Real World Data Analysis with Big Data Software Stack on TSUBAME2.5 Supercomputer

Overview

Higher performance analysis and more precise prediction of real-world/social data is required to support development of "smart-cities". Huge real world data generated by a number of sensors, including traffic data, motion data of people, status of infrastructure, should be analyzed to make daily-life more comfortable.

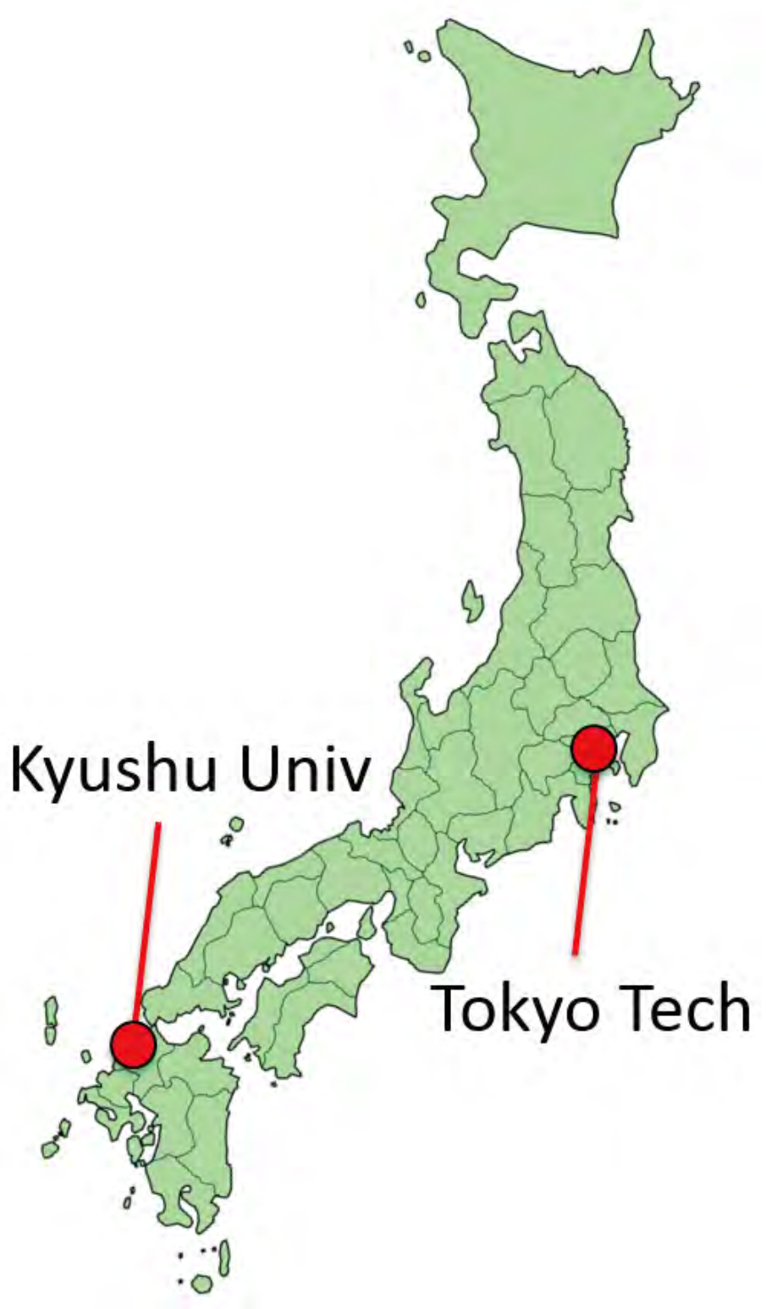
・Example: Passengers may avoid traffic congestions if they have good prediction

For this purpose, peta-scale supercomputing environment and software stack for big data analysis, especially for deep learning have to be available to big data scientists.

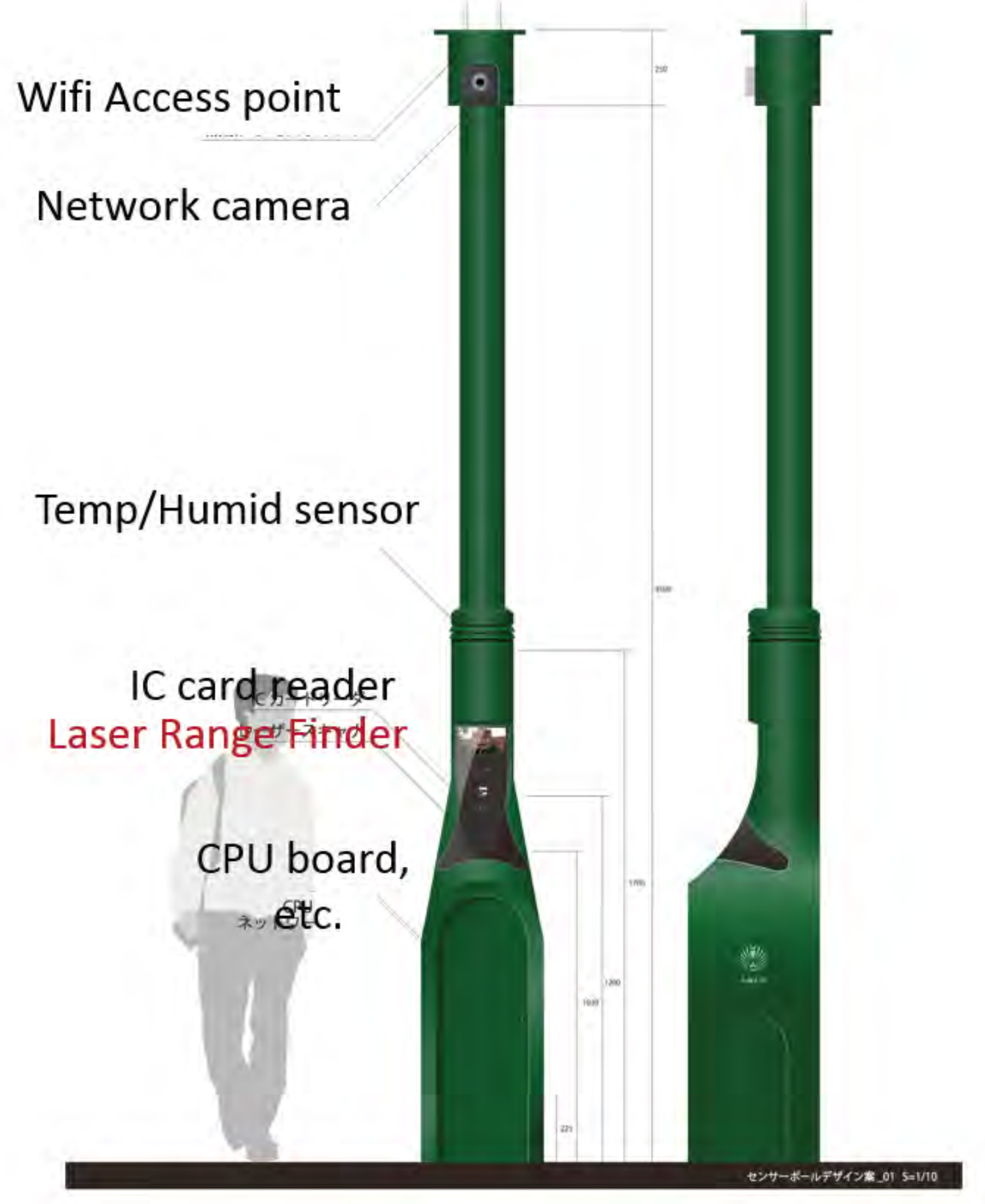
This work shows the preliminary experiments of such analyses:

・Target data: Data of people flow by P-sen sensors installed in Kyushu University

・Platform: Caffe, a deep learning tool, installed on TSUBAME2.5, GPU supercomputer in Tokyo Tech

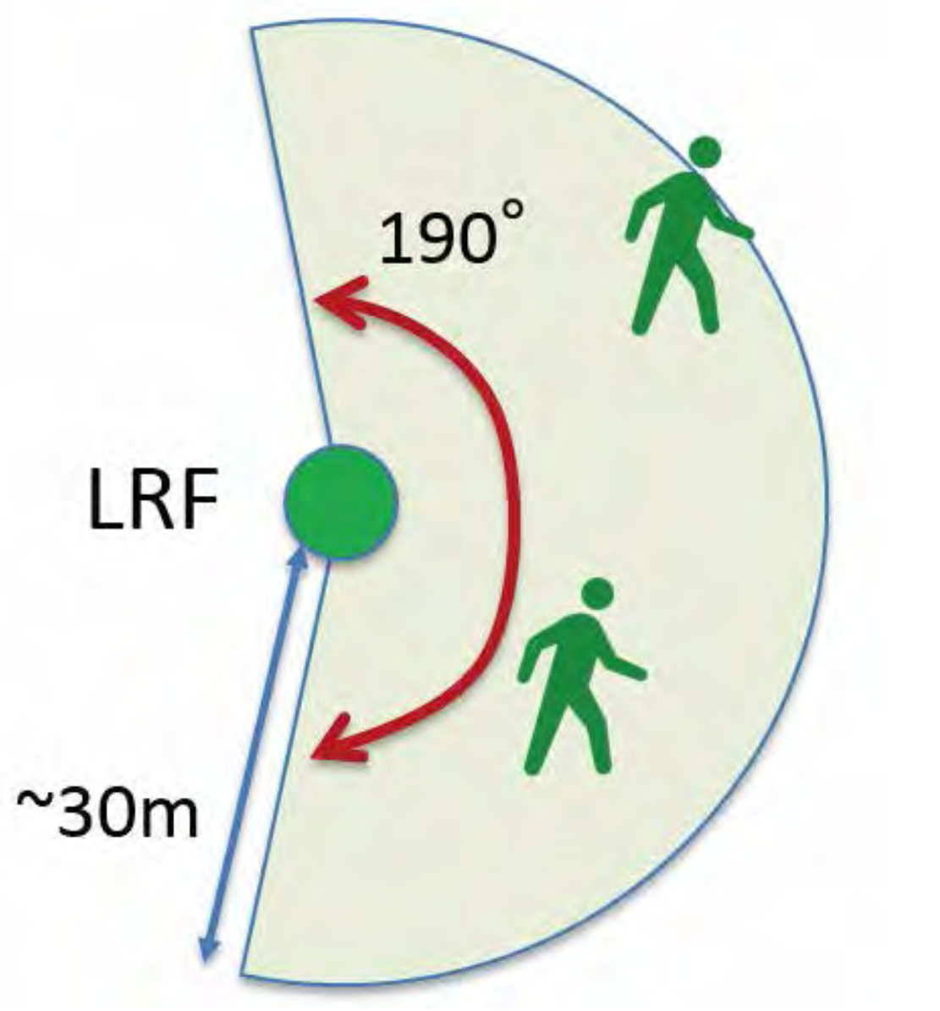


P-Sen: Petit Sensor Box

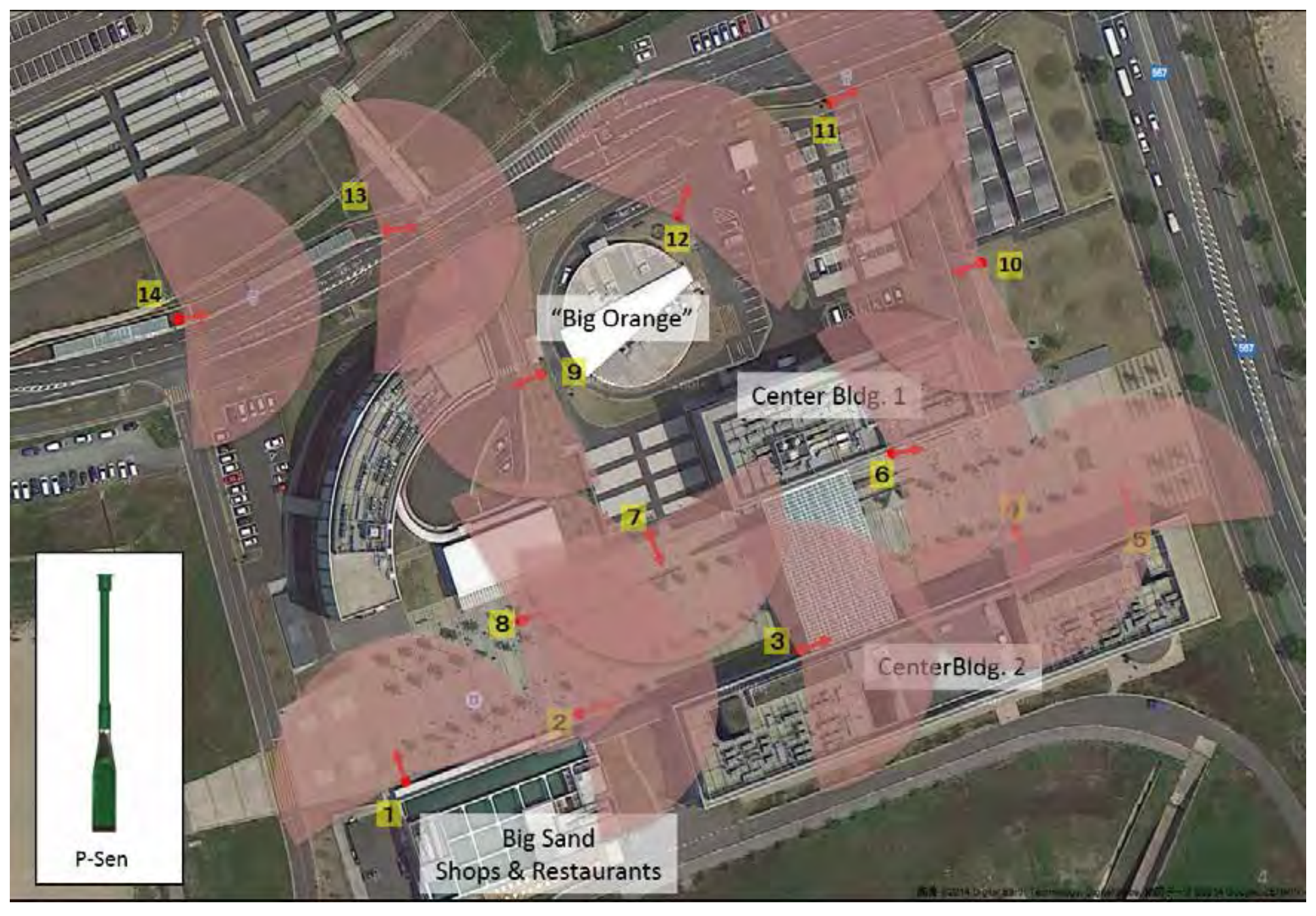


- 14 sensor poles installed in Ito campus, Kyushu Univ
  - To analyze people flow in the campus
- Sensors equipped:
  - Laser Range Finder (LRF) → Used in this work
  - Network camera
  - Temperature/humidity
  - Wifi access point
  - IC card reader

LRFs: distance sensors for people flow



14 P-Sen in the campus and coverage of LRFs



Privacy issues: LRFs are NOT cameras

・they know "there is a passenger at point (x, y) at time t"

・but do NOT know "who is"

People Flow Data from P-Sen and Conversion to Images

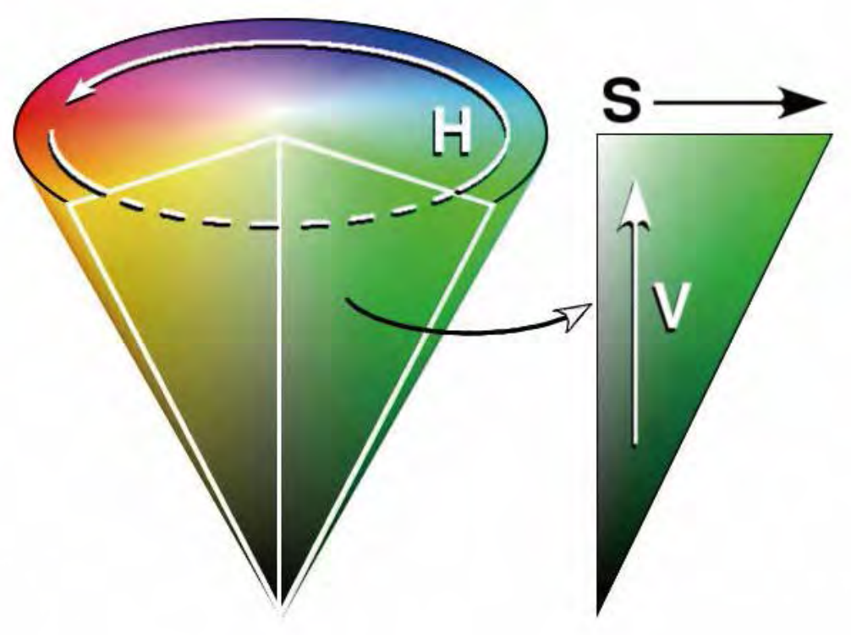
Raw data from P-Sen's LRFs (simplified examples)

P-Sen ID	pid	Time (unixtime)	x (m)	y (m)
1	533371	1438729200037	1.45	3.07
1	533371	1438729200137	1.46	3.14
1	533371	1438729200237	1.38	3.32
1	533371	1438729200337	1.18	3.55
1	533371	1438729200437	1.5	3.53

pid: IDs for passengers for tracking
they are local and volatile to respect privacy

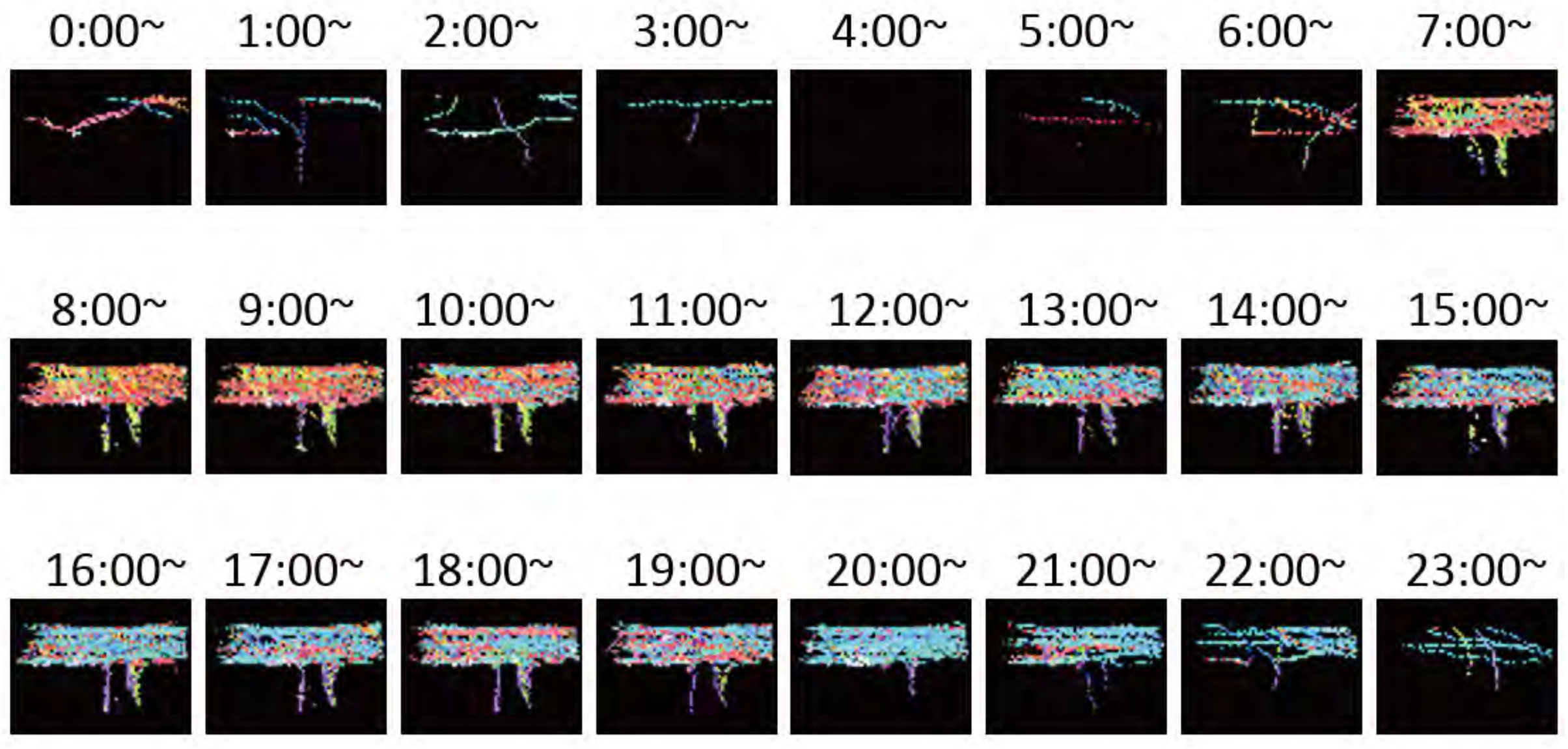
Convert to images

A passenger is recorded on the image per second



Color in HSV
H (hue): Walk direction
S (saturation): Walk speed
V (value): frequency of passengers existence

Images of flow data from P-Sen No. 10 on Aug 5

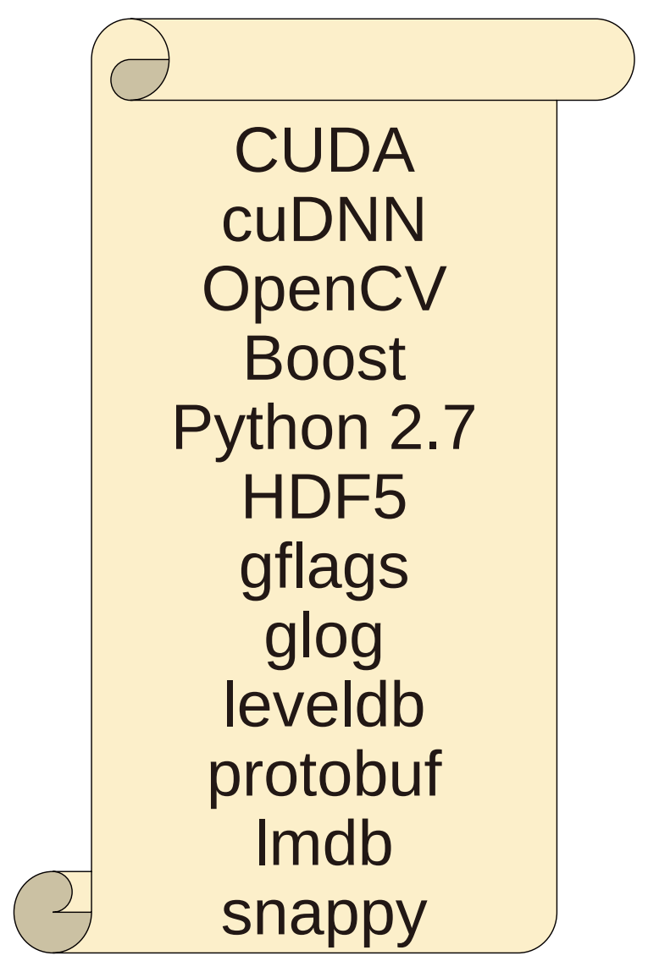


・Now we make an image per hour per sensor

・Size of each image is 64x48 pixels. 1pixel = 1meter

Towards Easy-to-use People Flow Analysis with Caffe on TSUBAME

Caffe by Berkeley is the most well-known deep learning package. However, there are still higher hurdles for usage — merely for installation!



CUDA
cuDNN
OpenCV
Boost
Python 2.7
HDF5
gflags
glog
leveldb
protobuf
lmdb
snappy

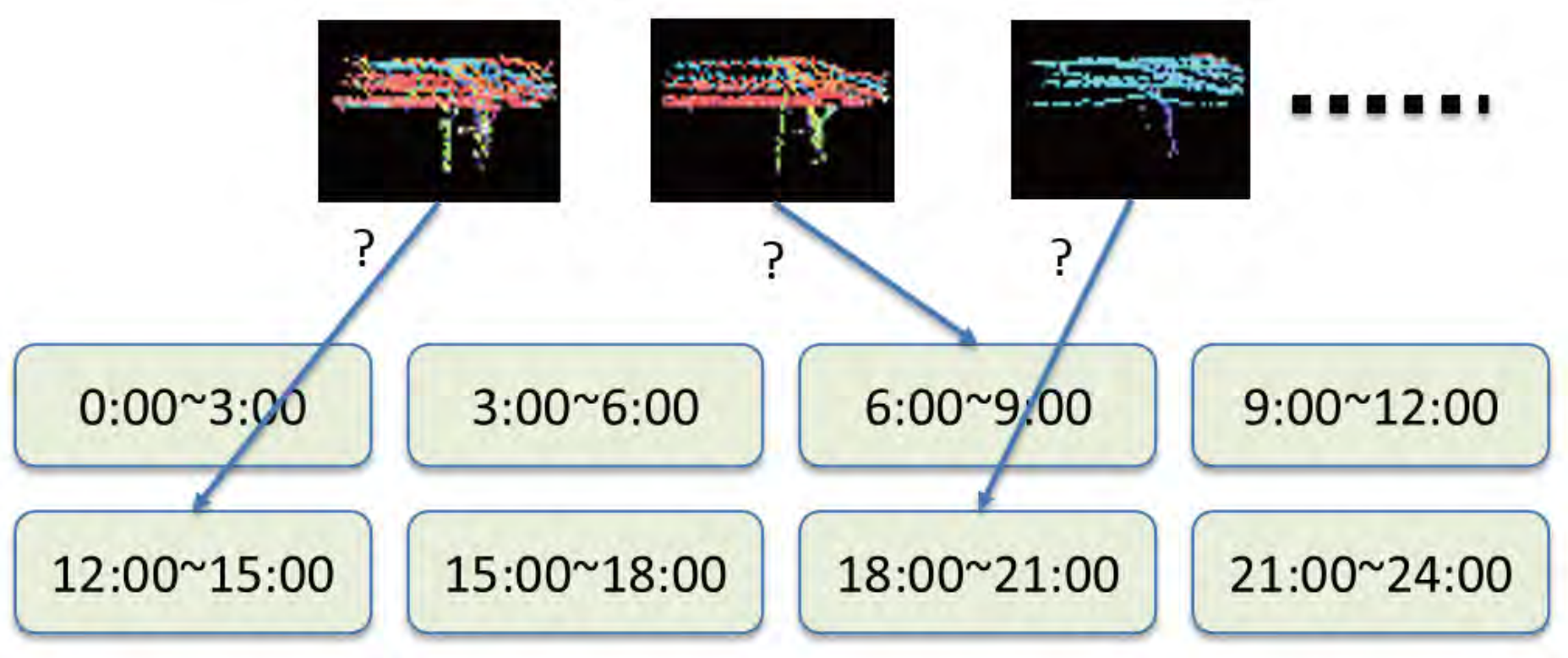
Do I have to install all the software packages…?

Caffe 0.13 is available on TSUBAME2.5!

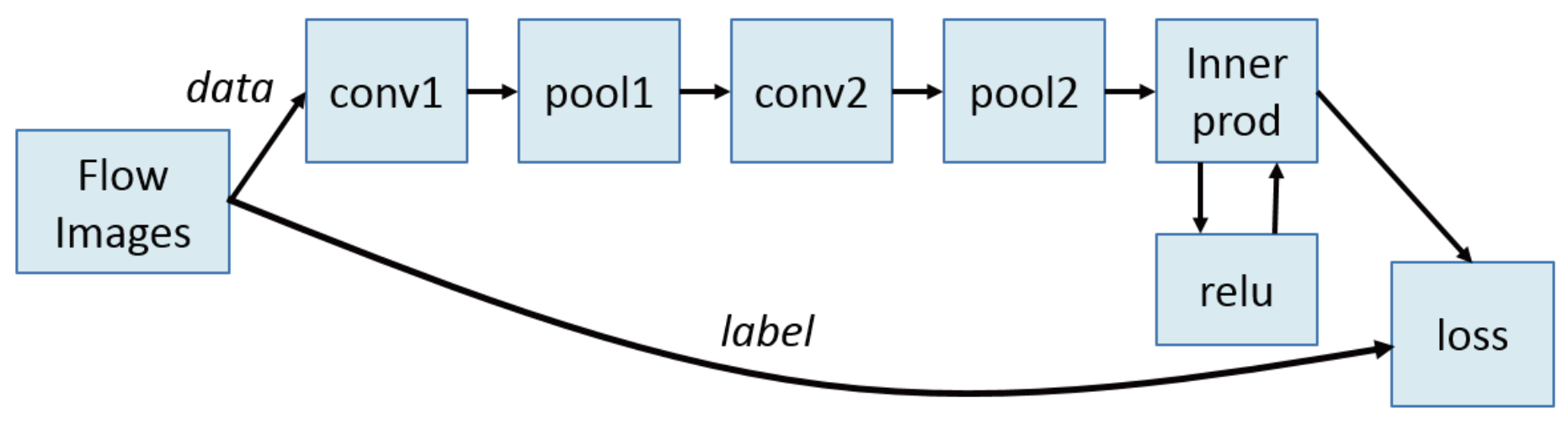
TSUBAME users can analyze their big data with lots of GPUs

Preliminary analysis: Time classification of flow images

Given a new flow image, the classifier tells "this image looks like the flow around XX:00



Neural network used in learning



Based on NN of "MNIST" example in Caffe

Experimental Conditions

Training data: Flow images during Aug 1 – Sep 23
(~1270 images per sensor)

Test data: Flow images during Sep 24 – Sep 30
(168 images per sensor)

1 NVIDIA K20X GPU is used per sensor

Experimental Results

	Training (1000iters)	Testing
X5670 CPU (8cores)	299sec	21sec
K20X GPU	19sec	0.7sec

Accuracy:

Psen1: 69.0%	Psen6: 49.4%	Psen11: 80.9%
Psen2: 60.7%	Psen7: 51.2%	Psen12: 73.8%
Psen3: 45.8%	Psen8: 60.1%	Psen13: 69.0%
Psen4: 53.6%	Psen9: 61.2%	Psen14: 75.0%
Psen5: 61.9%	Psen10: 48.3%	

Collaborative work with Institute of Mathematics for Industry, Kyushu University Center for Co-Evolutional Social Systems, Kyushu University

http://www.gsic.titech.ac.jp/sc15