Project Title: Event Analysis

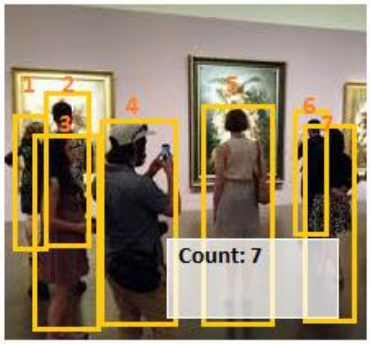
Applications of computer vision and AI in smart tourism industry

Background :

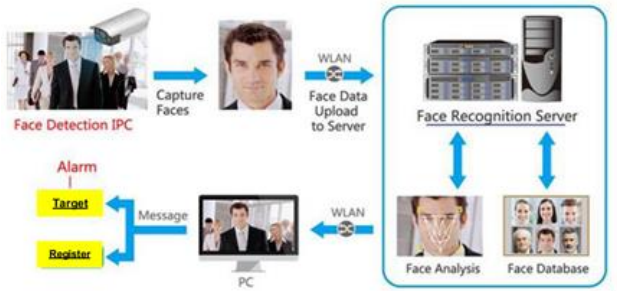
In summary, ‘event analysis’ is leveraged on two main technologies: HpVT provides the communication backbone and AI provides intelligent computing backbone. The goal of this project is to research and develop a cost-effective system that is capable of performing a smart visual analysis in the smart tourism domain in real-time or near real-time mode. The aim of this monitoring is to enhance safety and security. Participating countries will work on the same technological targets but may have different local applications. In this proposal, the event analysis will be employed in the following applications in the Smart Tourism area: (i) smart museums, (ii) smart surveillance and (iii) smart pedestrian and traffic monitoring.

Targets:

Smart museums



Smart surveillance



Smart pedestrian safety monitoring

Speaker:

Somnuk Phon-Amnuaisuk, Project leader, Universiti Teknologi Brunei

Project Title: **Event Analysis**

Project Members :

Name	Position/Degree	Department, Institution, Country	Email Address
Ah-Hwee Tan	Professor/Ph.D	Nanyang Technological University, Singapore	asahtan@ntu.edu.sg
Ken T. Murata	NICT Science Cloud Director/Ph.D	Integrated Science Data System Research Laboratory, NICT, Japan	ken.murata@nict.go.jp
La-Or Kovavisaruch	Head / NECTEC	Research Unit of Advanced Automation and Electronics, NECTEC, Thailand	La-or.kovavisaruch@nectec.or.th
Surachai Ongkittikul	Lecturer/Ph.D	Mahanakorn University of Technology, Thailand	surachai@mut.ac.th; s.ongkittikul@hotmail.com
Somsanouk Pathoumvanh	Deputy head/Ph.D	Lao-Japan Technical Training Center, Faculty of Engineering, National University of Laos, Lao PDR.	somsanouk@fe-nuol.edu.la
Myint-Myint Sein	Professor/Ph.D	University of Computer Studies, Yangon, Myanmar	myint@ucsy.edu.mm
Chin-Kuan Ho	Professor/Ph.D	Faculty of Computing & Informatics, Multimedia University, Malaysia	ckho@mmu.edu.my

Project Duration : 24 month, 1 September 2018 to 31 August 2020

Project Activities: **Bird's Eye View**

Concepts

2018

- September : kick off meeting
- November : 1st progress report

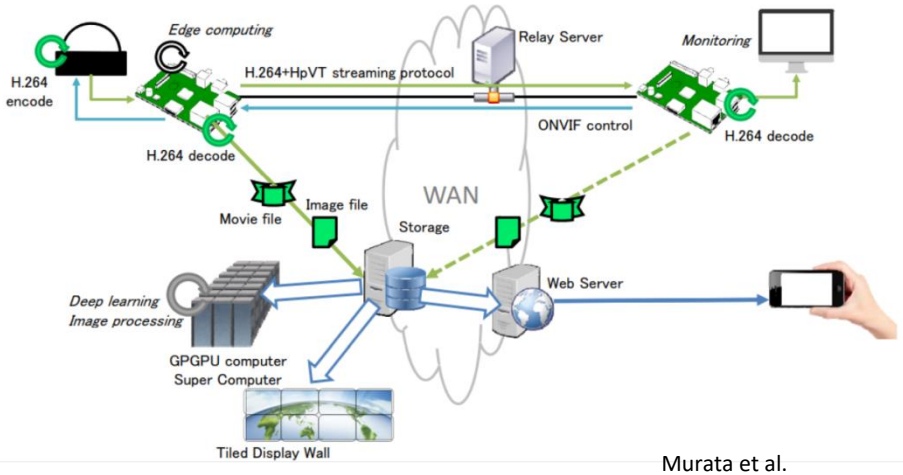
2019

- April – July : acquire cameras
- November : preliminary results

Main Activities:

- Logistics
- Installation & Data collection
- Algorithm development

Big picture on visual IoT

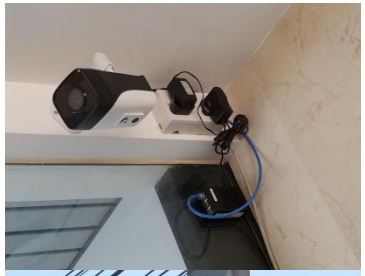


Murata et al.

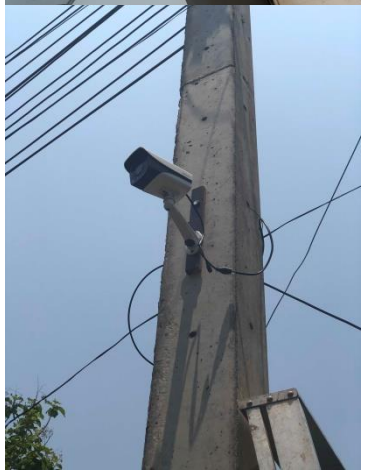
Project Activities: **Setup Camera Network**

Data Acquisition:
Thailand: MUT, NECTEC ← Museum
Laos: NUoL ← Pedestrian, cars
Myanmar: UCSY ← Tourist spots
Brunei: UTB ← Library

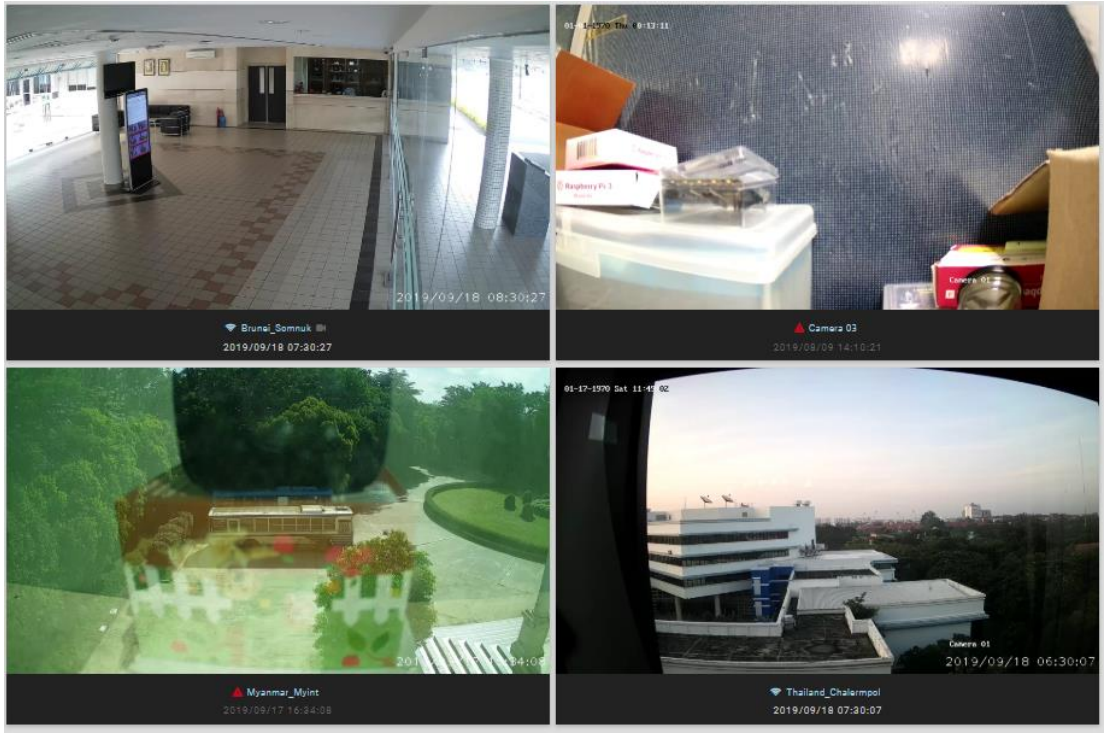
Algorithm Development:
UTB: Detection, Tracking, Counting
MUT: Face recognition
NUoL: Pedestrian, Car detection
NTU: Activity recognition



Indoor



Outdoor



Project Activities

January 2020 – August 2020

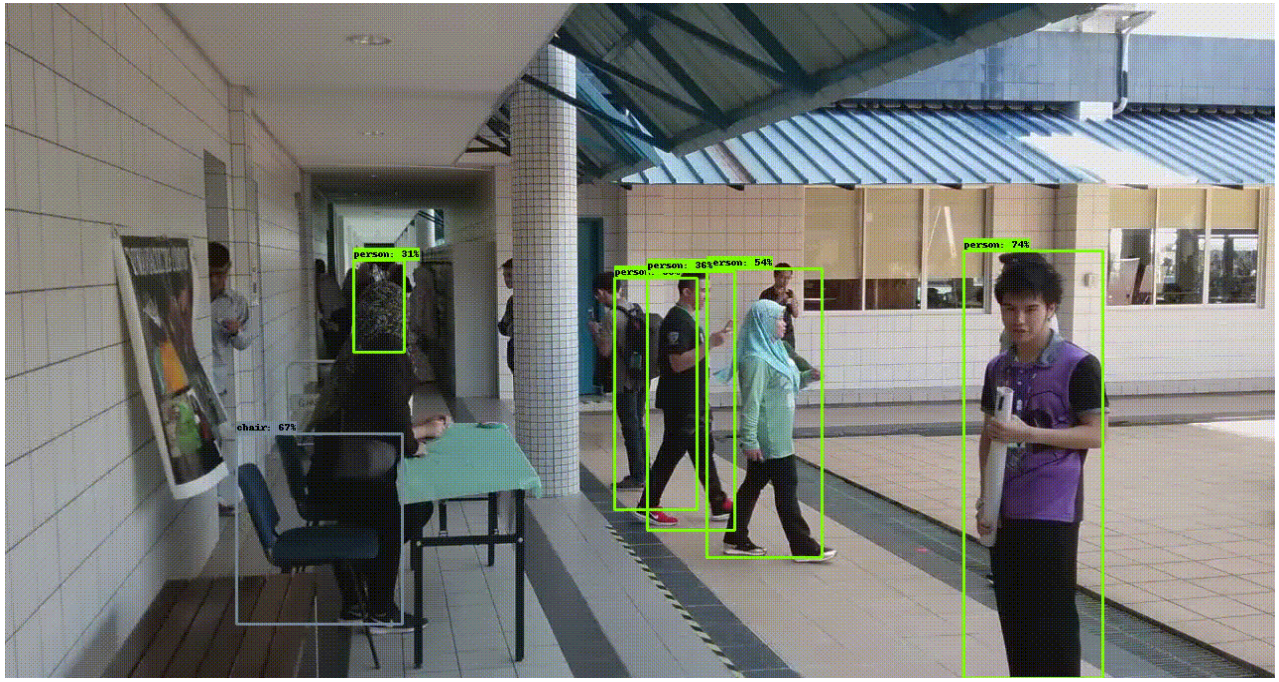
1. Meetings: three progress report meetings
- Will be hosted by UTB, tentative schedule early January 2020
- Will be hosted by NTU, tentative schedule March/April 2020
- Will be hosted by MMU, tentative schedule June/July 2020

2. Computing expenses such as servers, GPUs and TPUs

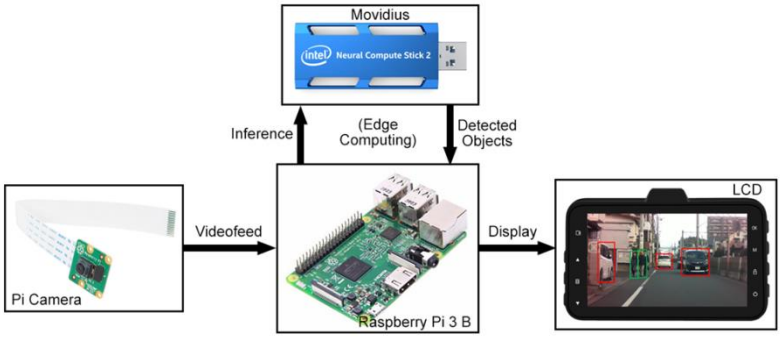
R&D results: Detection, Tracking & Counting (UTB)

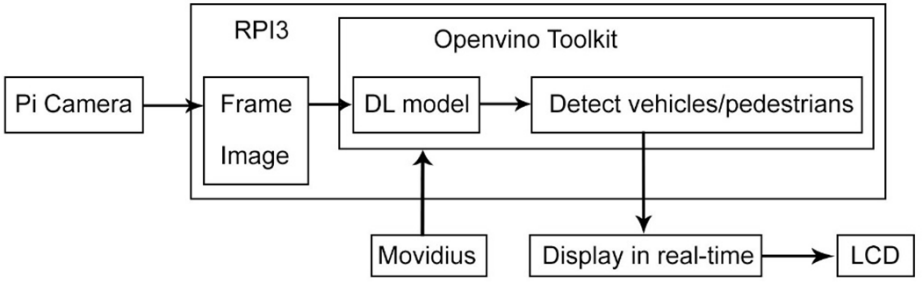
Table 1. Summary of sensitivity, specificity and average accuracy of the Faster R-CNN and the SSD detectors. Microsoft COCO is the training dataset for all models except the SSD VGG model which uses the VOC dataset.

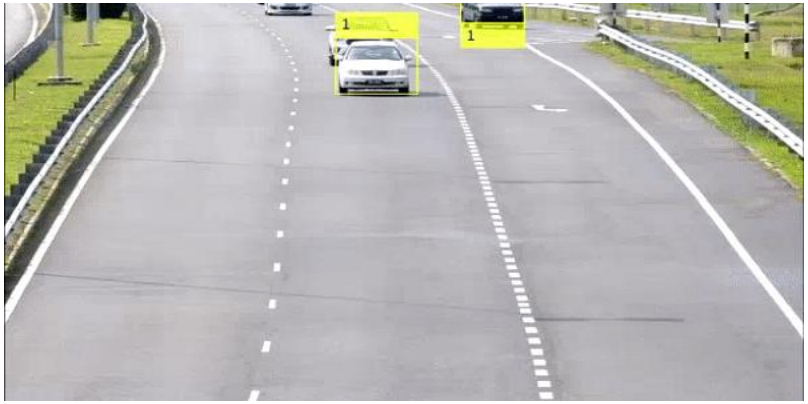
Model	Footage 1 Sens	Footage 1 Spec	Footage 2 Sens	Footage 2 Spec	Footage 3 Sens	Footage 3 Spec	Footage 4 Sens	Footage 4 Spec	Footage 5 Sens	Footage 5 Spec	Acc (%)
Faster R-CNN											
Inception model (COCO)	0.98	1.00	0.94	1.00	1.00	0.84	0.96	0.96	0.98	0.80	**95.5**
ResNet model (COCO)	0.98	1.00	0.96	1.00	1.00	0.98	0.99	0.86	0.98	0.96	**97.5**
SSD											
MobileNet model (COCO)	0.80	1.00	0.70	0.80	0.84	1.00	0.87	1.00	0.79	1.00	**85.3**
VGG model (VOC)	0.88	1.00	0.70	0.56	0.99	0.92	0.90	0.86	0.86	0.92	**86.1**

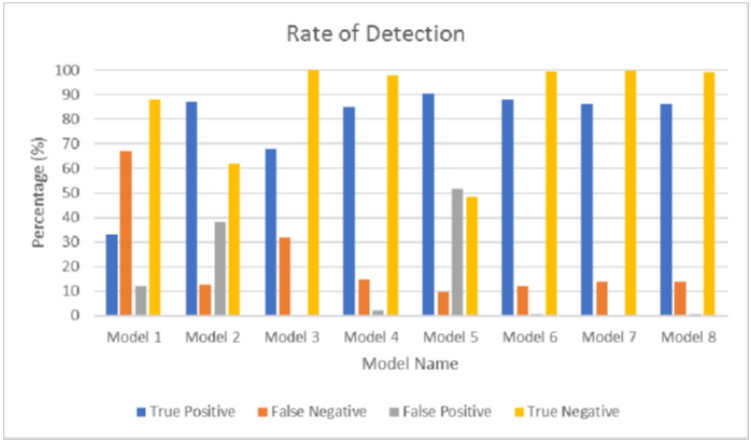


R&D results: Edge Computing (UTB & Science Cloud Lab)



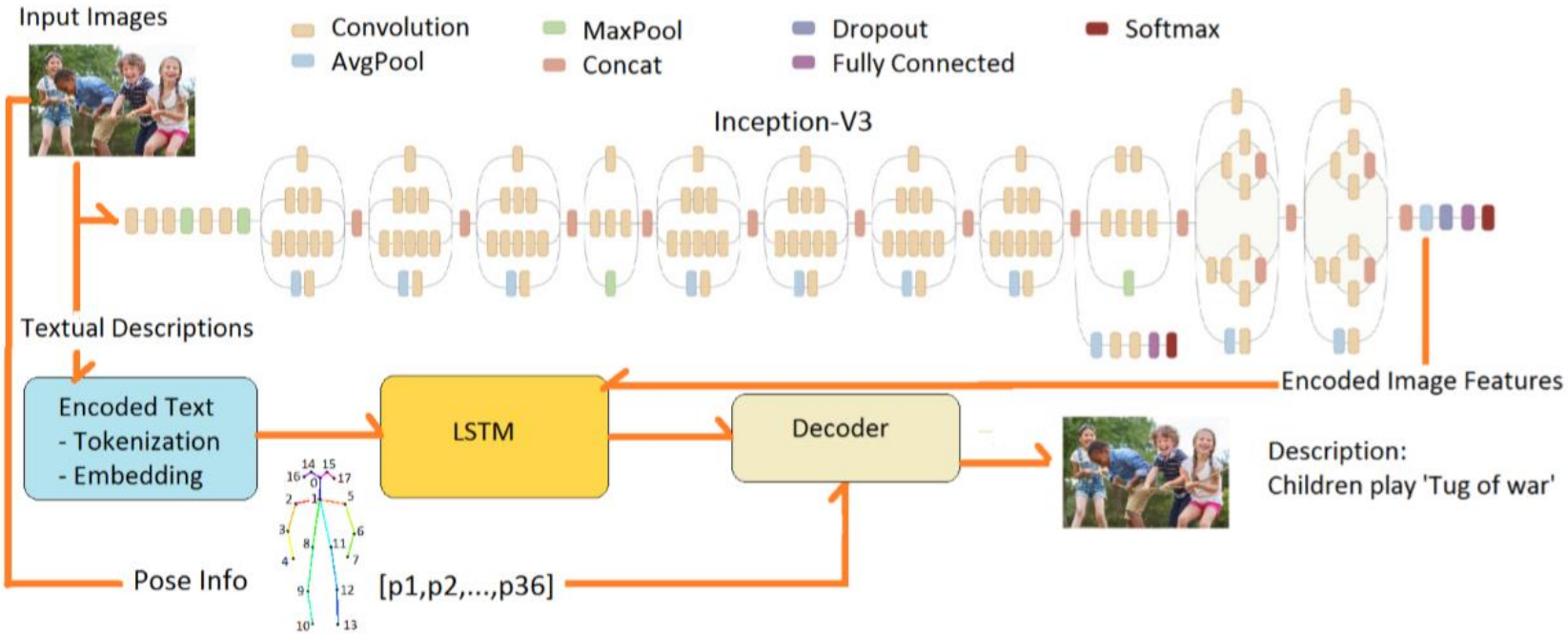






No	Models	Inference Time	Precision	Recall	F Score
	Intel Open Model Zoo				
1	person-vehicle-bike-detection-crossroad-0078	11.8 FPS	0.832	0.331	0.474
2	pedestrian-and-vehicle-detector-adas-0001	12.6 FPS	0.807	0.871	0.838
3	pedestrian-detection-adas-0002	13.4 FPS	1	0.68	0.810
4	person-detection-retail-0013	7.3 FPS	0.961	0.85	0.902
5	vehicle-detection-adas-0002	13.4 FPS	0.736	0.906	0.812
	TensorFlow Detection Model Zoo				
6	ssd_mobilenet_v1_coco	10.6 FPS	0.997	0.881	0.935
7	ssd_mobilenet_v2_coco	8.7 FPS	0.998	0.861	0.924
8	ssd_inception_v2_coco	7.6 FPS	0.995	0.862	0.924

R&D results: Visual and Textual Associations (UTB, Science cloud)



group of people are gathered around carnival game
group of kids are playing with water balloons
group of people are sitting around table with drinks

two children are playing in the grass
two boys are playing on the grass
two children are sitting on the grass

R&D results: Episodic Memory (NTU)

Episodic Memory for Video Captioning

Ah-Hwee TAN/NTU

Motivations

- Memory Networks storing episodic memory was shown to enhance Question/Answering tasks
- Most current state-0f-the-art video methods lacks Episodic memory module
- M^3 model attempts preliminary memory networks for video captioning.
  - Limits- Arbitrary number of memory slots => plasticity-elasticity dilemma

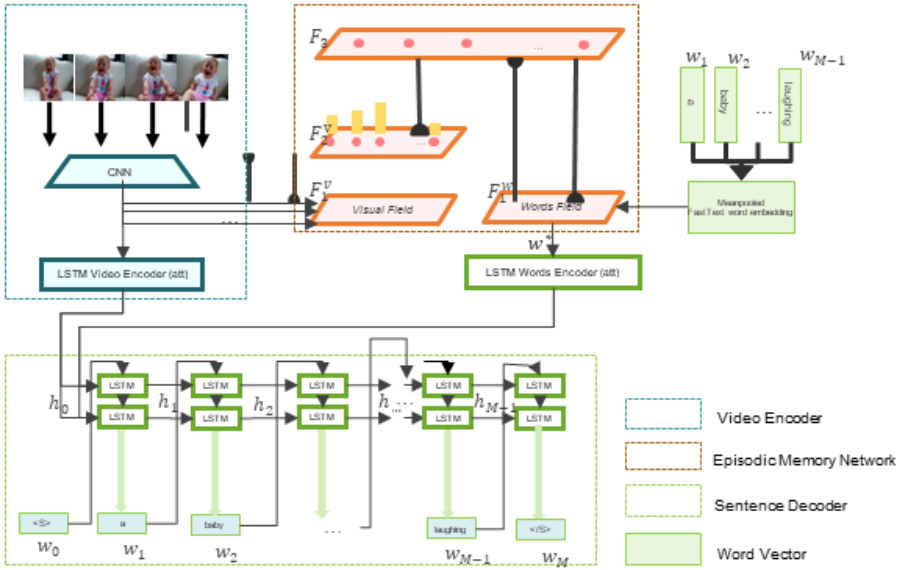
Objectives

- We propose Fusion ART based memory as a solution for plasticity-elasticity dilemma on the task of video captioning
  - Design mechanics to store and retrieve Episodic visual and textual representations
  - Utilizing memory for sentence decoding

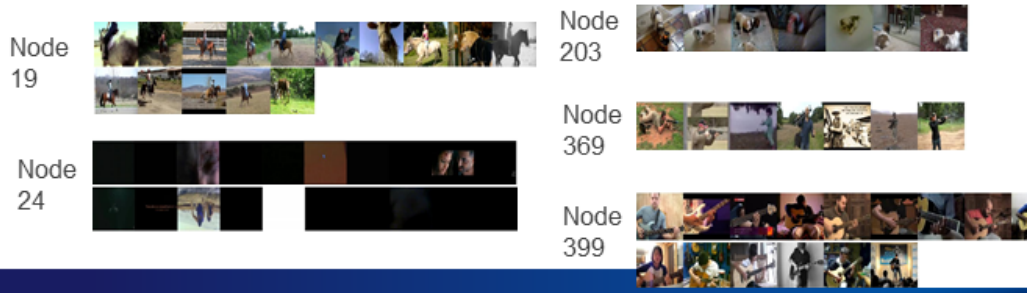
Problem Statement

- Given a video (Sequence of images), generate a natural language narration automatically .
- Formulation
  - Given image sequences
  - Decode sequences of words

Proposed Approach/Architecture



Examples of Visual Nodes Learned



NANYANG TECHNOLOGICAL UNIVERSITY | SINGAPORE

Scientific Contribution:

Presentations at International Conferences:

No:	Paper title:	Author names	Affiliation	Conference name:	The date of the conference	The venue of the conference
1	Children Activity Descriptions from Visual and Textual Associations	Somnuk Phon-Amnuaisuk, Ken T. Murata, Praphan Pavarangkoon, Takamichi Mizuhara and Shiqah Hadi	Universiti Teknologi Brunei National Institute of Information and Communication Technology CLEALINK Technology Co. Ltd.	The 13th Multi-Disciplinary International Conference on Artificial Intelligence (MIWAI 2019)	17-19/11/2019	Kuala Lumpur, Malaysia
2	Edge Computing for Road Safety Applications	Shiqah Hadi, Ken T. Murata, Somnuk Phon-Amnuaisuk, Praphan Pavarangkoon, Takamichi Mizuhara and Soon-Jiann Tan	Universiti Teknologi Brunei National Institute of Information and Communication Technology CLEALINK Technology Co. Ltd.	The 23rd International Computer Science and Engineering Conference (ICSEC 2019)	30/10/2019 to 1/11/2019	Phuket, Thailand

Societal Impact:

- Public Policy, Civil Right: This project creates a lot of discussion in terms of the privacy issues and how they should be handled.
- Social Entrepreneurship: The computer vision and AI technology provides great technological tools to SMEs.
- Environmental Sustainability: The outcome could provide a better understanding to our environment.

https://www.kellogg.northwestern.edu/social-impact/about.aspx

Conclusion:

Progress to date:

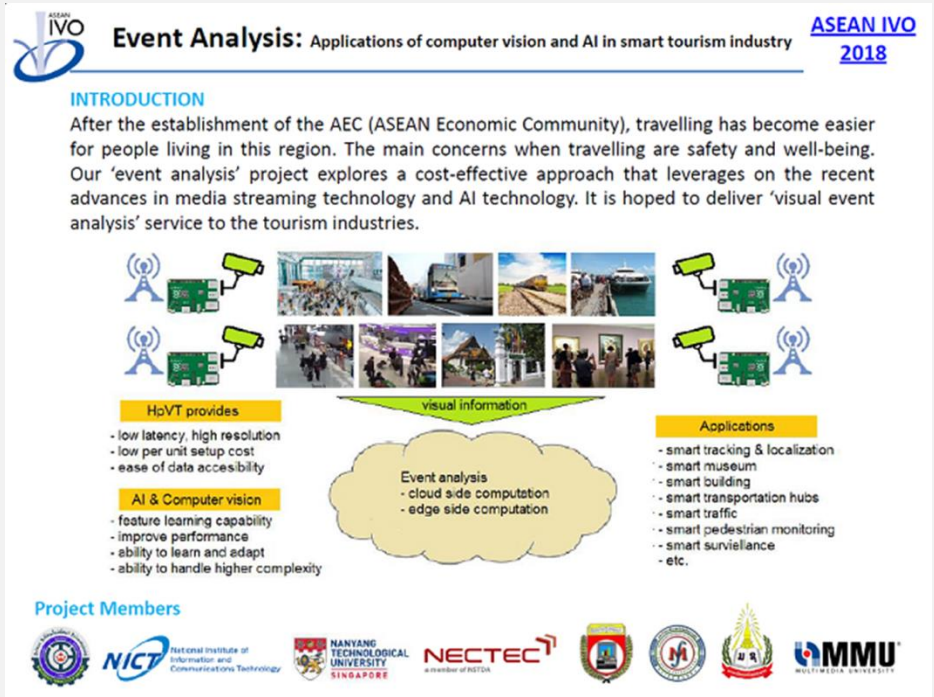
- Installation of camera and testing communications on the following sites: UTB, NECTEC, MUT, NUoL, and UCSY.
- Developing computer vision algorithms for the tasks:
  UTB: detection, image captioning
  NTU: video captioning
  NUoL, MUT: installation challenges

Challenges:

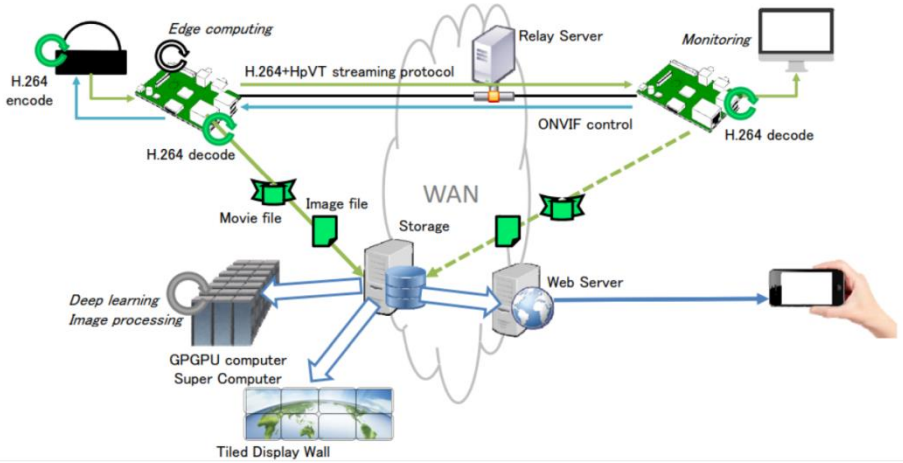
- Installed cameras at MUT and NUoL are not detected at NICT science cloud. Under investigation.
- Confusion in regulation to utilize budget e.g., the signing of CRDA takes too long to be processed.
- Privacy issues and non-existent or unclear local regulations on privacy issues.

Future works: Augment Visual Sensors with AI

Concepts



Big picture on visual IoT



Murata et al.

Visual sensors with AI augmentation could open a wide range of applications.

Future research in computer vision:

1. UTB: Anonymous objects (humans) localization and trajectory analysis.
2. NTU: Video captioning.
3. MMU: Anonymous pedestrian and vehicles analysis.