Data Handling
Data Compression Techniques for Wireless Sensor Networks

David Butler
Digital Manufacturing Group, NPL
Design Manufacturing & Engineering Management, University of Strathclyde

Introduction

- Recent advances in the field of information technology has resulted in the generation of huge amounts of data every second.

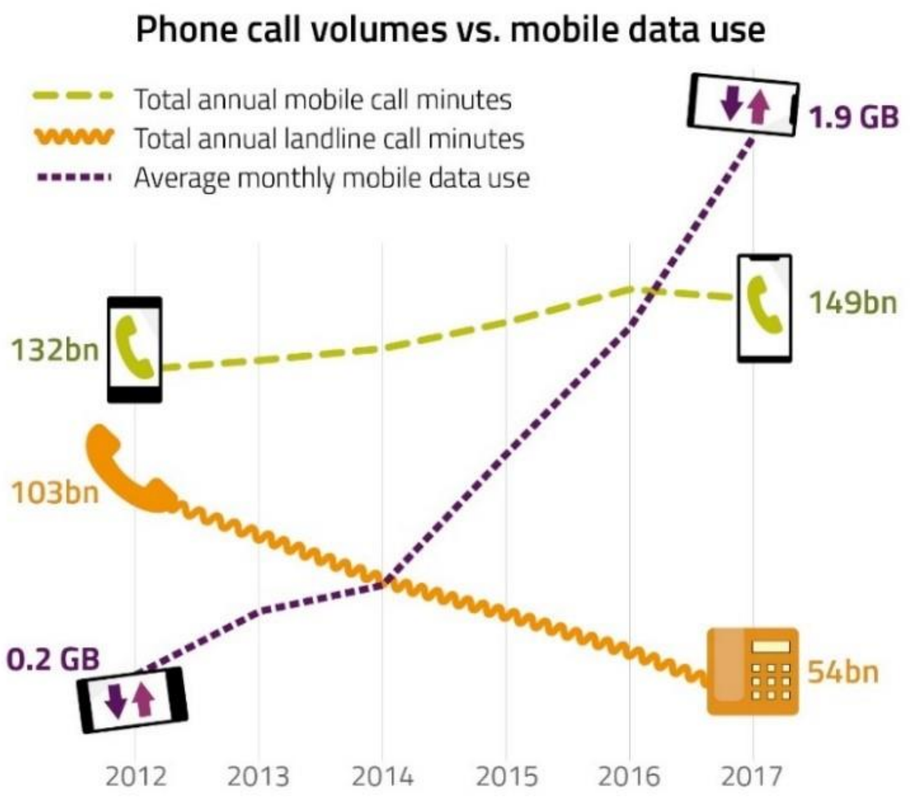
The necessity of storage and transmission increases at least twice as storage and transmission capacities increase.

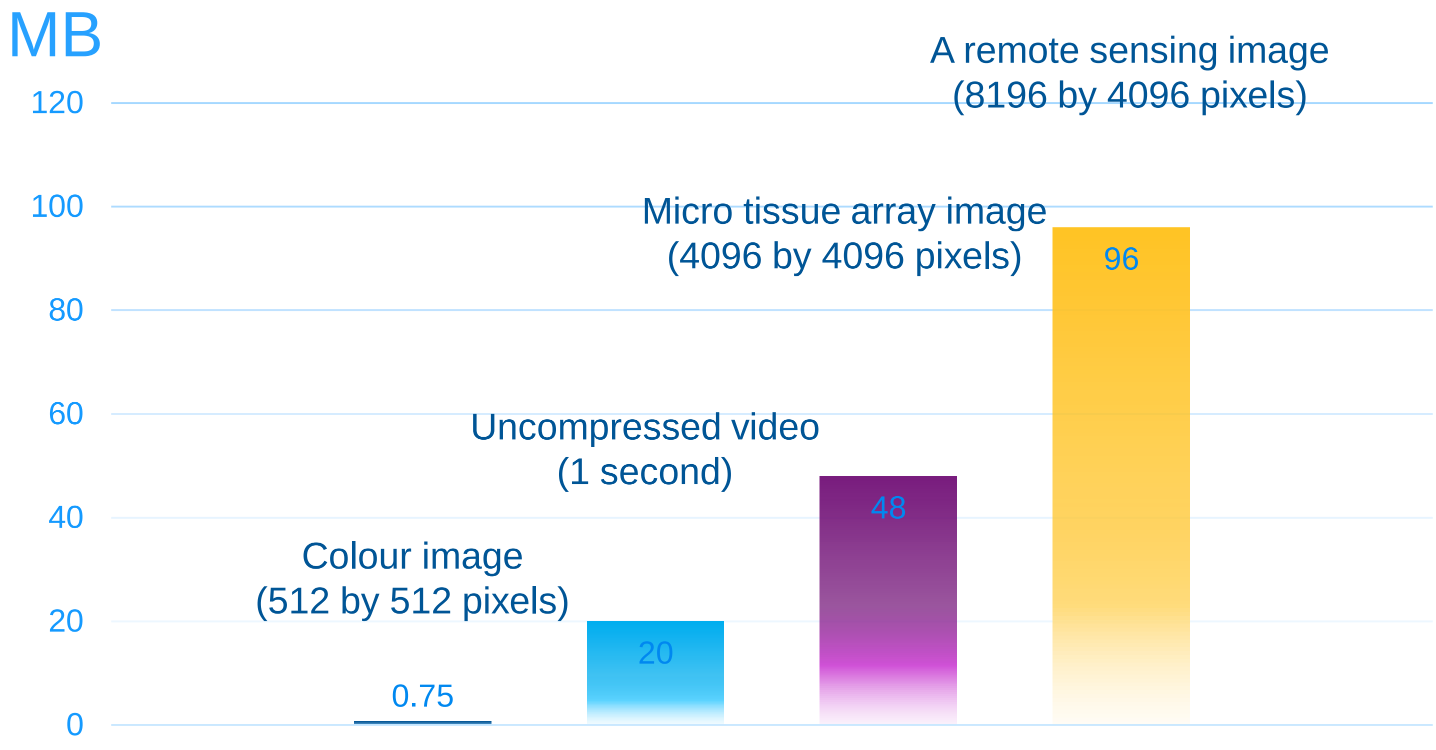
Based on Parkinson's First Law[1]

[1] Parkinson, C.N., 1957. Park Parkinson's First Law: "Work expands so as to fill the time available." In: Parkinson's Law and Other Studies in Administration. Ballantine Books, New York.

The Growth in Data Usage

NPL National Physical Laboratory





Data Transmission Costs

- A dedicated transmission of remote sensing data at uplink and downlink speed costs up to 1900USD

Data Compression

- In general, data can be compressed by eliminating data redundancy and irrelevancy
- Two levels in compressing data
  1. Data is analysed for any **redundant information** and extract it to develop a model
  2. The difference between the model and actual data called **residual** is computed and coded by the encoding technique

Data Compression Metrics

National Physical Laboratory

- Performance of Data Compression algorithms can be measured by
  - Algorithm complexity
  - Computational memory
  - Speed
  - Amount of compression
  - Quality of reconstructed data

Data Compression Metrics

$$CR = \frac{\text{No. of bits in uncompressed data}}{\text{No. of bits in compressed data}}$$

CR can be termed as bit per bit

$$Space\ saving = 1 - \frac{\text{No. of bits in compressed data}}{\text{No. of bits in uncompressed data}}$$

Defined as the reduction in file size relative to the uncompressed size

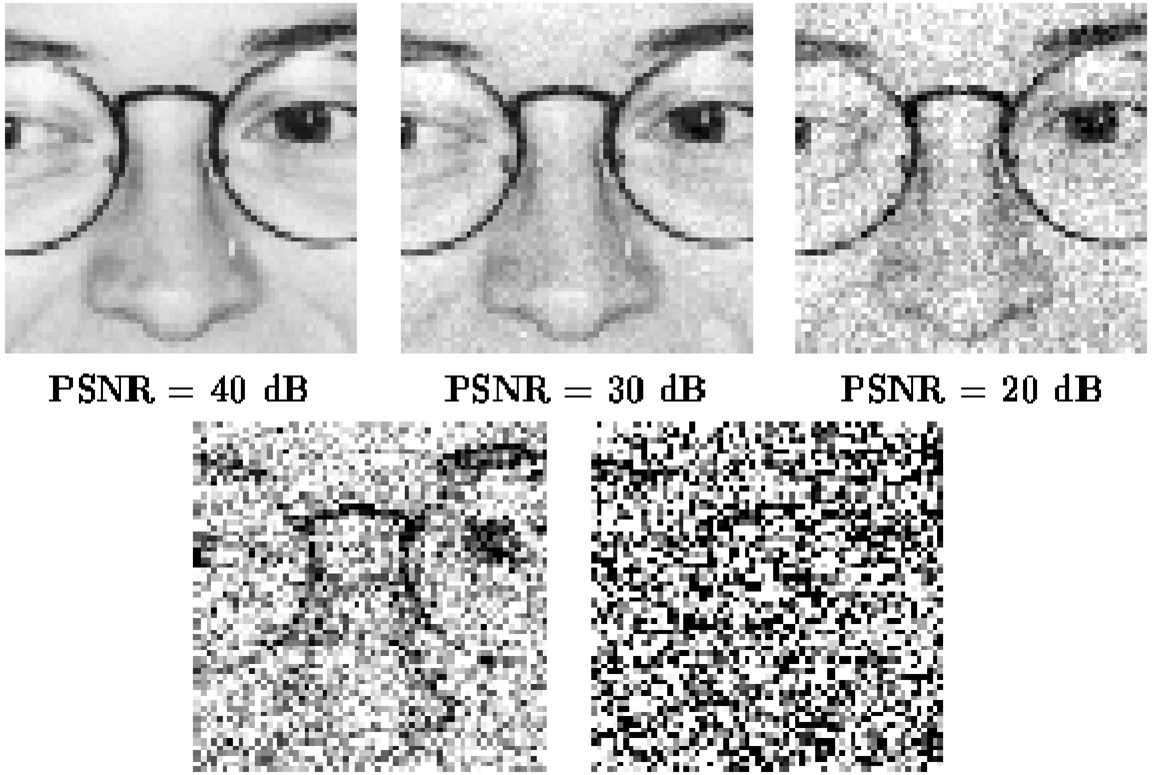
$$Compression\ gain = 100 \log_e \frac{original\ data}{compressed\ data}$$

Data Compression Metrics

- In lossy compression, reconstructed data actually varies from the original data, performance needs to consider the level of distortion, fidelity and quality.

$$PSNR = 20\log_{10}\frac{max|X_i|}{RMSE}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2$$



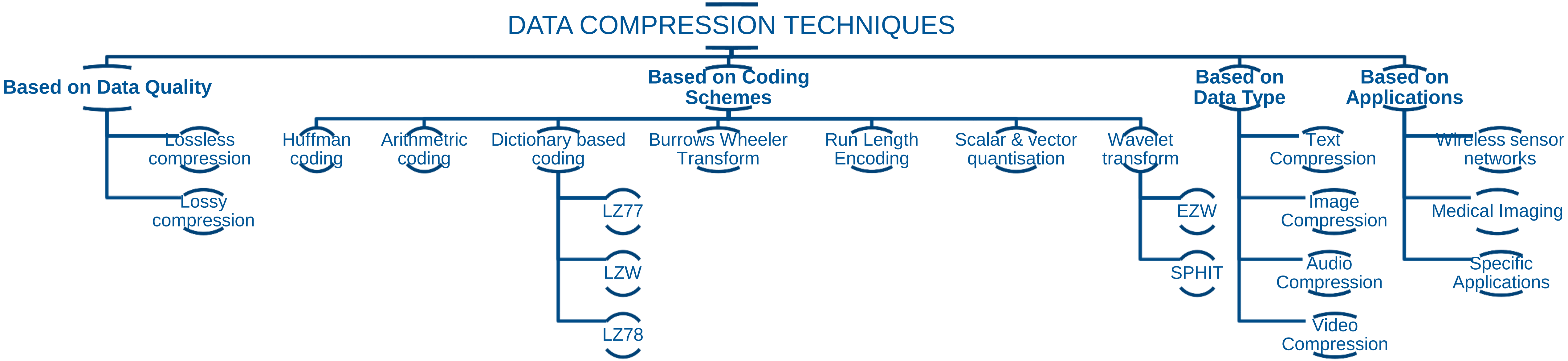
PSNR = 40 dB   PSNR = 30 dB   PSNR = 20 dB

PSNR = 10 dB   PSNR = 0 dB

Saeed Ahmed Sohag, Dr. Md. Kabirul Islam, Md. Baharul Islam, A Novel Approach for Image Steganography Using Dynamic Substitution and Secret Key, American Journal of Engineering Research (2013)

When original and reconstructed data are exactly identical, RMSE = 0, PSNR will be infinity

Data Compression Techniques



Data Quality

National Physical Laboratory

- The significance of the data quality of a DC technique is highly dependent on the type of data and application.
- DC techniques can be divided into:
- Lossless
  - No information is lost during compression
  - Medical imaging, law, forensics, military imagery etc.
- Lossy
  - Can be preferable where reconstructed data is not perfectly matched with the original data and an approximation is acceptable
- Near-lossless
  - Maximum absolute distortion

Data Coding Techniques

- Can be classified as predictive and transform coding techniques

- ***Predictive Coding***
  - Present information is exploited to predict the upcoming data
  - Actual difference is encoded
  - Simple, easy to implement, can be adjustable to diverse local image features
- ***Transform coding***
  - Converts and input data from one kind of representation to another kind
  - Transformed values (coefficients) are encoded by compression techniques
  - More suited to high computational complexity

Data Coding Techniques

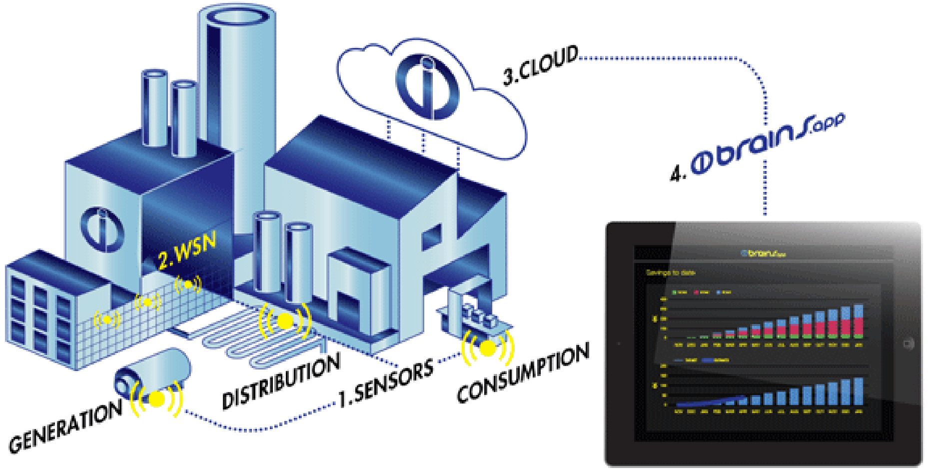
National Physical Laboratory

Reference	Coding	Feature	Compression Type	Versions	Advantages	Applications
Huffman (1952)	Huffman coding	Entropy based	Lossless	Minimum variance Huffman code, Length Limited Huffman code, Adaptive non-binary, Golomb-rice coding, Tunstall code	Effective in all file formats	ZIP, ARG, JPEG, MPEG, PKZIP
Langdon (1984)	Arithmetic coding	Entropy based	Lossy and Lossless	Adaptive arithmetic coding Binary arithmetic coding	Flexibility	JPEG, multimedia applications
Ziv and Lempel (1977)	LZ coding	Dictionary based coding	Lossless	LZ77, LZ78, LZW	Compress all kinds of data	TIFF, GIF, PDF, Gzip, ZIP, V.42, Deflate and PNG
Saupe and Hamzaoui (1994)	Fractal compression	Block based coding	Lossy	–	Suitable for textures and natural images	Live video broadcasting
Burrows and Wheeler (1994)	BWT	Block sorting compression	Lossless	–	No need to store additional data for compression	Bzip2
Capon (1959)	RLE	Employs in high redundant data	Lossless	–	Faster	TIFF, BMP, PDF and fax
Sayood (2006)	Scalar and Vector Quantization	Represents larger set of values to a smaller set	Lossless and Lossy	–	Less complexity	–

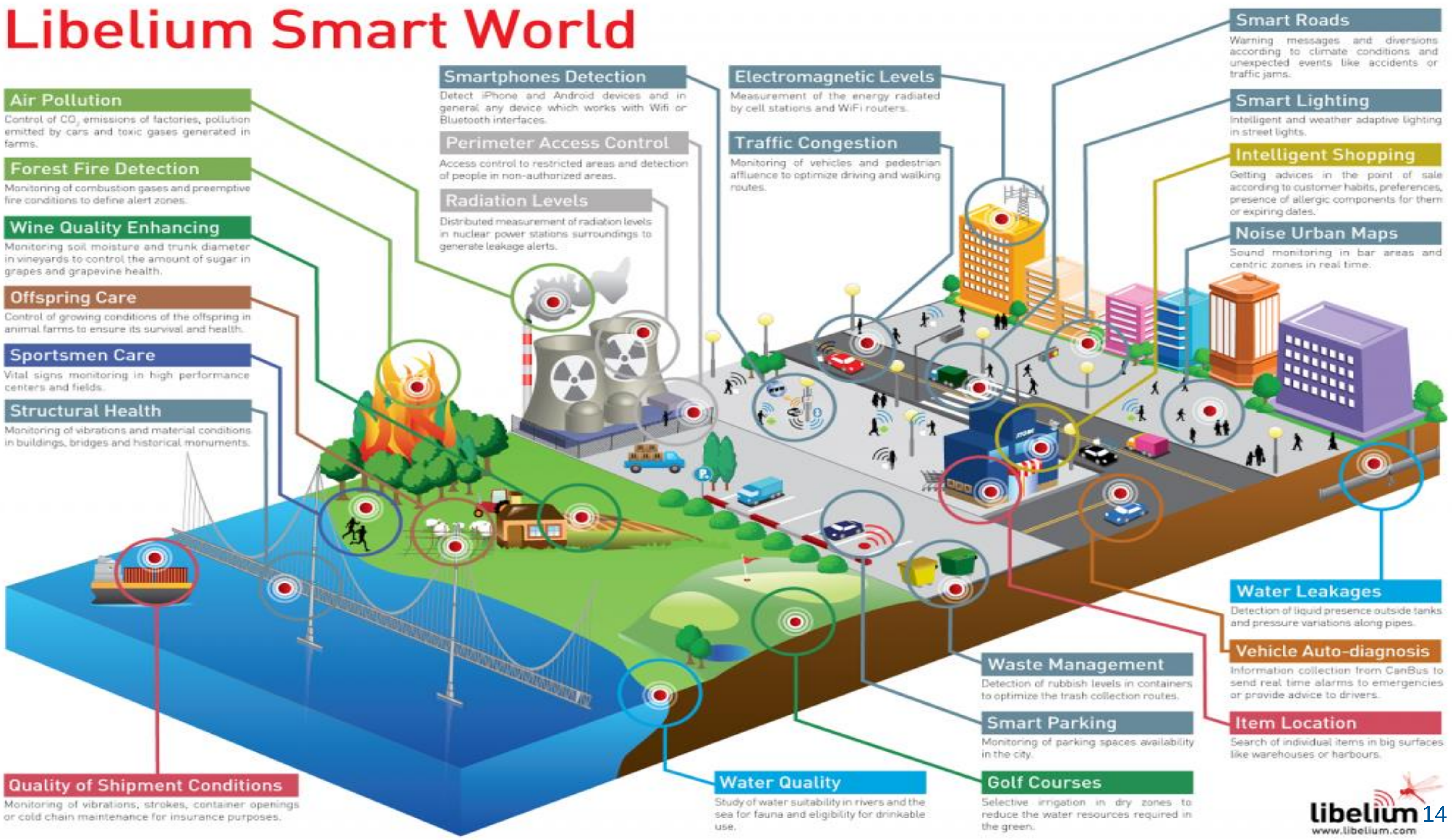
Uthayakumar, J., et al. A survey on data compression techniques: From the perspective of data quality, coding schemes, data type and applications. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.05.006

Wireless Sensor Networks

- Sensor nodes are energy constrained in Wireless Sensor networks thus data compression techniques play a major role in minimizing the energy consumption.
- By minimizing the amount of data reduces the number of data transmission



https://www.electronicdesign.com/iot/wireless-sensor-networking-industrial-iot



Source: http://www.libel m.com/resources/top_50_iot_sensor_applications_ranking

Recent Research

- **Marcelloni & Vecchio (2008)** proposed a simple data compression scheme to **reduce storage space** and **reduce computational resources** in WSN. The compression algorithm utilized high correlation between successive samples measured by the sensor node.
- Compression performance of 66.99% & 67.33% for temperature and relative humidity.

Marcelloni, F., Vecchio, M., 2008. A simple algorithm for data compression in wireless sensor networks. IEEE Commun. Lett. 12, 411–413. https://doi.org/10.1109/LCOMM.2008.080300.

Recent Research

- **Kolo *et al* (2012)** proposed an Adaptive Lossless Data Compression scheme (ALDC) . ALDC uses **predictive coding** for better capturing of underlying temporal correlations present in the sampled data for surveillance applications.
- The proposed system is applicable for real-time as well as delay insensitive situations.
- It achieved 74.02% better compression for real-world datasets when compared with other methods.

Kolo, J.G., Shanmugam, S.A., Lim, D.W.G., Ang, L.-M., Seng, K.P., 2012. An adaptive lossless data compression scheme for wireless sensor networks. J. Sens. https://doi.org/10.1155/2012/539638

Recent Research

- **Ruxananyasmin and Krishna (2013)** presented on LZW coding which is used to decrease energy consumption and maximise the network lifetime. LZW compresses the file to 1/3rd of its original size.
- LZW achieves a maximum CR for different file formats (e.g. text, speech and images)
- Widely used in mobile adhoc networks

Ruxanayasmin, B., Krishna, B.A., 2013. Implementation of data compression techniques in mobile Ad hoc. Networks 80, 8–12.

Recent Research

- **Incebacak et al (2015)** studied several DC techniques to enhance the lifetime of WSN working in stealth mode.
- A mathematical framework was developed to analyse the advantages of DC approaches to maximise the network lifetime

Incebacak, D., Zilan, R., Tavli, B., Barcelo-Ordinas, J.M., Garcia-Vidal, J., 2015. Optimal data compression for lifetime maximization in wireless sensor networks operating in stealth mode. Ad Hoc Networks 24, 134–147.

Recent Research

National Physical Laboratory

- **Abu Alsheikh *et al* (2016)** proposed a low-cost lossy compression approach with an error bound guarantee.
- The proposed method **reduces data congestion** and **minimizes energy** dissipation achieved by the utilization of spatio-temporal correlation between data samples.
- *Based on neural networks, a machine learning algorithm automatically predicts human activities and environmental conditions*.
- The algorithm has been tested with meteorological datasets and produced better results than other methods.

Abu Alsheikh, M., Lin, S., Niyato, D., Tan, H.P., 2016. Rate-distortion balanced data compression for wireless sensor networks. IEEE Sens. J. 16, 5072–5083. https://doi.org/10.1109/JSEN.2016.2550599.

Current Research

- **Wu et al (2016)** developed a novel energy efficient framework for clustered WSN which integrated **data prediction, compression and recovery techniques**.
- The main intention was to minimize the amount of **data transmission** using data prediction and compression with guaranteed delay.
- The proposed method was tested with a public sensor dataset
- Results showed that the method was efficient for continuous monitoring of WSN.
- Performance metrics
  - Prediction algorithm – prediction accuracy, convergence speed communication reduction
  - Compression algorithm – communication cost

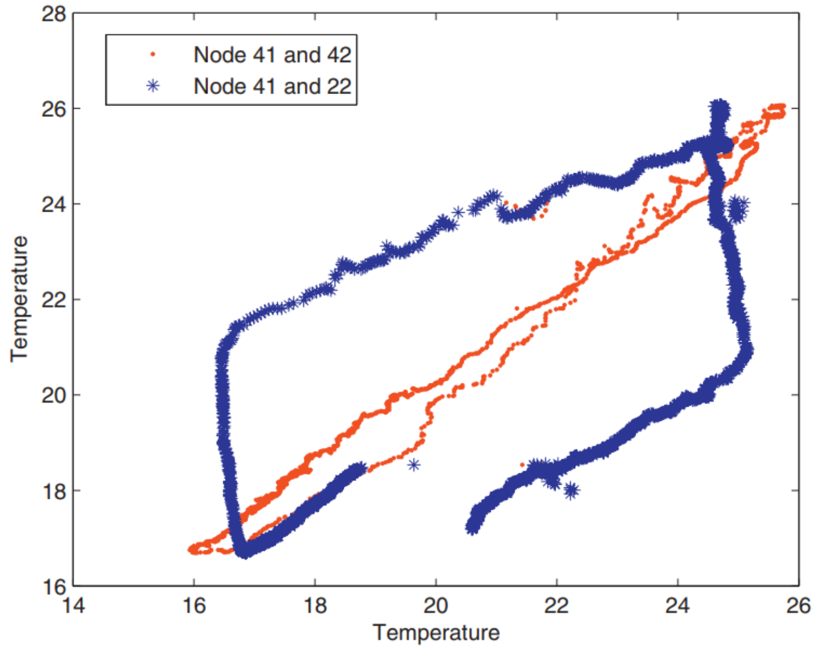


Fig. 3. Temperature readings collected by two pairs of sensor nodes over 3000 samples.

Wu, M., Tan, L., Xiong, N., 2016. Data prediction, compression, and recovery in clustered wireless sensor networks for environmental monitoring applications. Inf. Sci. (Ny) 329, 800–818. https://doi.org/10.1016/j.ins.2015.10.004.

Future Directions

National Physical Laboratory

Discrete Tone Images (DTI)

- Most image compression techniques are developed only for continuous tone images and very few are proposed for discrete tone images
- Discrete tone images are gaining in importance

  Machine generated images e.g. digital maps, GIS, logos tec.

Future Directions

Machine Learning in DC

- The application of compression algorithms in machine learning (ML) tasks such as clustering and classification has appeared in a variety of fields.
- The nature of ML can be used to predict the original input sequence from the past sequence – a useful tool for the decompression process.

Future Directions

- Framework for data strategies?

Inputs

- Level of information needed
- Frequency of information
- Cost of capturing data
- Value of captured data
- Cost of storage

Outputs

- Data sampling rate
- Data compression (transmission)
- Data compression (storage)
- Quality after data storage?

Forbes CommunityVoice   Connecting expert communities to the Forbes audience.   What is This?

837 views | Aug 21, 2018, 07:15am

How To Effectively Manage Data With Your Smart Manufacturing Strategy To Drive Profitability

John Clemons Forbes Council
Forbes Technology Council CommunityVoice ⓘ

POST WRITTEN BY

John Clemons

Director of Manufacturing IT for **Maverick Technologies**. Executive Committee of MESA International. Research services by Patricia Panchak.

NPL
National Physical Laboratory

Department for
Business, Energy
& Industrial Strategy

FUNDED BY BEIS

This work was supported by the UK government's Department for Business, Energy and Industrial Strategy (BEIS)