# Sensor-based Clustering for Indoor Applications

SECON 2008

林熙閔

09/18, 2008

# Outline

- Introduction
- Related work
- Analysis
- Distributed approach
- Evaluation
- Conclusions

# Introduction

- The lifetime requirements of wireless sensor network deployments continue to exceed the capacity of today's battery technology by orders of magnitude.
- Node clustering
  - Temporary cluster-node deactivation
  - Cluster-head rotation
    - LEACH (Low Energy Adaptive Clustering Hierarchy)

# Introduction

- Forming clusters in such a way that they reflect **real world semantics** that are **meaningful** to the application.
- This paper shows that it is feasible to automatically create **clusters** (groups of nodes) that adhere to **room boundaries** using inexpensive and broadly available sensors.

# Related work

- In most cases, clusters are formed on the basis of connectivity information or on the basis of geographical positions of sensor nodes.
- Rely on **acceleration sensors**. Their results show that a successful grouping can be established with high accuracy if the sensors are worn on the same part of the body. [6]

# Analysis

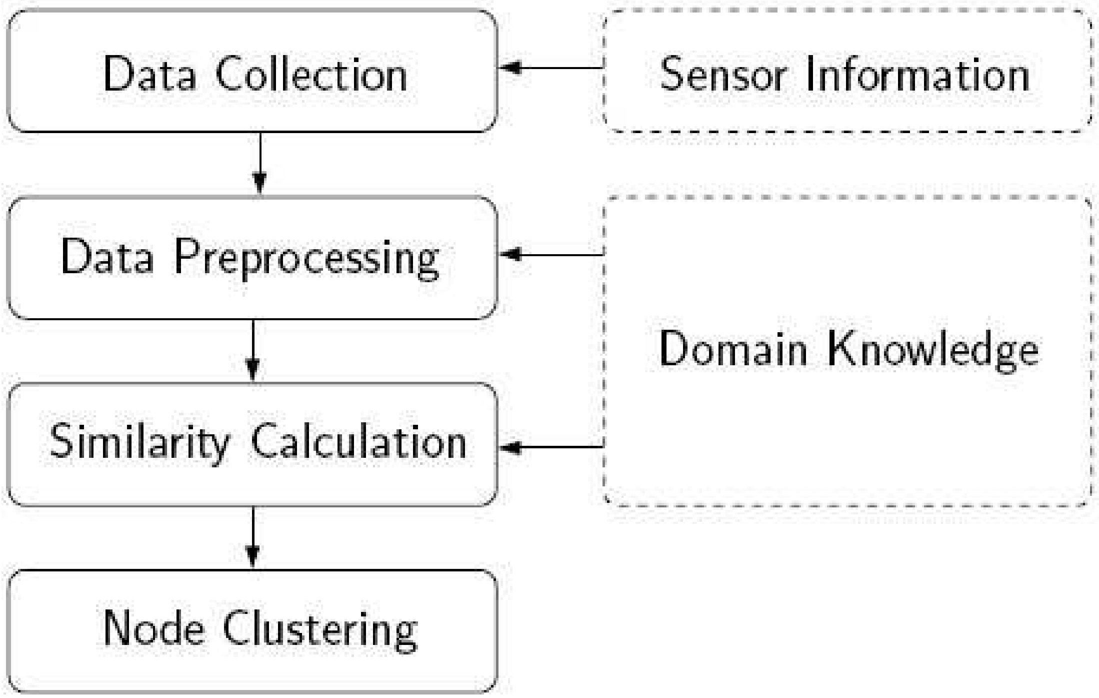



Fig. 1: Processing steps

# Analysis

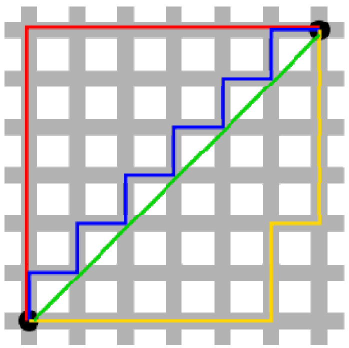
- Off-line analysis using a PC.
  - Not suitable for real world wireless sensor networks.
  - However, it eliminates a number of complications and ensures that multiple methods can be run on the same set of input data.

# Analysis

- **Data Preprocessing** ( or data filtering modifies )
  - Normalization filter
    - bringing the sensor data from all data sources to **one common scale**.
  - Data smoothing filter
    - calculating the **average of the previous x values** and uses this average as the output.
  - Curve tendency filter
    - solely records whether the **current sensor value is higher or lower** than the previous value, thereby producing **binary output**
    - required a **large number of samples** or **tended not to stabilize** which disqualifies it for use in real applications
  - Event detection filter
    - record **significant changes**
      - a light that is switched on or off
    - proved to be useful in a large set of cases across all experiments

# Analysis

- Similarity Calculation



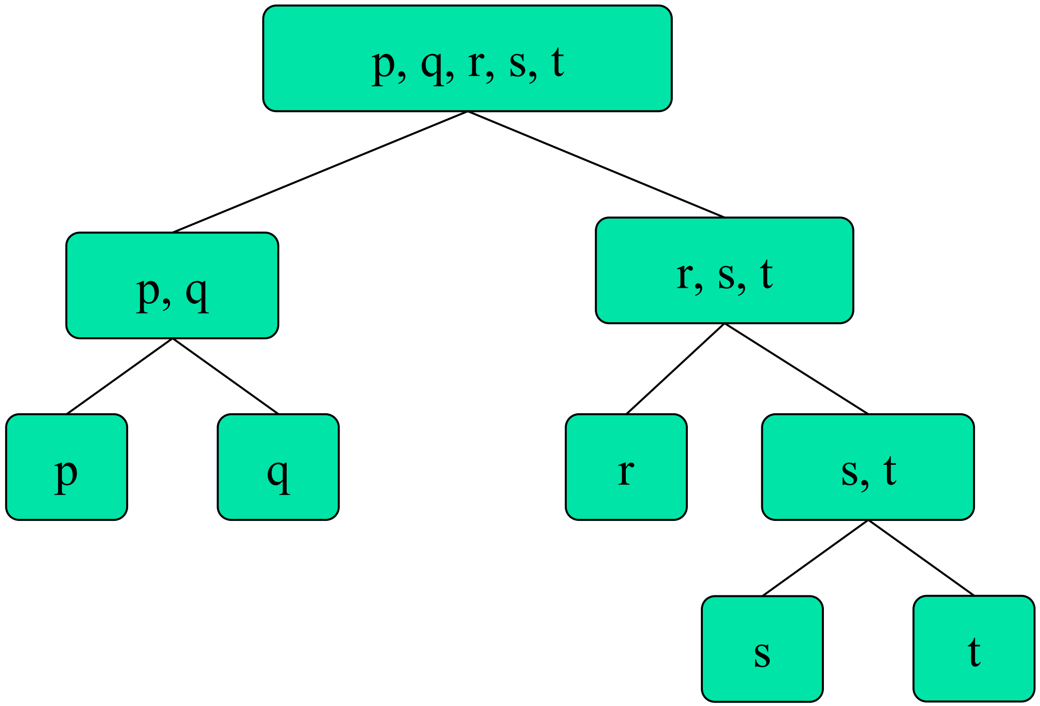
  - Distance metrics express how far apart two variables are according to a certain criterion
    - Euclidean distance
    - Manhattan distance
  - Correlation coefficient measures the strength of a relationship between two variables
    - Pearson coefficient
      - **most reliable** similarity metric
    - Phi coefficient
      - works well in combination with the event filter

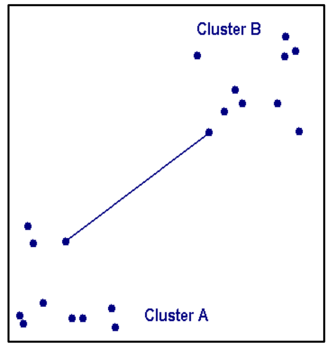
# Analysis

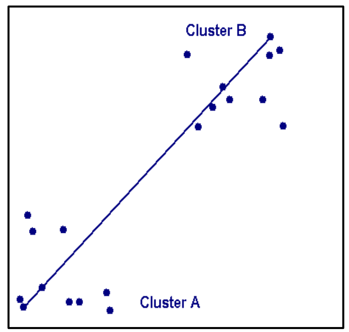
- Node Clustering
  - Hierarchical
    - arrange the individual elements of a set in a tree
    - can either be done agglomerative or divisive
  - Partitional
    - requires knowledge about the **number of clusters** to be found in the system
    - used the k-means clustering algorithm
      - most popular

# Analysis

- Hierarchical

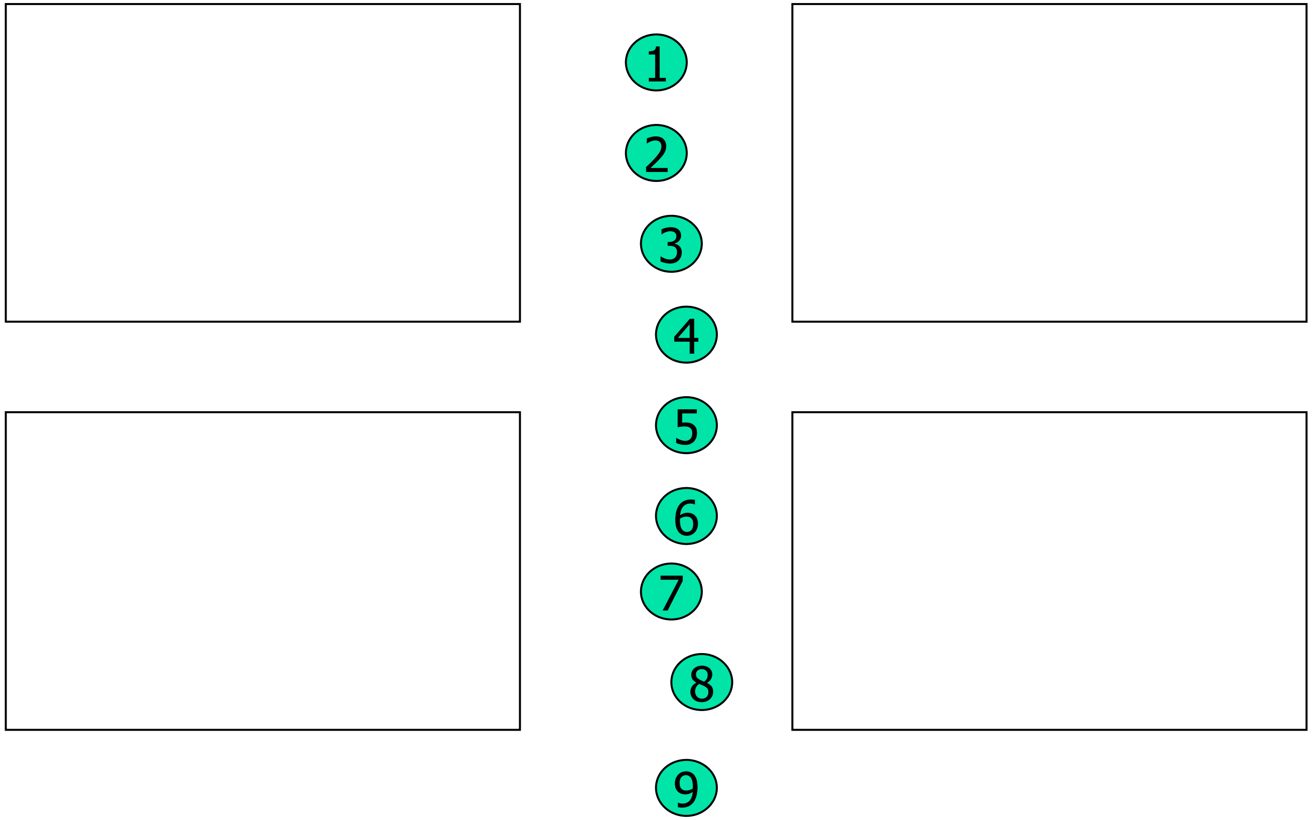


# Analysis

- k-means clustering algorithm



# Analysis

- Experimental Analysis
  - Tmote Sky sensor nodes
    - Temperature、Humidity、Light sensors : photosynthetically active radiation (**PAR**), total solar radiation (**TSR**).
  - Four different rooms with three nodes being placed in each room.
  - Nodes were in **different parts** of the monitored rooms at **different heights** and with **different orientations**.
  - Sensor chips were **not directly covered** by other artifacts of the room.
    - Deliberately **not avoided**
      - lying in the **shadow of an artifact**
      - lying in the **airflow of a window**

# Analysis

| | Home scenario 1 | | Home scenario 2 | |
|---|---|---|---|---|
| | Inside | Between rooms | Inside | Between rooms |
| Humidity | 0.87 | 0.22 | 0.86 | -0.02 |
| Light PAR | 0.96 | 0.50 | 0.93 | 0.20 |
| Light TSR | 0.98 | 0.39 | 0.95 | 0.34 |
| Temperature | 0.83 | 0.35 | 0.77 | 0.18 |

(a) inside rooms / between rooms

# Analysis

| | $T(n)$ | $S(n)$ | **Usefulness** |
|---|---|---|---|
| Normalization | $O(n)$ | $O(n)$ | Limited |
| Data smoothing | $O(n)$ | $O(1)$ | Limited |
| Curve tendency | $O(n)$ | $O(1)$ | Limited |
| Event detection | $O(n)$ | $O(1)$ | High |
| Euclidean | $O(n)$ | $O(1)$ | High |
| Manhattan | $O(n)$ | $O(1)$ | High |
| Pearson | $O(n)$ | $O(1)$ | High |
| Phi | $O(n)$ | $O(1)$ | High |
| Hier. single link. | $O(m^2 \log m)$ | $O(m^2)$ | High |
| Hier. compl. link. | $O(m^2 \log m)$ | $O(m^2)$ | High |
| Hier. avg. link. | $O(m^2 \log m)$ | $O(m^2)$ | High |
| k-means | $O(klmn)$ | $O(kn)$ | High |

(b) Result overview

# Analysis

- Combining Clustering Trees
  - **Improve** the resulting clustering **quality**
  - **Balance weaknesses** and strengths of different criteria.
  - Combining distance or correlation information from different sources **cannot be done by simply calculating** the average distance or correlation matrices
- **Average consensus supertree (ACS)** [14]

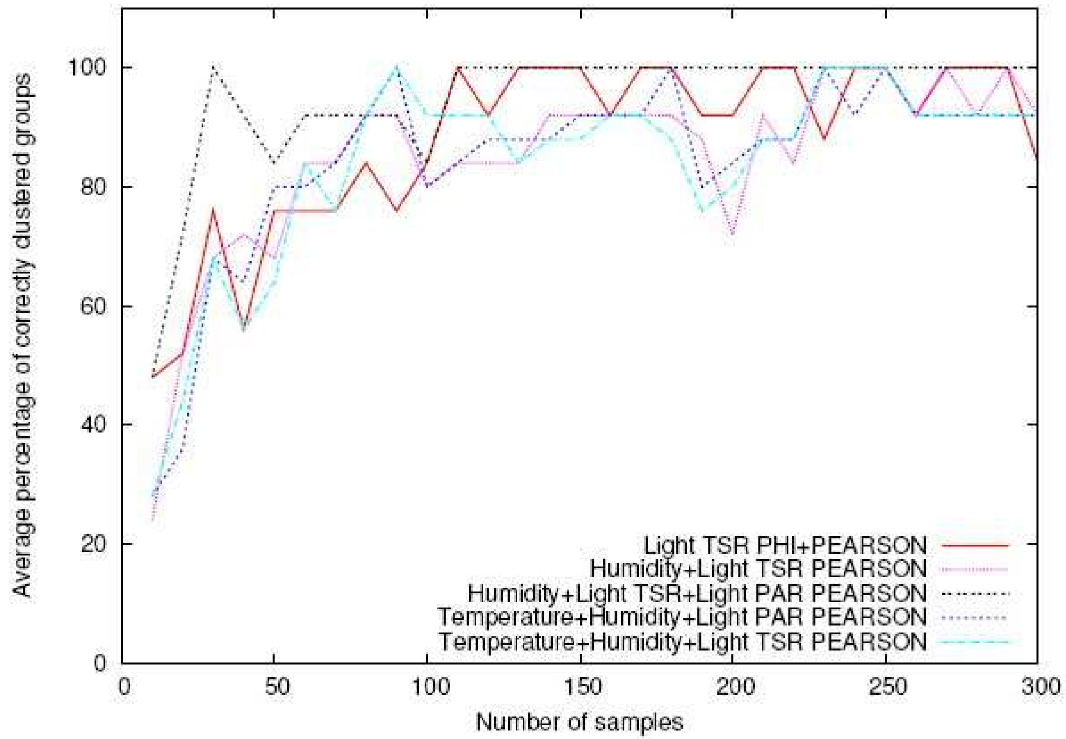
# Distributed approach

- Forwarding the collected sensor data from all nodes in the network **with a sufficiently high sampling rate** creates a **high traffic load**.

# Distributed approach

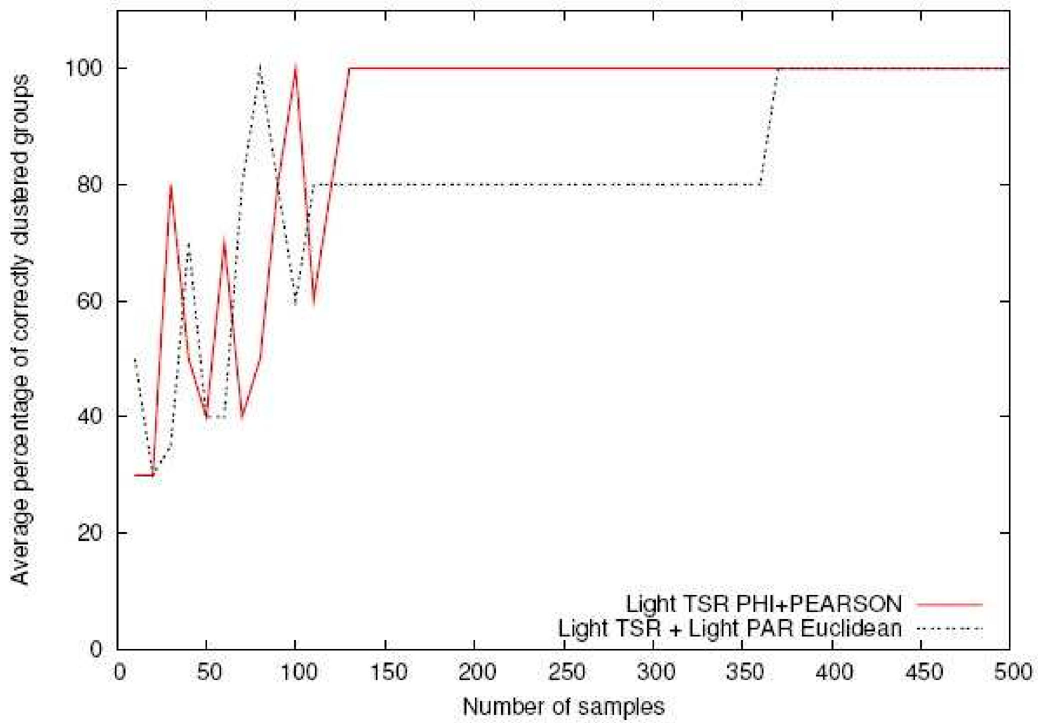
- Data preprocessing and similarity calculation are directly done on the individual sensor nodes.
  - each sensor node **periodically** reads its own sensor data and **broadcasts** the collected data in a **beacon message** to its **neighbors** every p seconds.
  - when **receives** a beacon message: Reads its own sensor data and compares the two data samples to **extract information** required for the similarity calculation.
- The actual computation of the node clustering.
  - the last step of the clustering process
  - cannot be distributed as it requires a **global view** on similarity information among all nodes.
- Collecting the **similarity information** at a central point is much **less expensive** than collecting **complete vectors** of sensor data.

# Evaluation



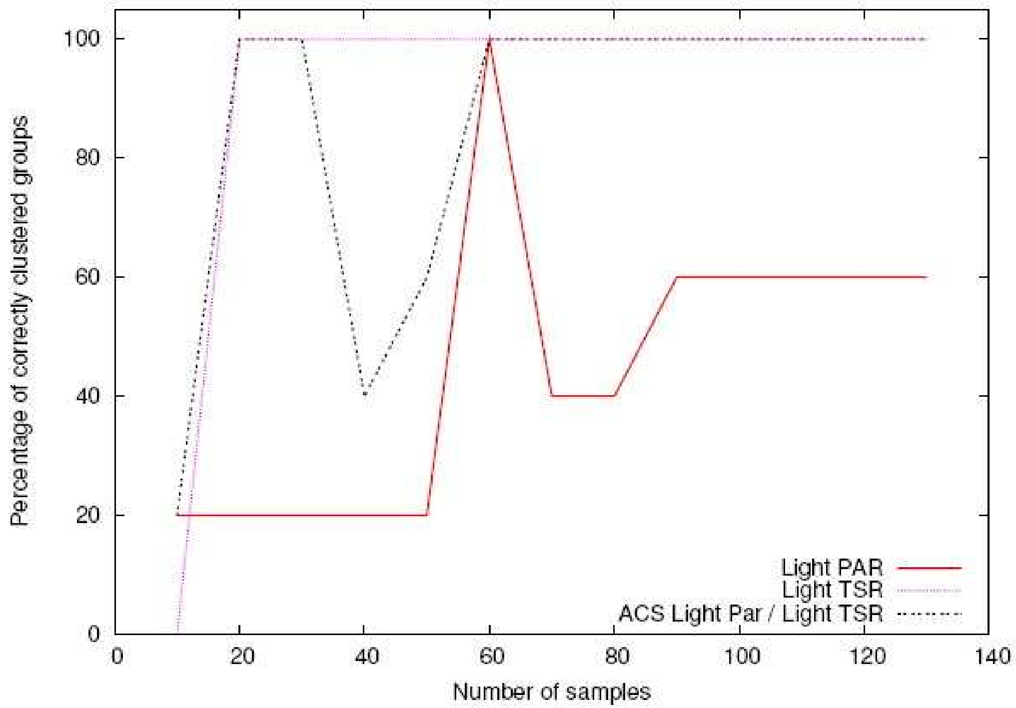
Average percentage of correctly clustered groups
0
20
40
60
80
100
0
50
100
150
200
250
300
Number of samples
Light TSR PHI+PEARSON
Humidity+Light TSR PEARSON
Humidity+Light TSR+Light PAR PEARSON
Temperature+Humidity+Light PAR PEARSON
Temperature+Humidity+Light TSR PEARSON

# Evaluation




(a) without activity

# Evaluation




(b) supported by explicit triggers

# Conclusions

- It is feasible to automatically create clusters that **reflect rooms** by analyzing the measurements of inexpensive and broadly available sensors.
- The idea of **clustering devices based on sensor data** does **not have to be limited** to sensor nodes
  - self-configuring home entertainment systems
  - home automation systems
  - alarm systems