ThermalIR SME building data:

Thermal IR data was captured once per minute and recorded as 24 x 32 pixel .jpg timestamped files. The data was captured in three different environments within a small medium sized enterprise (SME), the office, meeting room and kitchen. All of the files are labelled positive or negative to indicate whether occupants are present or absent. The labels are based on whether the files are saved in a directory “positive” or “negative” directory. Occupant numbers range from 0 – 6 at any time. These datasets were captured as part of a PhD Thesis entitled Occupant-Centric Energy Management for Small Commercial Buildings by Ruth Ande.

Summary of each dataset:

**Directory name – environment – machine learning test or train – capture period – Number of positive files – Number of negative files**

Data.i1 – office – test – 12:00-23:59 on 23/04 – positive: 443 files – negative: 273 files

Data.i1 – office – train – 00:00-23:59 on 24/04 – positive: 554 files – negative: 849 files

Data.i2 – office – train – 00:00-23:59 on 24/04 – positive: 580 files – negative: 860 files

Data.i2 – office – test AM – 00:00-11:59 on 29/04 – positive: 158 files – negative: 559 files

Data.i2 – office – test PM – 12:00-23:59 on 29/04 – positive: 540 files – negative: 180 files

Data.i3 – office – train – mixed 12:00 23/04 – 23:59 24/04 – positive: 683 files – negative: 757 files

Data.i3 – office – test – mixed 12:00 23/04 – 23:59 24/04 – positive: 340 files – negative: 377 files

Data.i4 – office – train – 12:00 23/04 – 23:59 03/05 – positive: 3191 files – negative: 6790 files

Data.i4 – office – test – 12:00 11/05 – 23:59 14/05 – positive: 1040 files – negative: 1235 files

Data.i5 – office – train – mixed 12:00 23/04 – 23:59 14/05 – positive: 3430 files – negative: 6206 files

Data.i5 – office – test – mixed 12:00 23/04 – 23:59 14/05 – positive: 1143 files – negative: 2068 files

Kitchen – kitchen – NA – 26/05 – 30/05 – positive: 548 files – negative: 2046 files

Meeting room – meeting room – NA – 03/06 – 05/06 – positive: 572 files – negative: 2810 files

Detail for each dataset:

Five data sets captured within the office labelled Data.i1 to Data.i5. The data sets are differentiated based on when the data was captured and data set size. For the purpose of machine learning, Data.i1 – Data.i5 are split into train and test directories, with 66.6% of the data for training.

Two additional test data sets are captured in two additional environments, the kitchen and meeting room.

Data.i1 – Data.i3:

Data.i1 to Data.i3 all comprise of ~1440 train samples and ~717 test samples and were captured over 36 hour period.

Data.i1 was captured 12:00 23/04 - 23:59 24/04 and separated into train and test data based on the image capture time:

* Test data: 12:00-23:59 on 23/04 (the outdoor temperature range 18 - 6 Degrees Celsius)
* Train data: 00:00-23:59 on 24/04 (the outdoor temperature range 21 - 9 Degrees Celsius)

Data.i2 was captured between 24/04 - 29/04. This data set differs from Data.i1 because the train and test data are captured a few days apart when the outdoor temperature varied significantly:

* Test data: 12:00-23:59 on 29/04. (the outdoor temperature range 11 - 4 Degrees Celsius)
* Train data: 00:00-23:59 on 24/04. (The outdoor temperature range 21 - 9 Degrees Celsius)

Data.i3 was comprised of the data previously captured for Data.i1, randomly separated into a train and test data set.

Data.i4 – Data.i5:

Data.i4 and Data.i5 are significantly larger data sets and comprise ~9300 train samples and 2675 test samples.

Data.i4 was captured between 23/4 - 14/05 and separated into train and test sets based on capture time:

* Test data: 12:00 on 11/05 - 23:59 on 14/05. (The outdoor temperature range highs 13 - 10 and lows 5 – 3 Degrees Celsius)
* Train data: 12:00 on 23/04 - 23:59 on 03/05. (The outdoor temperature range highs 21 - 10 and lows 9 – 4 Degrees Celsius).

Data.i5 was captured 12:00 on 23/04 - 23:59 on 22/05 and separated into a randomly mixed train data set.

Kitchen and Meeting room data:

These two data sets are labelled positive and negative, not split into test / train subsets. For the purpose of machine learning, these dataset can be used to test a model trained with office datasets, Data.i1 – Data.i5. Kitchen data was captured between 26/05 – 30/05. Meeting room data was captured between 03/06 – 05/06.