AFRL-AFOSR-JP-TR-2022-0035

An Infinitely Scalable Learning and Recognition Network

Milford, Michael
QUEENSLAND UNIVERSITY OF TECHNOLOGY
2 GEORGE ST
BRISBANE, , 4000
AUS

05/25/2022
Final Technical Report

Air Force Research Laboratory
Air Force Office of Scientific Research
Asian Office of Aerospace Research and Development
Unit 45002, APO AP 96338-5002

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**	
20220525	Final	**START DATE** 20190923	**END DATE** 20200922

4. TITLE AND SUBTITLE
An Infinitely Scalable Learning and Recognition Network

5a. CONTRACT NUMBER	**5b. GRANT NUMBER**	**5c. PROGRAM ELEMENT NUMBER**
FA2386-19-1-4079		
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Michael Milford

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
QUEENSLAND UNIVERSITY OF TECHNOLOGY
2 GEORGE ST
BRISBANE 4000
AUS

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)**
AOARD UNIT 45002 APO AP 96338-5002	AFRL/AFOSR IOA	AFRL-AFOSR-JP-TR-2022-0035

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Learning and recognition are fundamental process performed by animals, humans, robots and intelligent systems. Humans, for example, continually learn and recognize where they are in the world (place recognition), who is there with them (facial recognition) and what things are around them (object recognition). Recognition also plays a significant role in technology like smartphones, whether it be recognizing what you are saying (voice recognition) or what the consumer item in front of you is when using Google Goggles (object recognition). Google and other information aggregators perform recognition at a vast scale, recognizing and classifying billions of images in the cloud and house numbers in millions of kilometres of Google Streetview imagery. In security and surveillance, taskâ€•specific signatures (such as a specific person's voice, a bombâ€•carrying back pack or a person's face) must be automatically recognized amongst vast amounts of data. Common to all these artificial recognition processes are computational and storage requirements that grow with the magnitude of the task. Typically these storage and computational requirements grow linearly or worse with the size of the dataset, a critical problem in a world where data storage demand is outstripping capability (see Figure 1), and this gap is forecast to continue growing (1). There is currently no feasible solution to this problem – current techniques such as those used for video and image compression have plateaued in performance over the last decade, while the limits of hashâ€•based approaches are known and unlikely to provide an ultimate solution. This project combines modelling of and inspiration from the spatial memory encoding system in the mammalian brain with machine learning techniques to enable subâ€•linear storage growth; that is, as the number of "items" in the database (places, images, voice signatures etc.) that need to be encoded grows, the amount of storage space required per item continually decreases.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	SAR	5
U	U	U		

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
ALAN LIN	227-7009

An Infinitely Scalable Learning and Recognition Network

Principal Investigator: Professor Michael Milford
Deputy Director, QUT Centre for Robotics, Microsoft Research Faculty Fellow, Australian Centre for Robotic Vision Chief Investigator & School of Electrical Engineering and Robotics

Queensland University of Technology
2 George St Brisbane QLD 4051 Australia
michael.milford@qut.edu.au | +61 7 3138 9969

Key Collaborator: Professor David Cox
Director of the MIT-IBM Watson AI Lab
(Formerly Assistant Professor of Computer Science and of Molecular and Cellular Biology, Harvard University)

Key Collaborator: A. Professor Walter Scheirer
Department of Computer Science and Engineering,
University of Notre Dame

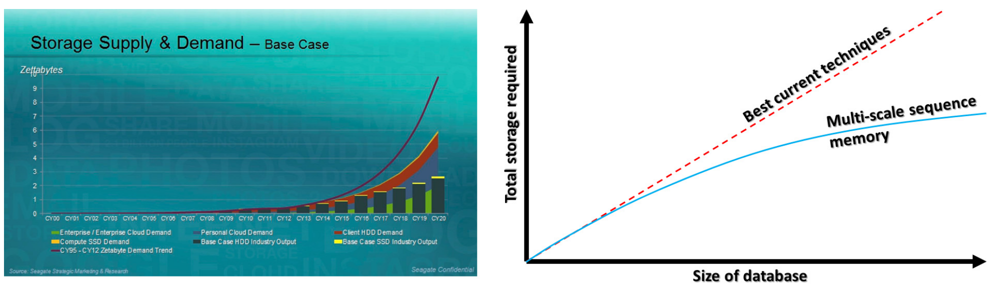
Final report for Period: June 22 2016 – Sep 30 2020 (non-continuous)

Proposed Total Cost (for each year)

Year	Amount in USD
Year 1	$90,034.62
Year 2	$90,322.62
Year 3	$92,937.62

Abstract (Publicly Releasable)

Learning and recognition are fundamental process performed by animals, humans, robots and intelligent systems. Humans, for example, continually learn and recognize where they are in the world (place recognition), who is there with them (facial recognition) and what things are around them (object recognition). Recognition also plays a significant role in technology like smartphones, whether it be recognizing what you are saying (voice recognition) or what the consumer item in front of you is when using Google Goggles (object recognition). Google and other information aggregators perform recognition at a vast scale, recognizing and classifying billions of images in the cloud and house numbers in millions of kilometres of Google Streetview imagery. In security and surveillance, task-specific signatures (such as a specific person's voice, a bomb-carrying back pack or a person's face) must be automatically recognized amongst vast amounts of data.



(a) (b)

Figure 1: (a) Data storage demand is rapidly outpacing storage availability (Source: Seagate Strategic Marketing & Research). (b) The proposed multi-scale sequence memory enables sub-linear storage growth.

Common to all these artificial recognition processes are computational and storage requirements that grow with the magnitude of the task. Typically these storage and computational requirements grow linearly or worse with the size of the dataset, a critical problem in a world where data storage demand is outstripping capability (see Figure 1), and this gap is forecast to continue growing (1). There is currently no feasible solution to this problem – current techniques such as those used for video and image compression have plateaued in performance over the last decade, while the limits of hash-based approaches are known and unlikely to provide an ultimate solution. **This project combines modelling of and inspiration from the spatial memory encoding system in the mammalian brain with machine learning techniques to enable sub-linear storage growth; that is, as the number of "items" in the database (places, images, voice signatures etc.) that need to be encoded grows, the amount of storage space required per item continually decreases.**

Enabler of Funding

The AOARD project has indirectly assisted in obtaining funding in the following schemes:

- US-Australian Multidisciplinary University Research Initiatives (MURI) bid, "Neuro-Autonomy: Neuroscience-Inspired Perception, Navigation, and Spatial Awareness". In brief:
  - 24 successful bids out of 295 proposals, 8% success rate.
  - a 3 year project extendable to 5 years
  - the U.S. side of the bid is worth approx $10M AUD over 5 years, the Australian side is about $5M over 5 years.
  - funded by ONR (US side) and DST (Australian-side).
  - collaboration between MIT, BU, Uni Melb, Macquarie, QUT and UNSW
- Advanced Terrain Detection (ATD) with Rheinmetall Defence Australia, https://research.qut.edu.au/qcr/Projects/rheinmetall-defence-australia-advanced-terrain-detection-atd/

The project and experience working in defence-related research has also formed a variable part of the foundation for funding proposals currently under review or in discussion with the following organizations:

- Australian Research Council (under review)
- Australian Department of Defence (under review)
- Amazon (confidential, under review)
- Ford (confidential, under review)

Students and Postdocs Funded by Proposal

The project has employed four postdoctoral research fellows, and also involved one PhD student.

- **Postdoc:** Dr Adam Jacobson, now working at Fortune 100 company Caterpillar in Brisbane, Australia (a long term collaborator). During Adam's PhD he also worked on some of the foundational research underlying the project.
- **Postdoc:** Dr Litao Yu, now working as a researcher at University of Technology Sydney, Australia, following a role at Griffith University, Brisbane, Australia.
- **Postdoc:** Dr Huu Le, now working as a researcher at Chalmers University, Sweden
- **Postdoc:** Dr Sourav Garg, now working as a Research Fellow at the QUT Centre for Robotics.
- **PhD student:** Mingda Xu, current second year PhD student at QUT, worked on the ICCV2019 paper and presented it in South Korea.

We would also add that the lead PI for this project, Professor Michael Milford, was during the project's duration first promoted from Associate to Full Professor at QUT, promoted to Deputy Leader of the Robotics Group (2016-2019) and then to Deputy Director and co-founder of the newly formed QUT Centre for Robotics. This project made a contribution towards this career development.

Other Dissemination of Project Outcomes

The project outcomes have been presented as part of approximately 40 presentations over the project duration: ranging from top international conferences, universities, and companies especially in the technological space. It's also been discussed in several dozen meetings with various defence departments from both Australia and the United States.

References

1. Pimentel A. http://recode.net/2014/01/10/stuffed-why-data-storage-is-hot-again-really/. http://recode.net/; 2014.
2. Chen Z, Jacobson A, Milford M, editors. Visual Segment-Based Place Recognition Inspired from Grid Cells [under review]. IEEE International Conference on Robotics and Automation; 2014; Hong Kong, China: IEEE.
3. Milford M, Wyeth G. Persistent Navigation and Mapping using a Biologically Inspired SLAM System. International Journal of Robotics Research. 2010;29(9):1131-53.
4. Milford M, Wiles J, Wyeth G. Solving Navigational Uncertainty Using Grid Cells on Robots. PLoS Comput Biol. 2010;6(11).
5. Milford MJ. Robot Navigation from Nature: Simultaneous Localisation, Mapping, and Path Planning Based on Hippocampal Models. Ditzinger T, editor. Berlin-Heidelberg: Springer-Verlag; 2008. 196 p.
6. Milford M, Wyeth G. Mapping a Suburb with a Single Camera using a Biologically Inspired SLAM System. IEEE Transactions on Robotics. 2008;24(5):1038-53.
7. Milford M, Turner I, Corke P, editors. Long exposure localization in darkness using consumer cameras. Proceedings of the 2013 IEEE International Conference on Robotics and Automation; 2013: IEEE.
8. Milford M. Vision-based place recognition: how low can you go? International Journal of Robotics Research. 2013;32(7):766-89.
9. Milford M, Wyeth G, editors. SeqSLAM: Visual Route-Based Navigation for Sunny Summer Days and Stormy Winter Nights. IEEE International Conference on Robotics and Automation; 2012; St Paul, United States: IEEE.
10. Pepperell E, Corke P, Milford M, editors. All-environment visual place recognition with SMART. IEEE International Conference on Robotics and Automation; 2014; Hong Kong, China: IEEE.
11. Burak Y, Brookings T, Fiete I. Triangular lattice neurons may implement an advanced numeral system to precisely encode rat position over large ranges. Quantitative Biology. 2006.
12. Burak Y, Fiete IR. Accurate path integration in continuous attractor network models of grid cells. PLoS Comput Biol. 2009;5(2).
13. Stensola H, Stensola T, Solstad T, Froland K, Moser M, Moser E. The entorhinal grid map is discretized. Nature. 2012;492(7427):72-8.
14. Hafting T, Fyhn M, Molden S, Moser M-B, Moser EI. Microstructure of a spatial map in the entorhinal cortex. Nature. 2005;11(436):801-6.