

CHAPTER 1

Evaluation

Now that the sensing system prototype has been realized, we are now able to perform experiments involving the system. In this chapter we create a set of experiments to test the sensor's properties and come to conclusions as to their effect on our ability to detect occupants. We then outline a process from taking raw sensor data and performing occupancy predictions with it. Using that process, we then devise a set of experiments involving recording occupancy data, and finally we process said data and produce accuracy results.

1.1 Sensor Properties

In order to best utilize the Melexis MLX90620 (*Melexis*), we must first understand the properties it exhibits, and their potential affects on our ability to perform person related measurements. These properties can be broadly separated into three different categories; bias, noise and sensitivity. A broad range of data was collected with the sensor in a horizontal orientation using various sources of heat and cold to determine these properties. This experimental setup is described in Figure 1.1 on the following page.

1.1.1 Bias

When receiving no infrared radiation, the sensor should indicate a near-zero temperature. If in such conditions it does not, that indicates that the sensor has some level of bias in its measurement values. We attempted to investigate this bias by performing thermal captures of the night sky. While this does not completely remove the infrared radiation, it does remove a significant proportion of it.

In Figure 1.2 on page 5 the thermal sensor was exposed to the night sky at a capture rate of 1Hz for 4 minutes, with the sensing results combined to create a

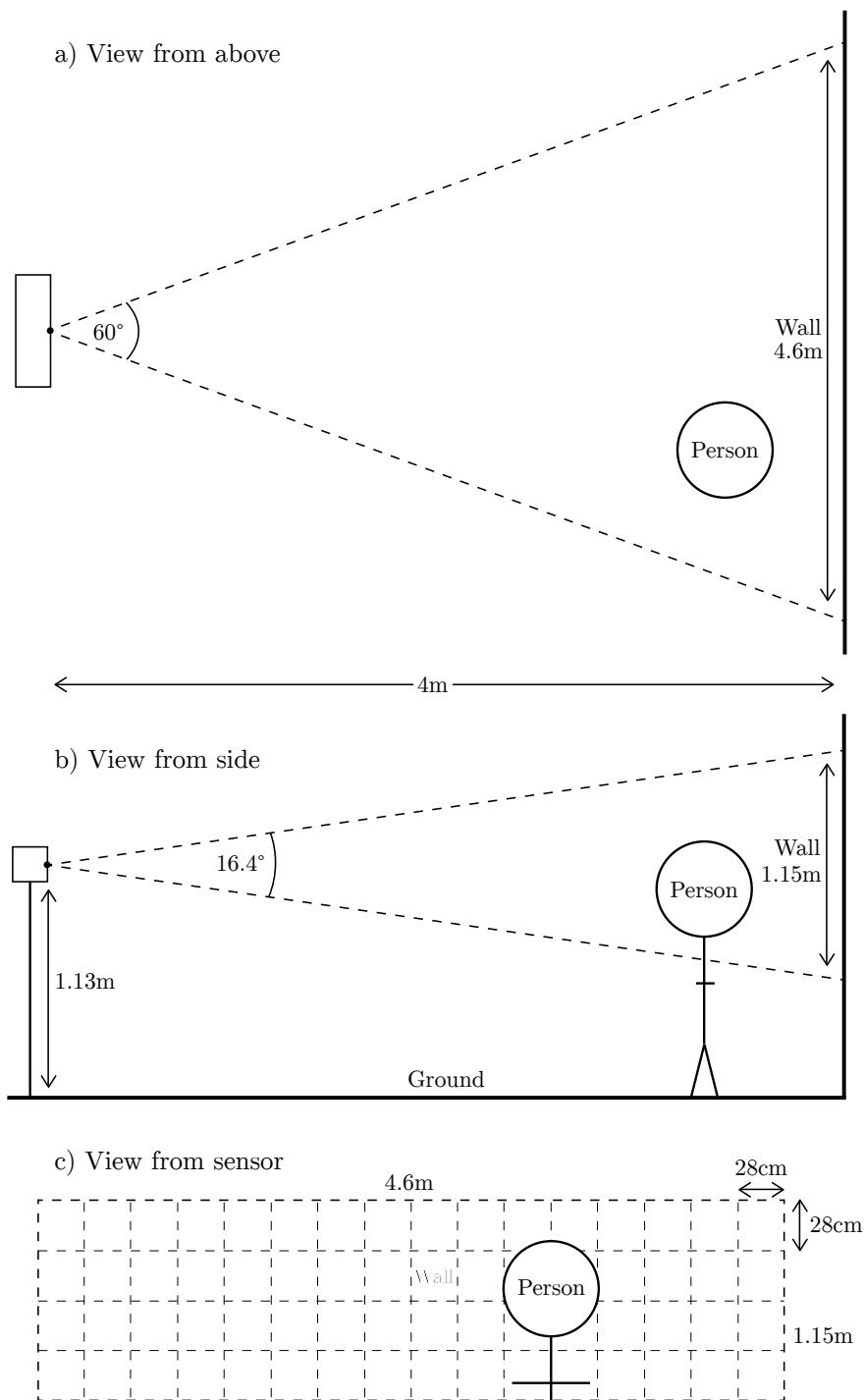


Figure 1.1: Experiment setup to determine sensor properties

set of means and standard deviations to indicate the pixels at “rest”. The average temperature detected was 11.78°C , with the standard deviation remaining less than 0.51°C over the entire exposure period. The resultant thermal map shows that pixels centered around the four “primary” pixels in the center maintain a similar temperature around 9°C , with temperatures beginning to deviate as they became further from the center.

The most likely cause of this bias is related to the physical structure of the sensor. The *Melexis* is a rectangular sensor which has been placed inside a circular tube. Due to this physical arrangement, the sides of this rectangular sensor will be significantly closer to these edges than the center. If these sides are at an ambient temperature higher than the measurement data (as they were in this case) thermal radiation from the sensor package itself could provide significant enough to cause the edges to appear warmer than the observed area of the sky. Such issues with temperature could be controlled for using a device that cools the sensor package to below that of the ambient temperature being measured, however, this is not a concern in this project, as the method of calculating a thermal background will compensate for any such bias as long as it remains constant.

1.1.2 Noise

One of the features of the *Melexis* is the ability to sample the thermal data and a variety of sample rates between 0.5Hz and 512Hz . However, it was noted in early experimentation that a higher sample rate resulted in an increase in the noise contained within the resultant images. As our experiments focus on separating objects of interest from a thermal background, it is important to determine the maximum level of noise tolerable before our algorithms are unable to separate the background from the objects of interest.

Figure 1.3 on page 7 plots one of the central pixels of the sensor in a scenario where it is merely viewing a background (shown in green), and when it is viewing a person (shown in red), at the 5 different sample rates achievable with the current hardware. We can see in these plots that the data becomes significantly more noisy as the sample rate increases, and we can also determine that the sensor uses a form of data smoothing at lower sample rates, as the variance in data increases with sample rate.

If the sample rate were to increase, it is likely that the ability for the sensing system to disambiguate between objects of interest and the background would diminish. However, in the current project, even the slowest sampling rate of 0.5Hz is sufficient, as occupancy estimations at a sub-second level present little

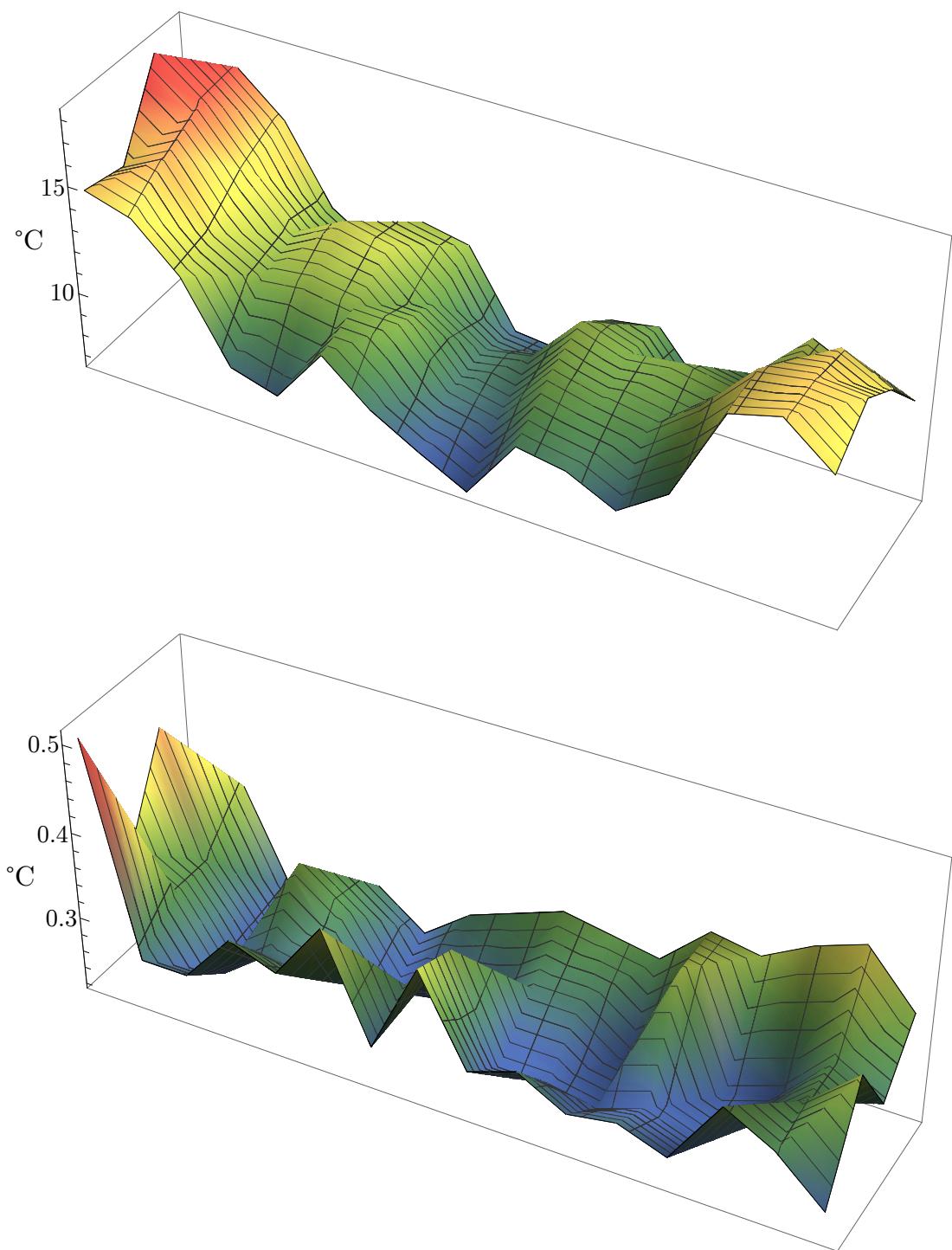


Figure 1.2: Plot of pixel means (top) and standard deviations (bottom) at rest

additional value and would require significant reforms in the efficiency of the software used.

1.1.3 Sensitivity

The *Melexis* is a sensor composed of 64 independent non-contact digital thermopiles, which measure infrared radiation to determine the temperature of objects. While they are bundled in one package, Figure 1.4 on page 8 shows that they are in fact wholly independent sensors placed in a grid structure. This has important effects on the properties of the data that the *Melexis* produces.

Figure 1.5 on page 8 shows a graph of the temperatures of the top row of 16 pixels of the *Melexis* as a hot object is moved from left to right at an approximately similar speed. One of the most interesting phenomena in this graph is the apparent extreme variability of the detected temperature of the object as it moves “between” two different pixels; there is a noticeable drop in the objects detected temperature. Further analysis of each of the pixel’s lines on the graph shows each pixel exhibiting a bell-curve like structure, with the detected temperature increasing from the baseline and peaking as the object enters the center of the pixel, and the detected temperature similarly decreasing as the object leaves the center.

This phenomenon has several possible causes. One likely explanation is that each individual pixel detects objects radiating at less favorable angles of incidence to be colder than they actually are: As the object enters a pixel’s effective field of view, it will radiate into the pixel at an angle that is at the edge of the pixel’s ability to sense, with this angle slowing decreasing until the hot object is directly radiating into the pixel’s sensor, causing a peak in the temperature reading. As the object leaves the individual elements field of view, the same happens in reverse.

While interesting, this phenomenon has little consequence to the effectiveness of the techniques used, as in experimental conditions the sensor will not be sufficiently distant that humans could be detected as single pixels. However, this phenomenon could be leveraged in future work to perform sub-pixel localization, discussed later on.

1.2 Energy Efficiency

TODO: This section will contain experimental descriptions relating to the power consumption of the prototype.

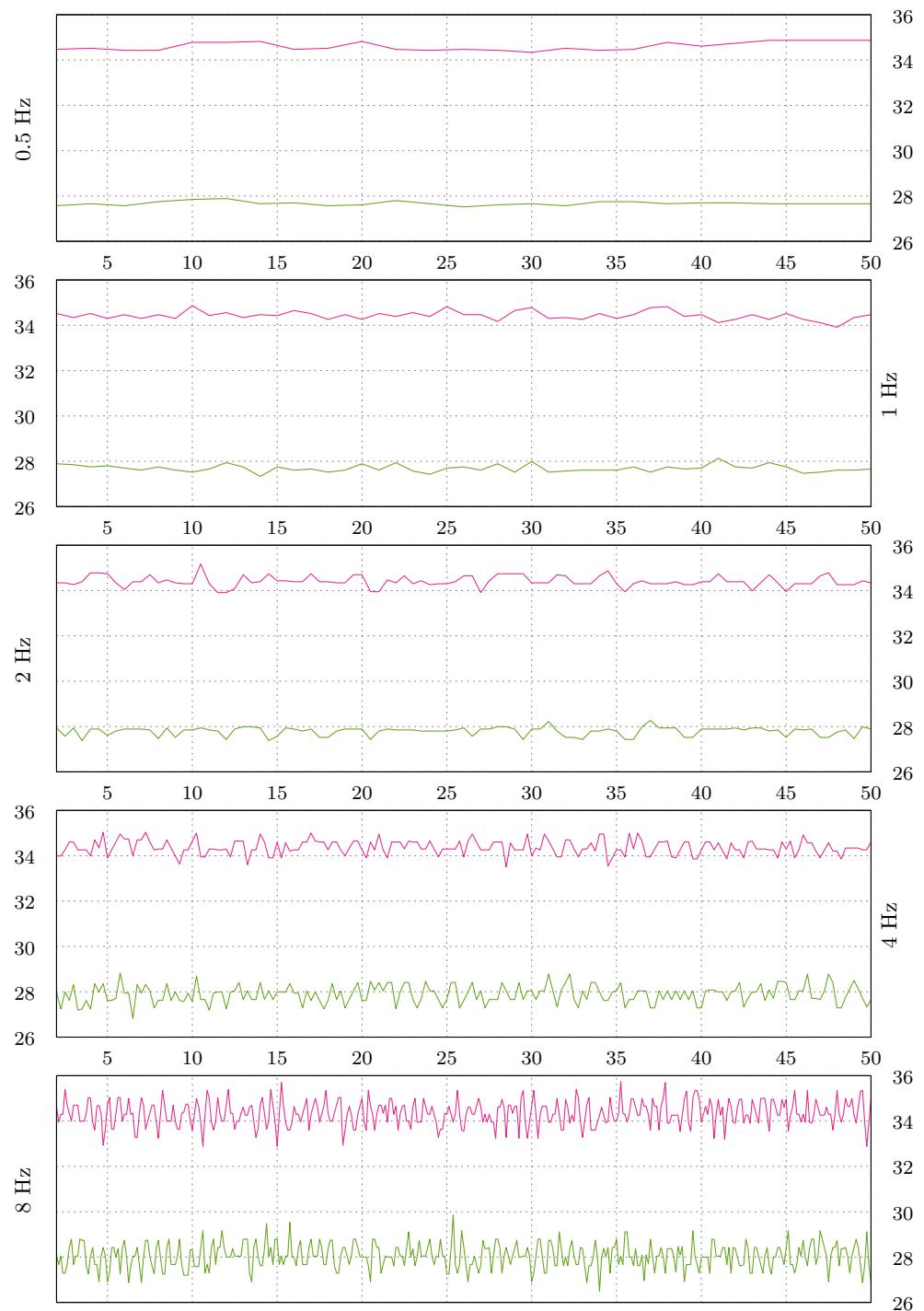


Figure 1.3: Noise levels of occupants (higher) and background (lower) at various sampling speeds

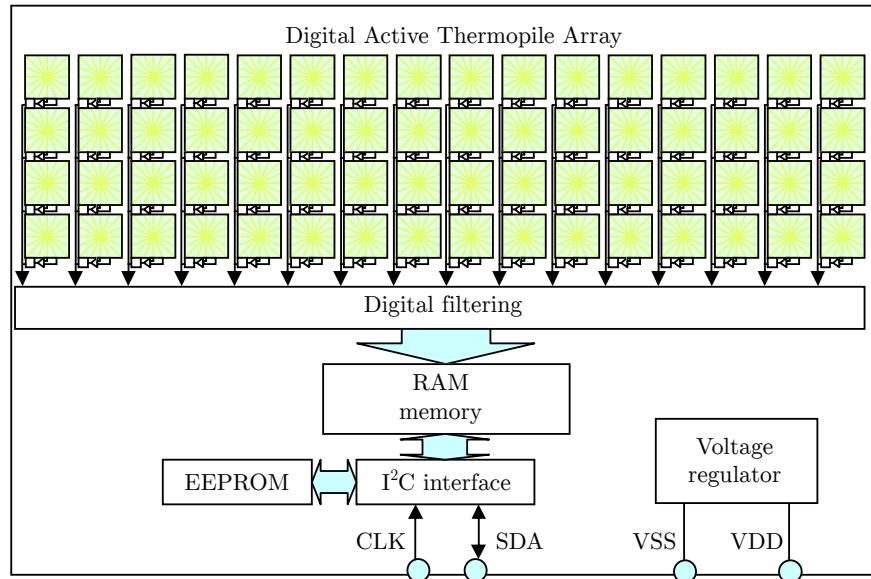


Figure 1.4: Block diagram for the *Melexis* taken from datasheet [2]

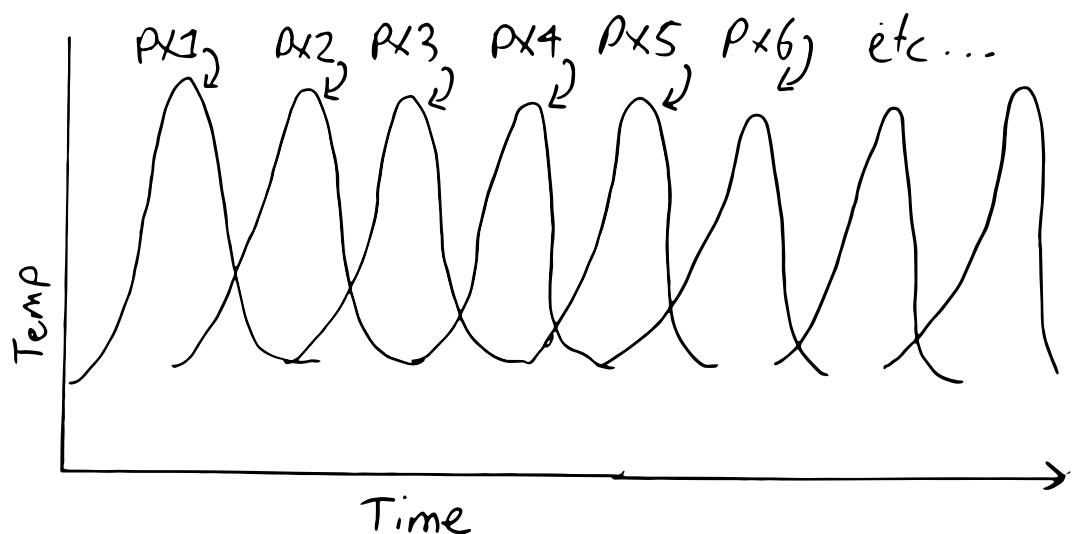


Figure 1.5: Different *Melexis* pixel temperature values as hot object moves across row

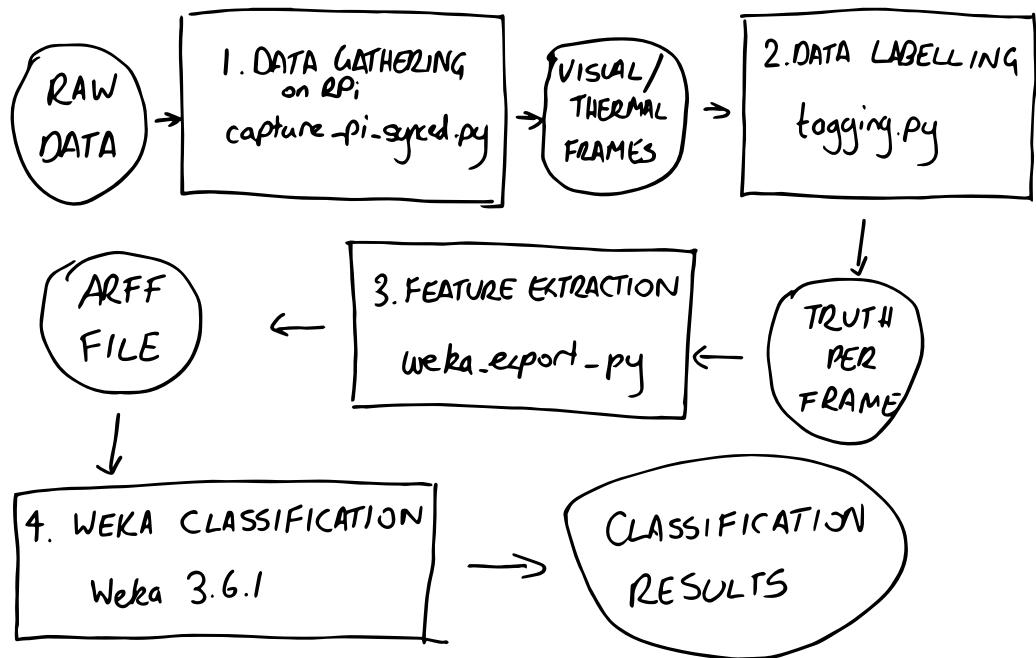


Figure 1.6: Flowchart of processing

1.3 Classification Accuracy

With the prototype, it is now possible to utilize the prototype to gather both thermal and visual data in a synchronized format. This data can be collected and used to determine the effectiveness of the human counting algorithms used. Due to the prototype's technical similarity to Thermosense [1], a similar set of experimental conditions will be used, with a comparison against Thermosense being used as a benchmark. To this end, several experiments were devised, each of which had its data gathered and processed in accordance with the same general process, outlined in Figure 1.6.

1.3.1 Data gathering

As the camera and the Arduino are directly plugged into the Raspberry Pi, all data capture is performed on-board through SSH, with the data being then copied off the Pi for later processing. To perform this capture, the main script used is `capture_pi_synced.py`.

`capture_pi_synced.py` takes two parameters on the command line; the name of the capture output, and the number of seconds to capture. By default, it

always captures at 2Hz. The script initializes the `picamera` library, then passes a reference to it to the `capture_synced` function within the `Visualizer` class. The class will then handle the sending of commands to the Arduino to capture data in concert with taking still frames with the Raspberry Pi's camera.

When the script runs, it creates a folder with the name specified, storing inside a file named `output_thermal.hcap` containing the thermal capture, and a sequence of files with the format `video-%09d.jpg`, corresponding to each visual capture frame.

1.3.2 Data labeling

Once this data capture is complete, the data is copied to a more powerful computer for labeling. The utility `tagging.py` is used for this stage. This script is passed the path to the capture directory, and the number of frames at the beginning of the capture that are guaranteed to contain no motion. This utility will display frame by frame each visual and thermal capture together, as well as the computed feature vectors (based on a background map created from the first n frames without motion).

The user is then required to press one of the number keys on their keyboard to indicate the number of people present in this frame. This number will be recorded in a file called `truth` in the capture directory. The next frame will then be displayed, and the process continues. This utility enables the quick input of the ground truth of each capture, making the process more efficient.

1.3.3 Feature extraction and data conversion

Once the ground truth data is available, it is now possible to utilize the data to perform various classification tests. For this, we use version 3.7.12 of the open-source Weka toolkit [3], which provides easy access to a variety of machine learning algorithms and the tools necessary to analyze their effectiveness.

To enable the use of Weka, we export the ground truth and extracted features to a Comma Separated Value (CSV) file for processing. `weka_export.py` takes two parameters, a comma-separated list of different experiment directories to pull ground truth and feature data from, and the number of frames at the beginning of each capture that can be considered as "motionless." With this information, a CSV-file file is generated on which the heading indicating the attribute names is added for Weka to recognize.

1.3.4 Running Weka Tests

Once the CSV file is generated, it is then possible to open the file in Weka for processing. Weka provides a variety of algorithms, but we choose a specific subset of algorithms based on those present in the Thermosense paper [1], as well others that we believe adequately represent the different approaches to classification.

Type	Attribute	Weka Class & Parameters
Neural Network (ANN)	Nominal, Numeric	weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 15 -S 0 -E 20 -H 5
<i>k</i> -nearest Neighbors (KNN)	Nominal, Numeric	weka.classifiers.lazy.IBk -K 5 -W 0 -F -A "weka.core.neighboursearch.LinearNNSearch -A \\"weka.core.EuclideanDistance -R first-last\\"
Naive Bayes	Nominal	weka.classifiers.bayes.NaiveBayes
Support Vector Machine (SVM)	Nominal	weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -C 250007 -E 1.0"
Decision Tree	Nominal	weka.classifiers.trees.J48 -C 0.25 -M 2
Entropy Distance	Nominal, Numeric	weka.classifiers.lazy.KStar -B 20 -M a
Linear Regression	Numeric	weka.classifiers.functions.LinearRegression -S 0 -R 1.0E-8
0-R	Nominal, Numeric	weka.classifiers.rules.ZeroR

Table 1.1: Weka parameters used for classifications

For those tests that are “nominal,” the `npeople` attribute was interpreted as nominal using the “NumericToNominal” filter, which creates a class for each value deleted in `npeople`’s columns. For those tests that are “numeric,” `npeople` is left unchanged, as by default all CSV import attributes are interpreted as such. For all tests where not specifically instructed, we use 10-fold cross-validation to validate our results.

As the data we are using is based on real experiments, the number of frames which are classified as each class may be unbalanced, which could cause the classification results to be affected. To that end, for each classification technique, we both classify the data in its raw, unbalanced form, and we also uniformly resample the `npeople` parameter using `weka.filters.supervised.instance.Resample -B 1.0 -S 1 -Z 100.0` in the pre-processing stage.

To help maximize the efficiency of the classification task, we use the Weka Knowledge Flow constructor to generate an encompassing flow that accepts an input CSV file of the raw data, and performs all resampling, numeric and nominal classification, returning a text file with the results of each of the different classification techniques run. The knowledge flow’s struture can be seen in . To enable maximum efficiency, the input and output elements of this flow are set to the environmental variables `UnifiedFlow.InputCSV` and `UnifiedFlow.OutputCSV`, a Jython script, `run_flow.py`, then sets those environmental variables to input and output file names, then calls the flow using Weka’s Java API. After this is complete, the script then runs a series of regexes on the output text data to generate summary spreadsheets with the relevant values.

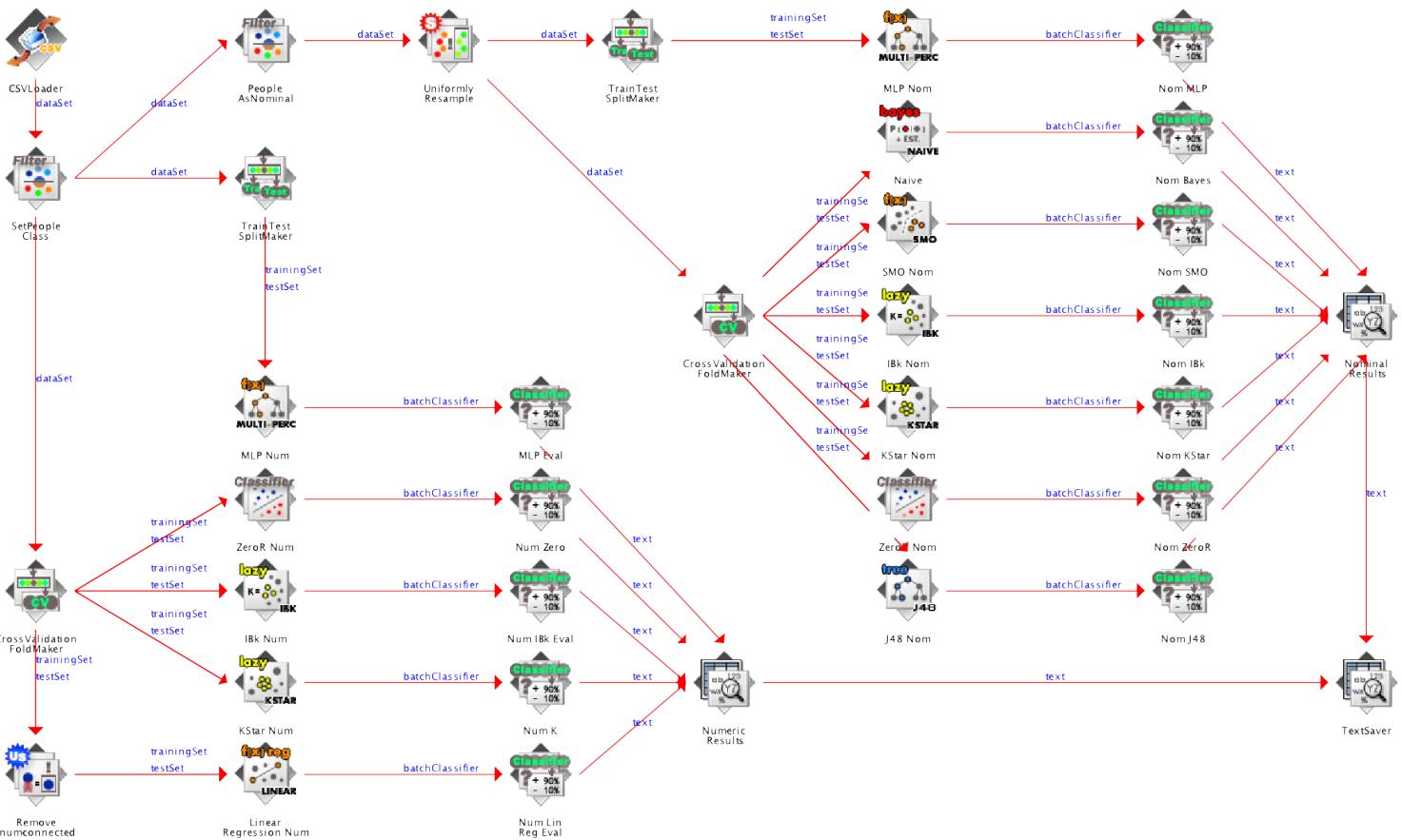


Figure 1.7: Unified Knowledge Flow for all experiments

1.3.5 Classifier Experiment Set 1

In our first set of experiments, a scene was devised in accordance with Figure 1.8 on page 17 that attempted to sense people from above, as did Thermosense. The prototype was set up on the ceiling, pointing down at a slight angle. For ease of use, the prototype was powered by mains power, and was networked with a laptop for command input and data collection via Ethernet. This set of experiments involved between zero and three people being present in the scene, moving in and out in various ways in accordance with the script in Table 1.2 on the next page.

1. (Remained standing) One person walks in, stands in center, walks out of frame. (sub-experiment 1)
2. (Remained standing) One person walks in, joined by another person, both stand there, one leaves, then another leaves. (sub-experiment 2)
3. (Remained standing) One person walks in, joined by one, joined by another, all stand there, one leaves, then another, then another. (sub-experiment 3)
4. (Remained standing) Two people walk in simultaneously, both stand there, both leave simultaneously. (sub-experiment 4)
5. (Sitting) One person walks in, sits in center, moves to right, walks out of frame. (sub-experiment 5)
6. (Sitting) One person walks in, joined by another person, both sit there, they stand and switch chairs, one leaves, then another leaves. (sub-experiment 6)
7. (Sitting) One person walks in, joined by one, joined by another, they all sit there, one leaves, one shuffles position, then another leaves, then another. (x2) (sub-experiment 7, 8)
8. (Sitting) Two people walk in, both sit there, both leave. (sub-experiment 9)

Table 1.2: Experiment Set 1 Script

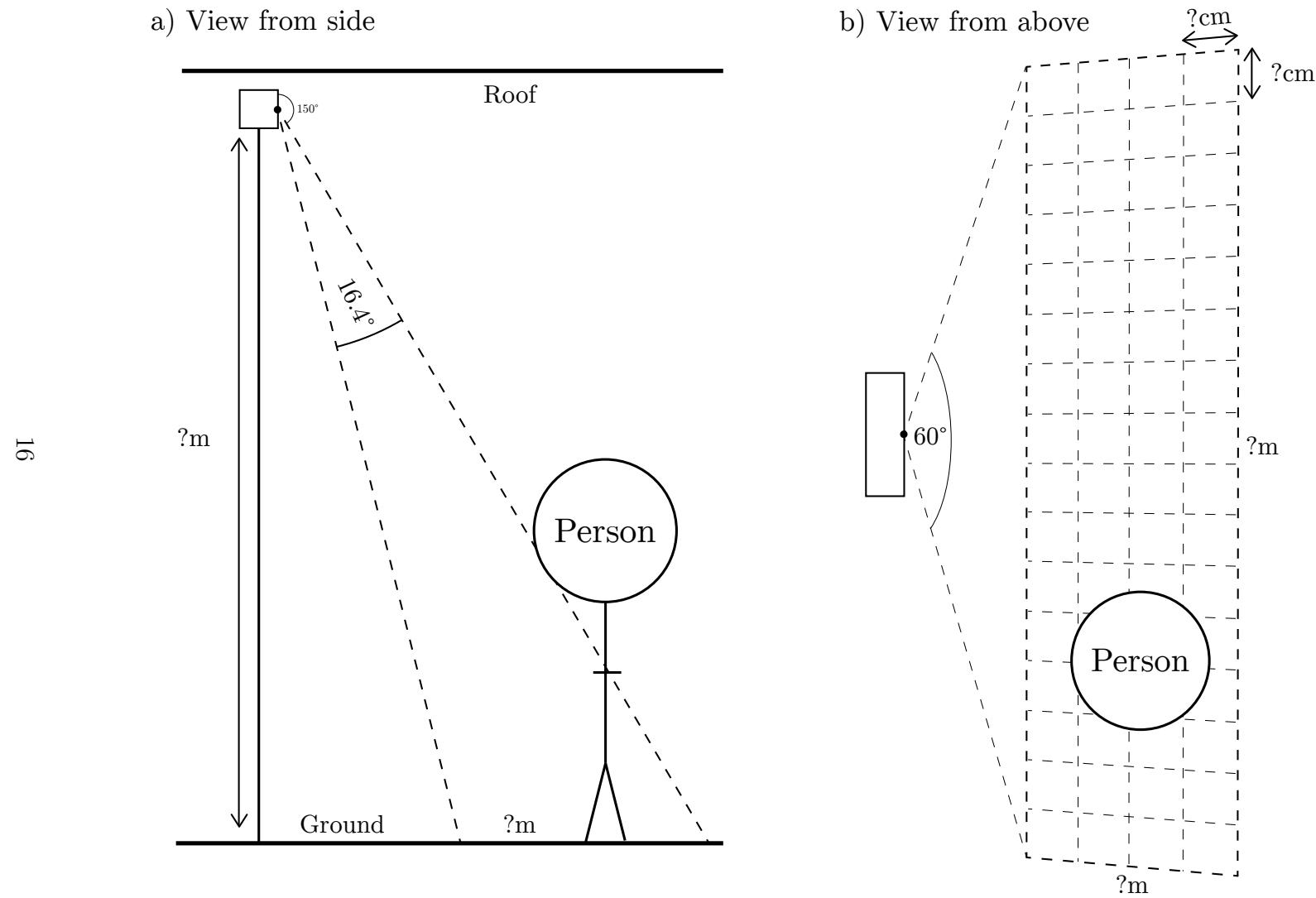


Figure 1.8: Classifier Experiment Set 1 Setup

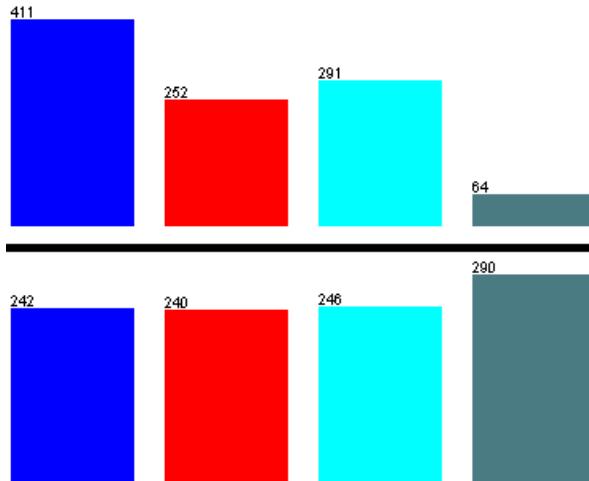


Figure 1.9: Experiment Set 1 Class Distribution Before and After Weka Resampling

In these experiments people moved slowly and deliberately, making sure there were large pauses between changes of action. The people involved were of average height, wearing various clothing. The room was cooled to 18 degrees for these experiments.

Each experiment was recorded with a thermal-visual synchronization at 1Hz over approximately 60-120 second intervals. Each experiment had 10-15 frames at the beginning where nothing was within the view of the sensor to allow the thermal background to be calculated. Each frame generated from these experiments was manually tagged with the ground truth value of its occupancy using the script mentioned previously.

The resulting features and ground truth were combined and exported to CSV allowing the Weka machine learning program to analyze them. This data was analyzed with the feature vectors always being considered numeric data and with the ground truth considered both numeric and nominal. All previously mentioned classification algorithms were run against the data set.

Experimental results from the first set of experiments were overall excellent, results from them can be seen in Table 1.3 on the following page. In the unbalanced results when including zero, an accuracy of 82.9% was observed, and in the balanced results, 78.2%. When excluding zero, the unbalanced results achieved 82.5% in the unbalanced and up to 84.9% in the balanced results.

Between the unbalanced and balanced classes when including zero, the ranking of different algorithms remained approximately the same, and consistently dropped in accuracy, with the exceptions being the SVM technique, which in-

Classifier	RMSE		% Excl. 0 Incl. 0		R^2 Excl. 0 Incl. 0	
	Excl. 0	Incl. 0	Excl. 0	Incl. 0	Excl. 0	Incl. 0
Thermosense Replication						
KNN (Nominal)	-	0.364	-	65.65	-	-
KNN (Numeric)	-	1.1235	-	-	-	0.3766
MLP	0.592	-	-	-	0.687	-
Lin Reg	0.525	-	-	-	0.589	-
Nominal Balanced						
C4.5	0.289	0.290	84.96	77.56	-	-
K*	0.296	0.293	84.27	78.25	-	-
MLP	0.359	0.320	70.86	72.13	-	-
Bayes	0.409	0.368	63.76	61.52	-	-
SVM	0.410	0.386	65.98	56.89	-	-
0-R	0.471	0.433	32.48	24.61	-	-
Numeric						
K*	0.423	0.550	-	-	0.760	0.828
0-R	0.651	0.972	-	-	-0.118	-0.129
Nominal Unbalanced						
C4.5	0.314	0.288	82.39	82.91	-	-
K*	0.304	0.285	82.56	82.61	-	-
MLP	0.362	0.286	77.14	78.69	-	-
Bayes	0.405	0.352	63.59	66.21	-	-
SVM	0.398	0.380	67.18	57.47	-	-
0-R	0.442	0.415	49.74	40.37	-	-

Table 1.3: Classifier Experiment Set 1 Results

creased in accuracy by about 1% in that instance. The drop in accuracy can be explained mostly by an over-representation of the zero class within the underbalanced data, as well as an underrepresentation of the three class (see Figure 1.9 on page 18). This is conformed by the fact that in the zero-excluded data, there is much less difference in the balanced and unbalanced set. These biases would enable classes to over-predict and under-predict these two classes respectively and achieve an artificially higher accuracy as a result. As discussed in the Methods, we performed re-sampling inside Weka to compensate for this.

For the numeric representation of the number of people, accuracy was consistently poor. From this data, we can see that all three classifiers used performed consistently poorly, with the Root Mean Square Errors being consistently double or more of comparable nominal results, and with correlation coefficients (R^2) indicating poor (or in the case of KNN) very poor correlations.

The two highest accuracy classifiers, C4.5 and K*, achieved quite similar results for both the balanced and unbalanced data while being quite different in implementation.

TODO: Discuss and compare this to ZeroR's RMSE. It's RMSE is quite close to some results, is this bad?

1.3.5.1 Individual sub-experiment results

In addition to the above aggregate classification results, in which each of the nine sub-experiments results are combined and fed into the classifier, each of the sub-experiments has been individually classified with each of the six balanced nominal classifiers above. The results for these classifications can be seen in Table ?? on page ?? and Table 1.4 on the next page.

TODO: We talk about points of interest in the sub-experiment results and see if we can draw any useful conclusions from them.

	1	2	3	4	5	6	7	8	9	Avg
Nominal										
KNN	0.584	0.422	0.375	0.425	0.651	0.352	0.396	0.338	0.244	<i>0.421</i>
C4.5	0.342	0.270	0.278	0.414	0.456	0.267	0.318	0.288	0.250	<i>0.320</i>
K*	0.305	0.249	0.260	0.412	0.407	0.245	0.299	0.265	0.196	<i>0.293</i>
MLP	0.345	0.275	0.272	0.399	0.466	0.232	0.389	0.246	0.242	<i>0.318</i>
Bayes	0.391	0.359	0.290	0.435	0.473	0.276	0.381	0.306	0.243	<i>0.351</i>
SVM	0.447	0.447	0.379	0.444	0.602	0.378	0.380	0.377	0.335	<i>0.421</i>
Numeric										
0-R	0.500	0.471	0.433	0.472	0.500	0.471	0.433	0.433	0.472	<i>0.465</i>
KNN	0.726	0.707	1.044	0.934	0.593	0.531	0.899	0.585	0.829	<i>0.761</i>
K*	0.256	0.422	0.598	0.806	0.377	0.432	0.586	0.627	0.472	<i>0.508</i>
Lin Reg	0.270	0.635	0.623	0.821	0.422	0.508	0.589	0.708	0.570	<i>0.572</i>
MLP	0.379	0.460	0.500	0.868	0.399	0.306	0.790	0.490	0.406	<i>0.511</i>
0-R	0.409	0.762	1.009	0.986	0.507	0.829	1.014	1.043	1.012	<i>0.841</i>
Avg	<i>0.413</i>	<i>0.457</i>	<i>0.505</i>	<i>0.618</i>	<i>0.488</i>	<i>0.402</i>	<i>0.539</i>	<i>0.475</i>	<i>0.439</i>	

Table 1.4: Classifier Experiment Set 1 Individual Sub-experiment RMSEs

1.4 Thermosense Comparison and Discussion

To aid in comparison with the Thermosense algorithm, in our experiment sets we chose three techniques similar to those used in Thermosense; an Artificial Neural Network (Multilayer Perceptron in Weka), k -nearest Neighbors (IBk in Weka) and Linear Regression. Our results for our approximations to their algorithms can be seen in the Thermosense Replication section of Table 1.3 on page 19. We also use an experimental setup with 0–3 or 1–3 people, depending on the classifier, just as the Thermosense experiments do. For KNN, it was ambiguous if Thermosense used numeric or a nominal class attributes in their experiments, so we have performed both options above. The specifics of this is discussed in the Methods chapter.

For those experiments where we tried to specifically replicate Thermosense' results, we found consistently poor results that did not come close to meeting the R^2 or RMSE values that Thermosense claimed. Their R^2 values were in the 0.8 range, while ours range from around 0.3–0.7. This can be put down to a multitude of factors, with one likely explanation being that differences with the Melexis MLX90620 (*Melexis*)'s narrower field of view being difficult for Thermosense' specific algorithmic selections to deal with.

Our Linear Regression (Equation (1.1)) underperforms particularly when compared to Thermosense' (Equation (1.2)), as it fails to find an adequate way to weight the two variables, instead opting to basically exclude them from consideration by picking very small weights and adding a large constant factor. We get a correlation of $R^2 = 0.589$ vs Thermosense' $R^2 = 0.858$.

$$n = 0.0456a + -0.024s + 1.1772 \quad (1.1)$$

$$n = 0.141a + -0.051s + 0.201 \quad (1.2)$$

However, for those algorithms we chose ourselves to test, we found results that were comparable or even better than Thermosense. Our nominally classed C4.5 decision tree achieved an balanced RMSE excluding zero of 0.289, compared to Thermosense' best result of 0.346. In the numeric classes, the K* implementation achieved correlations in the same ballpark of 0.760 (excl. 0) and 0.828 (incl. 0), compared to Thermosense' best 0.906 for their ANN.

However, none of the techniques used by Thermosense proved to be the best out of those algorithms tried. The Neural Network and k -nearest Neighbors techniques represent the middle-of-the-road of our results, with both being bested by the C4.5 and K* algorithms, which produced RMSEs of 0.289 and 0.298

respectively. Both these results are significantly better than Thermosense's best RMSE of 0.346 for k -nearest Neighbors, with our C4.5 algorithm representing a 28% improvement over that technique.

Bibliography

- [1] BELTRAN, A., ERICKSON, V. L., AND CERPA, A. E. ThermoSense: Occupancy thermal based sensing for HVAC control. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings* (2013), ACM, pp. 1–8.
- [2] MELEXIS. Datasheet IR thermometer 16X4 sensor array MLX90620. <http://www.melexis.com/Asset/Datasheet-IR-thermometer-16X4-sensor-array-MLX90620-DownloadLink-6099.aspx>, 2012. Accessed: 2015-01-07.
- [3] UNIVERSITY OF WAIKATO. Weka. <http://www.cs.waikato.ac.nz/ml/weka/>. Accessed: 2015-03-10.