A Collaborative Approach to In-Place Sensor Calibration

Vladimir Bychkovskiy, Seapahn Megerian, Deborah Estrin, and Miodrag Potkonjak

Department of Computer Science,
University of California, Los Angeles
{vladimir, destrin, seapahn, miodrag}@cs.ucla.edu

Abstract. Numerous factors contribute to errors in sensor measurements. In order to be useful, any sensor device must be calibrated to adjust its accuracy against the expected measurement scale. In large-scale sensor networks, calibration will be an exceptionally difficult task since sensor nodes are often not easily accessible and manual device-by-device calibration is intractable. In this paper, we present a two-phase post-deployment calibration technique for large-scale, dense sensor deployments. In its first phase, the algorithm derives relative calibration relationships between pairs of co-located sensors, while in the second phase, it maximizes the consistency of the pair-wise calibration functions among groups of sensor nodes. The key idea in the first phase is to use temporal correlation of signals received at neighboring sensors when the signals are highly correlated (i.e. sensors are observing the same phenomenon) to derive the function relating their bias in amplitude. We formulate the second phase as an optimization problem and present an algorithm suitable for localized implementation. We evaluate the performance of the first phase of the algorithm using empirical and simulated data.

Keywords: Sensor calibration, distributed calibration, consistency maximization, sensor networks, distributed algorithms, in-network processing, calibration routing

1 Introduction

The recent advent of sensor networks as enablers for completely new classes of applications, has not only captured the imagination of many a scientist and engineer in many a domain, but has also sparked the recognition of new classes of problems for the developers of sensor network systems and technology. Data inaccuracy and imprecision are two examples of inevitable challenges when dealing with the measurement of physical phenomena. These errors must be dealt with properly if sensor data are to be useful. Furthermore, these errors must ultimately be dealt with in the network to enable collaborative signal processing. Calibration traditionally refers to the process of correcting systematic errors

F. Zhao and L. Guibas (Eds.): IPSN 2003, LNCS 2634, pp. 301–316, 2003.

(biases) in sensor readings. The term has also often been used in reference to the procedure by which the raw outputs of sensors are mapped to standardized units. Traditional single-sensor calibration often relies on providing a specific stimulus with a known result, thus creating a direct mapping between sensor outputs and expected values. Consequently, such calibration for a sensor is often subject to specific ranges and operating condition restrictions, which are reported in the manufacturer specifications of the sensor. This type of calibration can be performed at the factory, during the production stage, and/or manually in the field. In addition to component level calibrations, sensors usually must be calibrated at the device level when used as part of a measurement system. Moreover, re-calibration is usually required in order to ensure proper operation of a measurement device, as ageing and other factors impact sensors and measurement hardware over time.

However, with large scale sensor networks, manual, single-sensor calibration schemes will not work well. In addition to the obvious scaling issues, the following are examples of factors that will also hinder such methods:

- Limited access to the sensors in the field
- Complex dynamic environmental effects on the sensors
- Sensor drift (age, decay, damage, etc)

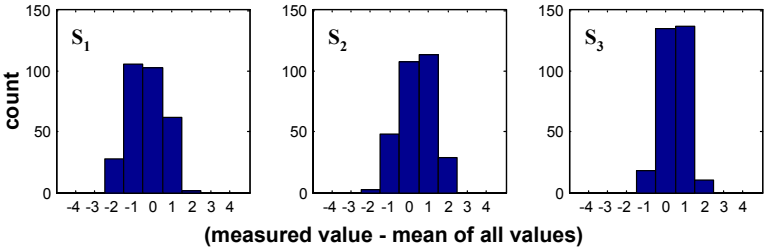

Fig. 1. Distribution of the noise in measured values reported by three light sensors measuring the same source

Consider the three histograms shown in figure 1 which correspond to the raw outputs of three photovoltaic elements connected to an analog to digital conversion circuit. Photovoltaic elements are small electronic devices that produce a voltage at their output pins based on the amount of incident light on their surface. This specific component is readily available in electronic supply stores and is quite inexpensive. It produces roughly 500mV in an average well-lit office space. The histograms correspond to the outputs of three individual sensors of this same type, each measuring the same light source under controlled conditions. The mean of the time-series data has been subtracted so that only the noise component in the measurement remains. The horizontal axis represents raw, uncalibrated values and thus, does not have a standard unit associated with it.

In the light-sensor example above, the errors in the measurements can be classified into 2 major categories as abstracted in figure 2. The vertical axis in the figure represents the probability and the horizontal axis represents the amplitude of a sensor output (i.e. the reported value). Note that figure 2 is only meant as an illustrative diagram and does not necessarily represent the exact characteristics of any specific sensor device. The two major classification of sensor errors are:

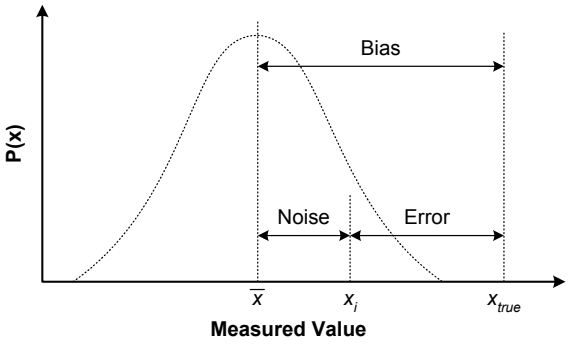

Fig. 2. Sensor measurement error terminology

- **Systematic Errors (*Bias*)**: The bias is an offset in the mean amplitude of sensor readings $\overline{x}$ from the true value x_{true}. The bias may depend on time, the sensed phenomena, the environment, or other factors.
- **Random Errors (*Noise)***: This random component in the error may be due to external events that influence sensor readings, hardware noise, or other difficult-to-predict transient events. In some cases, the noise in measurements may be modeled using a specific distribution (such as Gaussian).

Throughout our discussions, we assume that the output characteristics of sensors are of this general type. For a given measurement (at a given time), the error is the difference from the reported value of the sensor and the true value, which we refer to as ground truth. The goal of calibration in general, and our collaborative calibration in particular, is to determine and correct *systematic* biases in sensor readings. In case of a light sensor, the bias can be due to the sensor or supporting hardware, or external factors such as dust particles on the protective lens of the sensor. In our subsequent discussions, when we refer to sensor "readings" we assume that the measurement noise has been filtered out, for example by using averaging over time, but that the systematic bias remains.

The main challenge in detecting the systematic errors autonomously is the lack of known stimuli against which sensor measurements can be calibrated. In this paper, we present a collaborative calibration scheme that addresses this problem. The scheme exploits the redundancies in sensor measurements under

dense deployment scenarios and dynamically and autonomously derives calibration functions relating the biases of pairs of sensors. This scheme is in essence different than traditional calibration since it calibrates sensor outputs against the outputs of other sensors, utilizing redundancy. To achieve true calibrated results, one must have a reference point to the ground truth. In typical systems, this can be achieved by manually calibrating a subset of the sensors in the system, and allowing the calibration to adjust the remaining sensors based on the calibrated subset. Clearly the final calibration accuracy with respect to the real ground truth will also depend on the number and distribution of such reference points. However, in this paper, we focus our attention on the relative calibration errors between sensors.

1.1 Paper Organization

In the next section, we present the related work followed by the technical preliminaries section including several assumptions and definitions used in subsequent discussions. Section 4 contains the details of our two-phase calibration algorithm. Phase 1 is the main highlight of this paper which proposes an algorithm for deriving calibration relationships between pairs of co-located sensors. In phase 2, the goal is to improve the results of the first phase at a local level (including several nodes) since errors in measurements and inaccuracies in results will often yield inconsistent pair-wise relationships between different sensors. Results using both measured and simulated data are presented and discussed in section 5.2.

2 Related Work

The topic of sensor calibration is as old as sensors themselves. It is impractical to list all of the work that has been done in this area over centuries of human science. Thus, we focus on calibration techniques proposed for sensor networks

In [1] authors address calibration of transmission power in the context of a signal strength based localization system. Even though a radio transceiver is not usually considered a sensor per se, the technique described in that paper can, potentially, be applied to more traditional sensors. The approach described in [1] formulates signal strength calibration as a global optimization problem and as written was not intended for distributed on-line deployment. Sensor fault-tolerance is an issue that is closely related to sensor calibration. In [2] the author suggests a methodology for the design of a fault-tolerant sensor. The author advocates increasing the reliability of a virtual sensor through the use of appropriate models of the phenomenon and replication of physical sensors. Our approach is similar to [2] in that it assumes particular models for sensor failures and phenomena. However, we focus on calibration of a network of physically distributed sensors as opposed to the fault-tolerance of a single sensor.

A significant amount of calibration research work has been done in the context of array signal processing[3]. This research focuses on acoustic and radio

signals. Receiver *equalization* is defined as the calibration of the frequency response of a device. Issues such as time synchronization also come up in practice. Our approach is similar to *blind equalization*; it does not rely on known calibration sources. However, unlike blind equalization, our method does not assume any large-scale propagation models. Our method only requires understanding of the sensed phenomenon at a small scale, because sensors are assumed to be densely deployed.

The field of robotics has also contributed to in-place sensor calibration. Some of these calibration techniques take advantage of intrinsic sensor mobility. In [4], the authors suggest an approach to calibrating the perceived map of the world based on the data received from an inaccurate odometric sensor. Even though their work is drastically different for ours in its applications, the overall philosophy of the approaches is similar. The authors of [4] propose to derive the initial map of the world (a set of calibration functions in our case) based on the current information from a sensor. The inconsistencies of this map are later "relaxed" through a global optimization procedure.

3 Technical Preliminaries

We make the following assumptions about our target sensor systems

- Phenomenon
  - Known and limited spatial frequency (Nyquist)
  - High temporal frequency
- Sensors
  - Dense deployment: This indicates that we have multiple neighboring sensors sensing the same phenomenon and that calibration partitions do not occur.
  - Sensing is slow and has no drift within a calibration epoch with respect to the calibration process.
  - No angle-dependent gains in sensor measurements.
  - Due to time-synchronized nature of our calibration process, we assume there is no hysteresis or delay in sensor response.

Throughout this paper, we use the set S to denote the set of sensor nodes being considered. We denote the measurement reported by sensor $s_i \in S$ at time instance t as $s_i(t)$. We assume all reported measurements are real valued scalars.

Definition: A calibration function (CF), denoted as $F_{i,j}(x)$ is a real-valued function mapping the output x of sensor s_i to sensor s_j. For the sake of simplicity, we often omit the parameter x. Each function $F_{i,j}$ can also have an associated confidence weight $0 \leq w_{i,j} \leq 1$.

Definition: A calibration matrix (CM), denoted as $\mathbf{F}$, is a 2 dimensional $|S| \times |S|$ matrix such that each element $F_{i,j} \in \mathbf{F}$ is the calibration function mapping the output of sensors $s_i \in S$ to $s_j \in S$.

4 Calibration Algorithm

4.1 Overview

Our calibration algorithm consists of two phases. In its first phase, the algorithm derives relative calibration relationships between pairs of co-located sensors, while in the second phase, it maximizes the consistency of the pair-wise functions among groups of sensor nodes. The key idea in the first phase is to use temporal correlation of signals received at neighboring sensors when the signals are highly correlated (i.e. sensors are observing the same phenomenon) to derive the function relating their bias in amplitude. We formulate the second phase as an optimization problem.

4.2 Phase 1: Pair-Wise Calibration Functions

In the first phase of our algorithm we rely on a pairwise approach because of its scaling properties. This phase of the algorithm will perform well for any number and densities of sensors under the assumptions stated in section (sect. 3). Since all the computation here is based on only local data, scalability is unbounded.

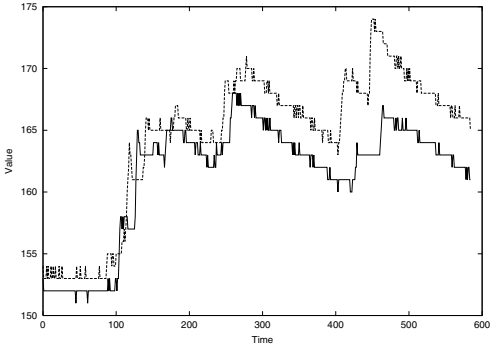

Fig. 3. Time-series data produced by a pair of uncalibrated sensors. Y-axis represents sensor values, X-axis depicts time at which a sample was taken.

Our algorithm consists of the following steps:

1. Collect time-series data in a synchronized manner.
2. Weight each potential data point.
3. Filter out irrelevant data points.
4. Fit a calibration function to the filtered data set.

Time synchronization of sampling is critical because we use temporal correlation to detect periods of time when sensors are observing the same event. A pair of values collected at exactly the same time by two sensors, i and j,

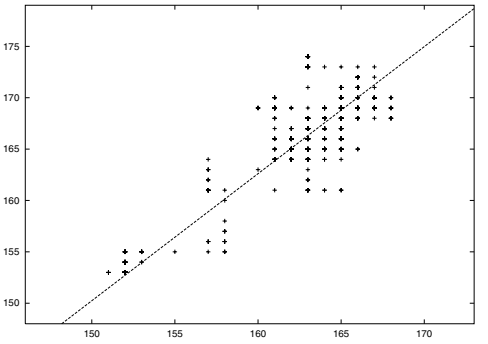

Fig. 4. Scatter plot of time synchronized signals. A data point on this plot is pair of data values taken by the two sensors at exactly the same instant in time. Y-axis corresponds to a value reported by the first sensor; X-axis corresponds to the value reported the second sensor.

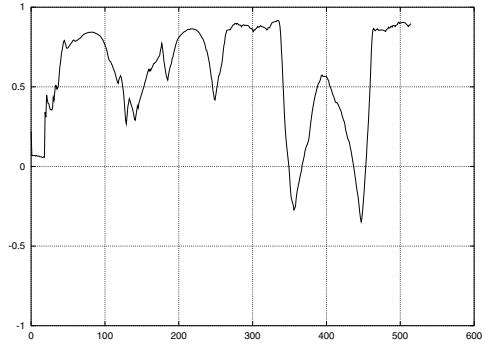

Fig. 5. Sliding window correlation of raw data as a function window time. Y-axis corresponds to the correlation value; X-axis corresponds to the window offset(time).

represents a *potential data point* of a calibration function $F_{i,j}$ between these sensors. It is *potential* because it may correspond to the same external stimulus. Fig. 3 shows two raw data streams aligned in time. If this pair of sensors had been observing the same phenomenon over the course of the whole experiment, we would be able to establish a relationship between them by fitting a line[1] through points on a scatter plot, Fig. 4. Uncorrelated events make it impossible to establish a relation from this data directly. For this reason, our algorithm filters out data points corresponding to periods of time when sensors observed uncorrelated phenomena.

[1] In general, a relationship between sensors can be represented by an arbitrary function. However, a line seems to be a good approximation for the class of sensor that we have used.

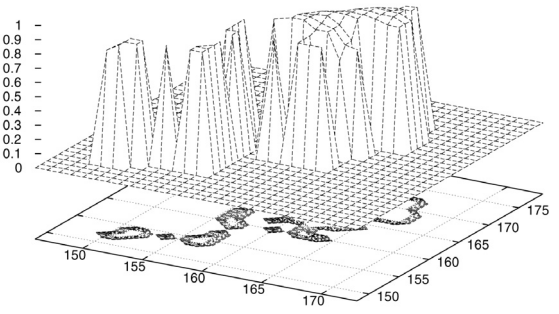

Fig. 6. 3D scatter plot. (x, y) value of each point of a surface corresponds to two sensor values taken at the same time. z-value of each point corresponds to the weight of the point derived through sliding window correlation.

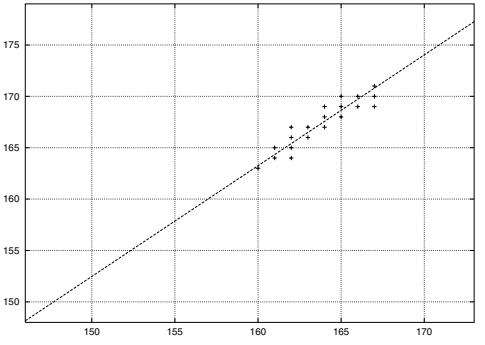

Fig. 7. Filtered scatter plot. It contains only the points believed to be relevant to the calibration relationship.

We use linear correlation to identify periods of time when co-located sensors observe the same phenomena. Our sensor devices exhibited linear behavior in the range of data values collected in the experiments[2]. High sensor density allows us to assume that neighboring sensors are likely to observe the same amplitude of the phenomenon. Fig. 5 shows a sliding window correlation of the same two sensors as a function of window offset. The size of the correlation window depends on the temporal frequency of the phenomenon. If correlated events come and go very quickly, large window size will fail to identify short periods of correlation. Small windows, on the other hand, may render the results of correlation insignificant from the statistical point of view. In our experiments we had the luxury of controlling the rate of change of the phenomenon; therefore, the window size was based on the known duration of correlation periods. A systematic method

[2] The ratio of the variances of the linear fit to a quadratic fit was around 2, whereas the ratios of quadratic to cubic and cubic to the forth power fit were close to 1

for deriving the correlation window size for different sensing applications is an open problem and is a subject of future work.

Based on the result of the sliding window correlation we establish weights of all potential data points. Initially weights of all potential data points are set to zero. Then, for all positions of a sliding window that produce a positive correlation coefficient we evaluate all potential data points included. If a point had positively contributed to the correlation of the current window, we increment its weight by the correlation coefficient of the window. The result of this procedure can be visualized as a 3D version of the scatter plot (see Fig. 6), where the height of each point determines its final weight.

The above heuristic provides means to rank potential data points according to their relevance to the relationship between the sensors. This allows us to filter out irrelevant points by picking potential data points with top ranks[3], *the top set*. The level of confidence in $F_{i,j}$ derived from this set can be related to the distribution of ranks in the top set. Choosing an appropriate size of the top set is not straight forward. Small set size may result in large error in the final relationship, due to filtering out of relevant data points. Large sets, on the other hand, are bound to contain more irrelevant data points, and thus may be noisy. In this study we have picked an arbitrary set size of 20 points. A systematic method for deriving a set size that maximizes "correctness" of the relationship is a subject of the future work.

After we have computed the top set, we proceed to fitting the calibration function, $F_{i,j}$. This procedure is similar to in-factory sensor calibration, but the stimuli are unknown. The nature of the calibration function depends on the type of sensors used[4]. For example, if the linear error in substrate doping in a semiconductor sensor are known to result in second order changes to sensor sensitivity, the calibration function for sensors of this type is very likely to be quadratic. For the purposes of this study we have assumed a linear calibration function. The result of fitting a line to the top set is shown in Fig. 7.[5]

4.3 Phase 2: Localized Consistency Maximization

Due to errors, the pair-wise calibration functions $F_{i,j}$ between pair of different nodes s_i and s_j, derived in the first phase of the algorithm, will not be globally consistent. More specifically, traversing the CFs along different paths will yield different calibrated results for a given node. In order to illustrate this problem,

[3] Direct thresholding may also be used. However, in some cases, the procedure may fail to establish the relationship due to the lack of potential data points above the threshold.

[4] If the difference in coupling of sensor to the environment needs to be accounted for, the relationship function should include corresponding terms.

[5] The pair-wise calibration algorithm described here does not limit the choice of calibration relationship. However, the use of linear correlation assumes that the inter-sensor relationship can be approximated by a linear transformation within the correlation window. In cases where this is not acceptable, it may be possible to use Spearman (rank) correlation instead of linear correlation.

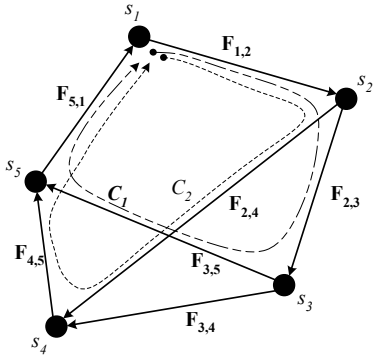

Fig. 8. An example of a calibration graph (CG)

consider the calibration graph (CG) depicted in figure 8. A CG in essence is the graphical representation of a calibration matrix. Each vertex in the CG represents a sensor node and each edge represents the corresponding CF $F_{i,j}$. The figure shows two possible calibration cycles C_1 and C_2 for the node s_1. For the sake of simplicity, we only show the CFs in one direction and omit the reverse mappings in the figure.

For a measured sensor value s_1:

$$\mathbf{C_1} : \mathbf{s_1'} = F_{5,1}(F_{3,5}(F_{2,3}(F_{1,2}(s_1)))) \tag{1}$$

$$\mathbf{C_2} : \mathbf{s_1''} = F_{5,1}(F_{4,5}(F_{2,4}(F_{1,2}(s_1)))) \tag{2}$$

and in general, due to errors, $s_1 \neq s_1' \neq s_1''$.

Our goal here is, given a calibration matrix **F**, to compute a new calibration matrix **F'** such that consistency is maximized. In order to formally discuss this problem, we must first establish our definition of consistency followed by an objective function which quantifies the consistency for **F**.

As mentioned above, traversing different paths in the CG can result in inconsistent calibrated values for a node. Furthermore, the number of such paths in a CG will often grow exponentially with the size of the CG. We should also note here that for linear calibration functions, traversing two consecutive edges of the CG results in a quadratic relationship, while for quadratic calibration functions, traversing two consecutive edges will result in a 4th-degree polynomial. It is easy to see that even for relatively simple calibration functions in low order polynomial form, even short CG path traversals results in very high-degree polynomials quickly. For example, consider the following CF:

$$F_{i,j}(x) = a_{i,j} \cdot x + b_{i,j} \tag{3}$$

Traversing a path from nodes s_1 to s_2 to s_3 we have:

$$F_{2,3}(F_{1,2}(x)) = a_{2,3} \cdot (a_{1,2} \cdot x + b_{1,2}) + b_{2,3} \quad (4)$$

$$= a_{2,3} \cdot a_{1,2} \cdot x + a_{2,3} \cdot b_{1,2} + b_{2,3} \quad (5)$$

which is quadratic in terms of the coefficients $a_{i,j}$ and $b_{i,j}$. However, we make two observations:

1) Since calibration relationships are inherently derived using local information, we should focus more on achieving consistency at a local level, i.e. consider paths in the CG of relatively short lengths.

2) Since each path in the CG is comprised of traversing CFs with different confidence levels, we should expect higher consistency levels from higher confidence traversals.

Definition: The calibration-matrix consistency objective function CMCOF for a calibration matrix $\mathbf{F}$, denoted as $\Gamma(\mathbf{F})$, is a real-valued function such that if $\alpha_1 = \Gamma(F_1)$ and $\alpha_2 = \Gamma(F_2)$, then F_1 is more consistent than F_2 if and only if $\alpha_1 > \alpha_2$.

The exact choice of the appropriate CMCOF depends in large part on the nature and types of errors of the sensors, the environment, and also on how the sensor results will be used. We have relied on the standard L_1, L_2, and L_∞ norms of the discrepancies resulting from different paths in the CG as candidates for experimentation. However, the exact choice of a CMCOF function will depend on the application at hand, as well as analysis of experimental data from sensors under real-life conditions. The data from our experiments which have been of relatively small size with a small number of nodes, have not provided convincing indications for us regarding which function performs better. In general, in a laboratory setting, creating experiments which can truly capture the full effects of the real errors on sensors is in itself a difficult undertaking.

4.4 Consistency Optimization Algorithm

Solving the non-linear programming problem which results by trying to maximize a general CMCOF, directly, under the constraints of given pairwise relation functions is computationally intractable. In the next subsection we present our heuristic-based algorithm that attempts to improve the consistency of the calibration functions on a local scale, subject to the computational and storage limitations of typical sensor nodes. In addition, in order to be practical, our goal has been to create an algorithm which lends itself well to localized and distributed implementation in sensor networks.

The algorithm generates a set of data-point values for each sensor based on the derived calibration functions in phase 1. Each value is obtained by picking a starting value, and calculating the weighted averages of the result produced

by traversing different paths in the CG. The weights correspond to combined confidence levels of each traversal and is obtained by multiplying the confidence values along each segment of a path. Thus, higher confidence traversals have a higher weight in the averaging process. The resulting data-points are used to derive a new pair-wise calibration matrix.

```
Enumerate calibration paths P given the CM F
Step 1: Pick a random starting value x
   For every node s_i ∈ S, s'_i = 0 and count_{s_i} = 0
   For every path p_i ∈ P {
       s_prev = first node in p_i
       CurrentValue = x, α = 1
       While cycle not done {
           s_curr = next node in p_i
           CurrentValue = F_{s_prev,s_curr}(CurrentValue)
           α = α x the confidence of F_{s_prev,s_curr}
           s'_curr = s'_curr + α · CurrentValue
           count_{s_curr} ++
           s_prev = s_curr
       }
   }
   For every node s_i ∈ S, s'_i = s'_i/count_{s_i}

Repeat step 1 n times to get n "data points" for each sensor
Step 2: Compute new CM F' using the data-points
```

Fig. 9. Pseudo-code for localized calibration matrix consistency optimiztion

The algorithm is presented as pseudo-code in figure 9. The initial part of the algorithm enumerates the calibration paths which will be used in the rest of the algorithm. A user specified parameter is used to indicate the maximum length of paths which we consider. We enumerate the paths by exhaustively searching the CG graph using breadth-first-search, starting from each node. We discard paths whose confidence levels are below a user-specified threshold. Note that due to the sparsity of the **F** matrix, and these threshold values, the number of paths that are enumerated can be kept to manageable levels (dictated by available memory, speed of processing, and allotted runtime), given the strict resource constraints of the sensor nodes.

After the data points (n) have been generated by the averaging process, the calibration matrix **F** is recalculated by fitting pairwise relationship functions based on the new data points, similar to the corresponding step in phase 1.

5 Experiments

We ran our algorithms on temperature data collected using a set of uncalibrated sensors. Because an important contributor to sensor error may be the differences in electronics supporting the sensing components, we used a COTS wireless sensor node system as our experimental uncalibrated sensor node. Co-located calibrated sensor components directly wired to a data aquisition system were used to collect ground truth data. The experimental setup and results are described below.

5.1 Experimental Setup

We chose to use real wireless sensor nodes (MICA motes [5]) to collect the raw data from uncalibrated sensors. This decision is motivated by our plans to run distributed versions of our algorithms on these nodes in future deployments. The heart of the mote is an Atmel [6] ATMEGA103L micro-controller. This chip has a builtin analog to digital converter and can be connected to a resistive sensor through a voltage divider. We have used a YSI44006 [7] precision thermistor to perform our measurements. This is a very stable sensor, but it is not factory calibrated. The observed calibration error was as high as 10%. The ATMEGA 103L is also equipped with 4KB of EEPROM, which we used to store data collected during the experiments.

In order to verify the quality of the calibration algorithms, we collected the ground truth measurements using an industrial quality data acquisition system(DAQ) from National Instruments [8]. We used the SCXI-1001 chassis with SCXI-1102 module for analog input. Type J thermocouples were used as temperature sensors.

Since the pair-wise algorithm (sect. 4.2) uses temporal correlation of the sensed data, we had to synchronize all sampling. In order to synchronize motes among themselves, we implemented the Reference Broadcast Synchronization algorithm [9]. We also implemented *time routing* to make sampling requests more robust to packet loss and enable synchronized sampling across multiple broadcast domains. We synchronized mote sampling with the DAQ by using one of the DAQ's input channels to trigger the sampling. This channel was controlled by flipping a GPIO pin on the mote.

The experiment was conducted on a flat surface of a table indoors. Separation between the mote's thermistor and the DAQ's ground truth sensor was less then 5mm for all sensor nodes. 9 sensor nodes were placed on the flat surface in a 3×3 square grid. Each square of this grid had dimensions of 5cm$\times$5cm. The ambient temperature in the lab during the experiment was approximately 26° Celsius. The sampling rate was set to 2 samples/sec.

We used a commercial hand-held hair dryer as a heat source. This heat source was positioned 10-20cm above the surface of the table. During the experiment the nozzle of the hair dryer was directed towards the sensors. In order to create temperature variations we moved the heat source over the sensor grid in a ran-

dom fashion at a velocity no greater than 1cm/sec. We limited the velocity of heat source to avoid undersampling.

5.2 Experimental Results

After collecting the data as described above, we analyzed it in the following way. We derived the relationships between all uncalibrated sensor using the pairwise algorithm described in Sect. 4.2. We also mapped each uncalibrated sensor onto its corresponding ground truth sensor[6]. In order to verify the quality of the calibration relationships we calculated the difference between a value derived though direct translation and a value derived through the ground truth.

The result of the above procedure is shown in Fig. 10. The vertical axis corresponds to percentage of the conversions. The horizontal axis corresponds to the difference between a value derived directly through the calibration relationship and a value derived through the ground truth sensor. In our experiment 0 80% of the translations were off by less than 5°C.

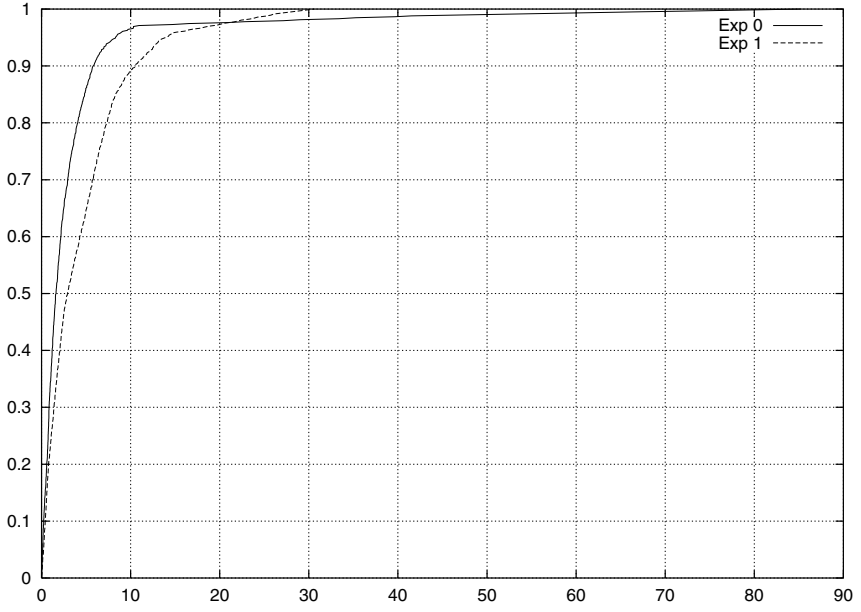

Fig. 10. Calibration error CDFs for two experiments. Y-axis is a fraction(percentage) of inter-sensor translations. X-axis is a error in the translation in degrees C.

[6] For the purposes for this study we neglected the difference among the calibrated sensors since we measured it previously to be less than 1.0°C.

5.3 Discussion and Future Work

While this scheme requires significant development and study before it will become deployable in the field, the results are promising. The distribution of calibration error is such that in 70% of the cases we were able to derive calibration relationships for the sensors with less 5°C. It is important to note that our algorithm does not make *any* assumptions about initial sensor calibration.

However, more than 10% of all translations were greater than 10°C. We have identified several potential sources of error, each of which will be investigated in future work:

- A possible reason for these errors is undersampling of the phenomenon. The velocity of the heat source may have been too high at times resulting in undersampling of the signal and associated aliasing. This, in turn, would have resulted in additional, possibly correlated, noise in the scatter diagram.
- Another source of error may come from our inability to determine the correct ground truth value. In particular, type J thermocouples have much higher mass than precision thermistors; therefore, thermocouples have slower response to changes in temperature. This may have invalidated ground truth data collected during the periods of high variability of the phenomenon.

As mentioned above, a systematic method for deriving the correlation window as well as choosing an appropriate size of the "top" set in the correlation process are subjects of future research. In addition, our future work also aims to address the applicability of these techniques to higher frequency phenomena such as light, acoustic, and seismic, where our assumption about neighboring sensors sensing the same phenomena and no angle-dependent gains may not hold. In such cases, we believe insights about the nature of the phenomena, the environment, and sensor response characteristics will help in building the appropriate calibration models.

Development of evaluation metrics for calibration quality is another important issue. Different applications may have different requirements. For example, an isotherm finding application may not be concerned with the RMS calibration error. For this application one can define a utility function in terms of the overlap between the isotherm based on the ground truth data and a measured isotherm. We are currently developing this and other methods for calibration quality evaluation.

6 Conclusions

We presented a method that can be used to address the difficult problem of sensor calibration in large-scale autonomous sensor networks. The scheme relies on redundancy in senor measurements due to overdeployment, and assumptions about the nature of the phenomena being sensed, to derive functions relating the output discrepancies (biases) of neighboring sensors. Due to inaccuracies and processing based on purely local information, the pairwise relationship functions

will be inconsistent in the network. In the second phase, new pairwise relationships are derived by a heuristic method that is designed to increase the consistency in the system. Early experimental results indicate that the pairwise relative calibration scheme is promising. However, significant experimentation with relatively larger scale sensor networks are required to determine the true performance, especially for the second phase of the algorithm.

References

1. Whitehouse, K., Culler, D.: Calibration as parameter estimation in sensor networks. In: 2002 ACM International Workshop on Wireless Sensor Networks and Applications (WSNA'02), Atlanta, Georgia (2002) 67
2. Marzullo, K.: Tolerating failures of continuous-valued sensors. ACM Transactions on Computer Systems (TOCS) **8** (1990) 284–304
3. Cevher, V., McClellan, J.: Sensor array calibration via tracking with the extended kalman filter. In: 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Volume 5. (2001) 2817–2820
4. Andrew Howard, M.M., Sukhatme, G.: Relaxation on a mesh: a formalism for generalized localization. In: IROS 2001. (2001)
5. URL: Crossbow corporation, http://www.xbow.com (2002)
6. URL: Atmel corporation, http://www.atmel.com (2002)
7. URL: Ysi corporation, http://www.ysi.com (2002)
8. URL: National instruments, http://www.ni.com (2002)
9. Jeremy Elson, L.G., Estrin, D.: Fine-grained network time synchronization using reference broadcasts. In: Proc. OSDI 2002 (to appear), Boston, MA (2002)