobile sensing nowadays emerges as a promising direction. Mobile sensing applications analyze the collected signal data from the surroundings and thus understand the physical environment. Various applications have been developed through this new paradigm, such as mobile location estimation, audio recognition, and augmented reality. However, many practical issues need to be addressed when deploying mobile sensing applications in real-world scenarios. In this thesis, we investigate three practical problems related to mobile sensing applications and design customized frameworks and methods to provide low-cost and real-time mobile sensing services, particularly in the area of mobile location sensing.

First, we investigate the high data collection cost problem in WiFi fingerprint-based mobile location sensing. We propose a general framework with a low-cost offline data collection while maintaining high localization accuracy. In particular, we reduce the number of reference points to obtain a sparse fingerprint. Our framework adopts the clustering method to reduce the adverse effects when applying regression-based approaches on the sparse fingerprint. The proposed framework can provide high localization accuracy through extensive experiments on the campus.

Second, we study the heterogeneous mobile devices problem in passive mobile location sensing systems. We propose a customized localization approach that automatically infers a signal-strength-to-distance function for every device on the fly and simultaneously estimates its location with the Expectation-Maximization algorithm.

A real-world pilot test at an exhibition center is conducted, and heterogeneous mobile devices can be localized and tracked accurately.

Third, we present a mobile deep learning inference framework to schedule multiple DNN jobs with real-time requirements for deep mobile sensing tasks. Considering characteristics of DNN workloads and mobile hardware, we design the framework with mobile GPU/CPU collaboration by DNN partitioning and CPU offloading. The proposed framework can better utilize all computational resources on mobile phones. We evaluate our system on the mobile platform by extending TensorFlow Lite. The evaluation results indicate that our framework can support real-time deep mobile sensing tasks.

Keywords: Mobile Sensing, Location Estimation, WiFi Signal Strength, Real-time Scheduling, Deep Learning Inference

Table of Contents

Declaration	i
Abstract	ii
Acknowledgments	iv
Table of Contents	v
List of Tables	ix
List of Figures	x
Chapter 1 Introduction	1
1.1 Low-cost mobile location sensing with WiFi fingerprint	2
1.2 Low-cost location sensing for heterogeneous mobile devices	4
1.3 Real-time scheduling for deep mobile sensing	6
1.4 Thesis organization	9
Chapter 2 Background and Related Works	10
2.1 WiFi-based mobile location sensing	10
2.1.1 The fingerprint approach	11
2.1.2 The multilateration approach	13
2.1.3 The center of gravity approach	14
2.2 Deep mobile sensing	15
2.2.1 Mobile sensing with deep learning	15
2.2.2 DNN inference acceleration on mobile devices	17
2.2.3 Scheduling for DNN inference	18

Chapter 3	Low-cost Mobile Location Sensing with WiFi Fingerprint	20
3.1	Introduction	20
3.2	Motivation	21
3.3	The proposed framework	25
3.3.1	Cluster discovery	26
3.3.2	Regression analysis	28
3.3.3	Online estimation strategy	31
3.4	Experimental evaluations	32
3.4.1	Experiments setup	32
3.4.2	Experiments results	34
3.5	Summary	36
Chapter 4	Low-cost Location Sensing for Heterogeneous Mobile Devices	38
4.1	Introduction	38
4.2	Motivation	39
4.3	System architecture	41
4.4	Proposed localization approach	43
4.4.1	Modeling signal-strength-to-distance relationship	44
4.4.2	Estimating location of the targeted device	45
4.4.3	Inferring parameters of the piecewise model	47
4.4.4	The EM algorithm for localization	48
4.4.5	Complexity analysis	49
4.5	Practical considerations	50
4.5.1	Number of sniffers	50
4.5.2	3D localization model	50
4.6	Experiments	51
4.6.1	Localization experiments in testbed	52
4.6.2	Localization experiments in HKCEC	56

4.7	Summary	59
Chapter 5	Real-time Scheduling For Deep Mobile Sensing	61
5.1	Introduction	61
5.2	Motivation	62
5.2.1	Deep neural networks	62
5.2.2	DNN inference on mobile devices	64
5.3	System design	66
5.3.1	Two-stage execution	66
5.3.2	System architecture	68
5.3.3	Utility maximization	68
5.3.4	Execution latency profiling	69
5.4	Problem formulation	70
5.5	A dynamic programming approach	71
5.5.1	System state	73
5.5.2	State transition equations	75
5.5.3	Algorithm design	76
5.6	Benchmarks	78
5.7	Implementation	78
5.8	Evaluation	82
5.8.1	Evaluation metrics	82
5.8.2	Performance results	83
5.9	Summary	85
Chapter 6	Conclusion and Future Work	87
6.1	Conclusion	87
6.1.1	Low-cost mobile location sensing with WiFi fingerprint . . .	87
6.1.2	Low-cost location sensing for heterogeneous mobile devices	88
6.1.3	Real-time scheduling for multi-DNN inference on mobile . .	88

6.2	Future work	89
6.2.1	Mobile location sensing with indoor map information fusion	89
6.2.2	Energy-efficient DNN inference on mobile	89
	Bibliography	91
	List of Publications	101
	Curriculum Vitae	103

List of Tables

3.1	Part of RSS data in three reference points	23
3.2	The mean localization error of the-state-of-art classification and regression approaches	33
3.3	The mean localization error of our proposed framework	34
4.1	Mean localization error of different approaches for different devices (the fingerprint model is constructed based on the Nexus phone) . .	55
4.2	Mean localization error of the fingerprint with different training devices	56
5.1	DNN models' execution latency on mobile GPU and CPU (Samsung A71)	64
5.2	Notations adopted in Chapter 5	72

List of Figures

2.1	The fingerprint approach	11
2.2	The multilateration approach	13
2.3	The general framework of deep mobile sensing	16
3.1	The sparse fingerprint collected in Hong Kong Baptist University and the DBSCAN clustering result	22
3.2	The negative effect on regression analysis from noise data	24
3.3	The overall process of proposed localization framework	24
3.4	The clustering process of DBSCAN algorithm	25
3.5	The testing points in the testbed of HKBU	33
3.6	The mean localization error of regression-based approaches and corresponding approaches in our framework	35
3.7	The mean localization error of state-of-the-art approaches and our framework approaches	36
3.8	The mean localization error of GBT based approaches	37
3.9	The mean localization error of SVM based approaches	37
3.10	The mean localization error of RF based approaches	37
4.1	The measured RSS of the packets originating from two mobile phones (Nexus 6p and Samsung C9) versus the distance from the measuring device	40
4.2	System architecture of the localization system.	42

4.3	Our proposed localization approach.	44
4.4	Illustration of different sniffer layouts (in height) resulting in different horizontal distances, given the same device to sniffer distance.	51
4.5	The FSC801 testbed in HKBU	52
4.6	Localization results in estimating the location of a Nexus phone with different approaches	53
4.7	Localization results in estimating the location of an iPhone with different approaches (The fingerprint model is constructed based on the Nexus phone)	54
4.8	Localization results in estimating the location of a Sony phone with different approaches (The fingerprint model is constructed based on the Nexus phone)	54
4.9	(a) Mean localization error with EM and without EM (b) Mean localization error with weights and without weights	56
4.10	Sniffers in our real-world pilot test	57
4.11	Heap map in the exhibition venue	58
4.12	Actual trajectory and estimated trajectory by our localization method of a Sony phone in HKCEC	58
4.13	Actual trajectory and estimated trajectory by our localization method of an Asus phone in HKCEC	59
4.14	Actual trajectory and estimated trajectory by our localization method of a Nexus phone in HKCEC	60
5.1	An example of DNN with the multi-layer structure	62
5.2	FLOPs of layers in some DNN models	63
5.3	DNN models' layer-wise execution latency on mobile GPU and CPU (Samsung A71)	65
5.4	Multiple DNN inference in TFLite	66

5.5	Multiple DNN inference with CPU offloading	67
5.6	System architecture	69
5.7	The illustration of DNN Partitioning	70
5.8	Example for state transition with two cases	74
5.9	Latency profiling results for different DNN models ((a) Inception (b) DenseNet-121 (c) FaceNet (d) Resnet-50 (e) Posenet (f) Deeplab)	81
5.10	The deadline miss rate under different numbers of tasks in a scheduling period	84
5.11	The completion rate under different numbers of tasks in a scheduling period	84
5.12	CPU utilization rate under different numbers of tasks in a scheduling period	85