Information Fusion towards Multi-sensor Data using Cognitive Computing Approach

1 Parikshit N Mahalle, 2 Saudagar S Barde, 3 Poonam N Railkar, 4 Pankaj R Chandre

1,2,3,4 Department of Computer Engineering, Savitribai Phule Pune University,
Smt. KashibaiNavale College of Engineering,
Pune, Maharashtra, India

Abstract - As Internet of things and allied applications are spreading their facets at faster rate, use of sensors is increasing rapidly. Every sensor produce big amount of data which needs processing, reliable delivery. There are many use cases where information is to be sensed from multiple sensory resources and fused together. An objective of information fusion from multiple sources is to interpret single parameter for effective decision making. In the sequel, this paper presents an overview of information fusion and its mathematical model. Next part of the paper presents motivation and challenges in information fusion for different uses of Internet of Things. Comprehensive literature survey and evaluation of the literature survey is also presented in order to compare existing potential works with respect to performance parameters like reliability, scalability, computational time etc. The proposed data science approach using cognitive computing is also presented and discussed in detail. Experimental results of information fusion for one sensor and three sensors are also presented which produces significant contribution towards information fusion. Results show that there is marginal difference in the access time to sense information from one sensor and multiple sensors. Finally, paper also presents challenges and future outlook.

Keywords - Data fusion, Information fusion, Multi-sensor.

1. Introduction

The integration of data and knowledge from several sources is known as data fusion [1]. Data fusion methods are used extensively for target tracking, automated identification of targets, remote sensing, battlefield surveillance, condition monitoring of weapons, soldiers and many more [2]. Techniques to combine or fuse data are drawn from a diverse set of more traditional disciplines, including digital signal processing, statistical estimation, control theory, artificial intelligence, and classic numerical methods etc. [2]. Multiple observations if correctly fused and associated then the combination of two or more sensors provides a better determination of location that could not be possible if sensors are used independently [3]. Figure 1 show data fusion of radar and FLIR sensors. In smart environments, there are locations equipped with variety of multiple sensors and often these sensors serve distinct purposes and respective data is processed by separate systems and set of algorithms. Consider the use case smart building where occupancy or temperature sensors are used for lighting or heating efficiency can be useful to the security system or vice versa. Recent standard such ZigBee make it possible to receive and aggregate data from multiple sensors; however integrated information processing with diverse set of sensor data is a big challenge [4]. The fundamental problem in remote sensing and GIS application is the way the collected information is processed. In the sequel, information fusion is concerned with how this multi-sensor data can be processed to increase the relevance of this big data. Aim of this project is to design and develop lightweight and reliable information fusion method with knowledge ecosystem.

![Fig. 1 Schematic Representation of Data Fusion of Two Sensors (RADAR and FLIR)][5]
2. Information Fusion and Mathematical Model

In remote sensing and GIS applications, there is continuous collection and processing of data from the environment for the better understanding of entities behavior [6]. It involves continuous monitoring of the occurrence of the possible events so that the proper action may be taken. In remote sensing and GIS application, there is less human intervention and intelligent machine communication drives the communication world. Machine learning and cognitive computing is an integral part of this environment which consists of location tracking, proximity awareness etc [7]. Integration of this intelligence with machine learning algorithms and techniques results into adaptive and dynamic algorithms. Mathematically it can be represented as:

Algorithm Remote_Sensing_GIS = f (I, MLT)

Where CAI – Intelligence
MLT - Machine learning techniques
I = {Identity, Location, Time, Environment, Type…} 
MLT = {Supervised, Unsupervised, Reinforcement, Learning to rank,…..}

In remote sensing and GIS application, sensor data is growing at faster rate as the use of wireless sensor network and sensors is increasing at faster rate and scale. To use this big information in real-time, first step is to understand information so that we can perform some statistics and filtering. Essentially in the context of remote sensing application, there is a need of supervised or un-supervised methods to transform the data from multiple sensor sources in order to provide effective un-supervised decision making. In the sequel, Information fusion is defined as:

“Supervised or semi-supervised transformation of information from various sources into single parameter for effective un-supervised decision making”.

From the single parameter, we should be able to produce specific and comprehensive data about the underlined entity or event. Consider Mi is different measurement from different sensor sources and P is the parameter from which the accurate inferences can be drawn. Hence, mathematically information fusion can be represented as:

\[ P = \sum_{i=1}^{n} M_i \]  

The function which will do the task of data fusion from different heterogeneous sensors is represented by f which takes data generated by specific sensor.

\[ CFD = \sum_{i=1}^{n} f(SN_iD) \]  

Where,

CFD= Complete Fused Data
f = function for data processing, data cleaning etc. Input to function is data from specific sensor and output will be structured and formatted data for given sensor,

\[ SN_iD = i^{th} \text{ sensor node data} \]

Once data is integrated from different sensory format, fusion method will convert it into representational format (RF).

\[ g(CFD) = RF \]  

Where, 
\( g \) = function which converts given CFD into RF
Therefore,

\[ RF = g(\sum_{i=1}^{n} f(SN_iD)) \]  

By equations (2), (3), (4) we can come to conclusion that, RF is composition of 2 functions g and f, such that

\[ RF = (g \circ f^n) x = g(f^n(x)) = g(\sum_{i=1}^{n} f(SN_iD)) \]  

Where,
\( o \) = composition symbol for any 2 functions
\( x = SN_iD, i = 1 \text{ to } n \) (n sensors)
\( f^n = n \text{ functions of same type (f) } \)

The main objectives of information fusion are detection and classification of tasks in applications like robotics, intrusion detection, and location estimates etc [8]. In space applications, information is provided by multiple sensory devices in order to perform intended task by the system.

Information fusion from multi-sensor data deals with combination of this multi-source sensory information into unique mimetic format. Figure 2 shows the clear relationship between multi-sensor fusion, multi-sensor integration, data aggregation, data fusion and information fusion.
Following are the observations from figure 2.

- The terms information fusion and data fusion can be used alternatively.
- Multi-sensor fusion is the subset that deals with only sensory sources.
- Data aggregation represents another subset of information fusion where the objective is to summarize or reduce the data volume.
- Multi-sensor integration deals with the application of information fusion to make inferences using sensory devices and other information.

3. Motivation

In remote sensing and GIS applications, satellites continuously provide earth observations which include spatial, temporal and spectral information. In order to get increased interpretations and better decision making, data with different characteristic needs to be combined. Consider the use case where the aim of information fusion is to integrate different data in order to obtain more information than can be obtained from single source individually. Consider the information generated by following sensors:

1. Sensor sensitive to infrared
2. Synthetic aperture radar

The information received from these two sources has completely different characteristics and there are few questions which need to be answered:

1. What is the objective of the user?
2. Which type of data is most appropriate to meet the user demands?
3. What is best methodology to fuse the information?
4. Which pre-processing steps are required?

<table>
<thead>
<tr>
<th>Area</th>
<th>Use Case</th>
<th>Scale of Users</th>
<th>Data Speed</th>
<th>Data Scale</th>
<th>Data Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space</td>
<td>Remote Sensing application</td>
<td>Big</td>
<td>Fast</td>
<td>TB</td>
<td>Complex</td>
</tr>
<tr>
<td>Scientific Computing</td>
<td>Bioinformatic</td>
<td>Small</td>
<td>Slow</td>
<td>TB</td>
<td>Simple</td>
</tr>
<tr>
<td>Financial Analysis</td>
<td>Online transactions</td>
<td>Big</td>
<td>Fast</td>
<td>GB</td>
<td>Simple</td>
</tr>
<tr>
<td>Social Networks</td>
<td>Web 2.0</td>
<td>Very Big</td>
<td>Fast</td>
<td>PB</td>
<td>Complex</td>
</tr>
<tr>
<td>Mobile Networks</td>
<td>Smartphones</td>
<td>Very Big</td>
<td>Fast</td>
<td>TB</td>
<td>Complex</td>
</tr>
<tr>
<td>Multimedia</td>
<td>Audio/Video</td>
<td>Very Big</td>
<td>Fast</td>
<td>PB</td>
<td>Simple</td>
</tr>
</tbody>
</table>
All the examples stated above show that the information generated by all the sensors is big in size, varied in type and dimensions. Equally it is also important to address issue like assurance for the information exchange between things, secure and private management of distributed data spread across multiple things, personal data auditing and enhanced audit data visualization for users to make them understand the usage of their identities and data by things and secure storage and deletion of audit data in a distributed space environment. Data science is fundamentally the science of prediction which deals with investigating different variables which can predict better performance in business perspective. Due to large scale of devices and data in the space, data scientists have large opportunities to create predictive scores across various space applications for different business horizons. Table I depicts this motivation more clearly.

4. Challenges

In the space domain, the criticality of the data and its real time processing is at prime importance. Due to economics of scale, there are large numbers of remote devices which are continuously generating big amount of data. In the view of these design issues, following are the key challenges for information fusion essentially in the context of remote sensing and GIS applications towards space domain:

1. Resources scarcity in terms of computational power, memory and other resources
2. Heterogeneity of the devices which ranges from tiny sensor node to the GPS device.
3. Scalability of the devices in the space domain.
4. Temporal and spatial correlation among the sources when information fusion and dissemination takes place at the same time.
5. Data loss in communication medium challenges information fusion because losses mean that input data may not be completely available.
6. Big data processing and management.

5. Literature Survey

There are various ways in which information fusion has been addressed. The techniques include filters, Bayesian and Dempster-Shafer inference, aggregation functions, interval combination functions, and classification methods.

In [9] authors investigate Bayesian network for data fusion for integration of sensors information into the real-time decision making process in a surveillance context. But they have only presented a simulation tool for a Smart environment. According to [9] Data fusion is “the theory, techniques and tools which are used for combining sensor data, or data derived from sensory data, into a common representational format.” Fusing data from different sources can improve the quality and the utility of information and help improve efficiency, security and functionality.

In [10] authors have proposed method to implement a distributed data fusion acquisition using a lightweight service composition model, which ensures the correctness of collaborations without a cyclic behavior. This method allows working with data in a distributed, decentralized manner.

[11] Proposes a process for military information system design. They have merged existing fusion framework and author’s experience in data fusion. They considered constrained by a determined sequence of events, detection, identification, tracking and estimate of future states (DITE) and proposes a five-dimension structure for data modeling and management: space (x,y,z), time (t) and possible worlds (w).

In [12] author implemented a multi-sensor data fusion algorithm for wildfire monitoring. Author also proposes warning based on adaptive weighted fusion algorithm (AWFA) and Dempster–Shafer theory (DST) of evidence. But author perform experiment in an indoor environment which is not as complex as outdoor environments.

In [13] author discusses three different approaches namely: Pre-Filtering, Post-Filtering and Pre-Post-Filtering are described based on how filtering is applied to the sensor data, to the fused data or both. They provide multi-sensor data fusion by combining a modified Bayesian fusion algorithm with Kalman filtering. Here they handle the problem of data uncertainty and inconsistency in a mobile robot. But they have not considered fusion of data with nonlinear dynamics.
### Table 2. Evaluation of Literature Survey

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Approach</strong></td>
<td>Bayesian network</td>
<td>Existing fusion frameworks</td>
<td>author’s experience in data fusion</td>
<td>multi-sensor data fusion algorithm based on adaptive weighted fusion algorithm (AWFA) and Dempster–Shafer theory (DST) of evidence</td>
<td>Bayesian data fusion algorithm with pre- and post-data filtering</td>
<td>Recursive fusion algorithm</td>
<td>Cluster-based Wireless Sensor Networks Using Fuzzy Logic Theory</td>
</tr>
<tr>
<td><strong>Application</strong></td>
<td>Simulation tool for a “smart environment”.</td>
<td>Process for military information system Design</td>
<td>Wildfire Monitoring</td>
<td>Handle the uncertainty and inconsistency problems of sensory data.</td>
<td>Monitoring the concentration of ozone.</td>
<td>Reduce traffic and enhance the performance of the sensor networks.</td>
<td>Lightweight and Reliable Algorithms for Information Fusion towards Multi-sensor Data</td>
</tr>
<tr>
<td><strong>Multi sensor Data fusion</strong></td>
<td>Supported</td>
<td>-</td>
<td>Supported</td>
<td>Supported</td>
<td>Supported</td>
<td>-</td>
<td>Support</td>
</tr>
<tr>
<td><strong>Heterogeneity of devices</strong></td>
<td>Heterogeneous set of visual and non-visual sensors are used. Also motion detector, smoke detector, daylight sensor are used</td>
<td>-</td>
<td>Supported</td>
<td>Supported</td>
<td>-</td>
<td>-</td>
<td>Tiny sensor node to the GPS device support</td>
</tr>
<tr>
<td><strong>Accuracy</strong></td>
<td>-</td>
<td>Supported</td>
<td>Supported</td>
<td>Supported</td>
<td>Supported</td>
<td>-</td>
<td>Support</td>
</tr>
<tr>
<td><strong>Reliability</strong></td>
<td>-</td>
<td>Supported</td>
<td>Supported</td>
<td>Supported</td>
<td>Supported</td>
<td>-</td>
<td>Support</td>
</tr>
<tr>
<td><strong>Scalability</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Supported</td>
<td>Support</td>
</tr>
<tr>
<td><strong>Robust and flexible</strong></td>
<td>Supported</td>
<td>Supported</td>
<td>Supported</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Support</td>
</tr>
<tr>
<td><strong>Computational Time</strong></td>
<td>Less</td>
<td>-</td>
<td>Less</td>
<td>Recommended to use Modified Bayesian Fusion with Post-Filtering in applications where time is an</td>
<td>Less</td>
<td>Less</td>
<td>less</td>
</tr>
<tr>
<td>Important Factor</td>
<td>Big data Processing Management</td>
<td>Saving Energy</td>
<td>Real time monitoring</td>
<td>Human intervention</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-------------------------------</td>
<td>---------------</td>
<td>----------------------</td>
<td>--------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Supported</td>
<td>Not Supported</td>
<td>-</td>
<td>Supported</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Supported</td>
<td>Not Supported</td>
<td>Reduce energy consumption</td>
<td>Not Supported</td>
<td>Supported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Supported</td>
<td>Not Supported</td>
<td>Require less computational Power</td>
<td>Not Supported</td>
<td>Support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Supported</td>
<td>Not Supported</td>
<td>Support</td>
<td>Support</td>
<td>No, Unsupervised</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 6. Un-supported Decision Making

To build the architecture based on knowledge ecosystem and make the process un-supervised, context-aware decision discovery is required in order to enhance the framework with the ability to take into account various context information when presenting to an application or software. This ability is necessary since application is not aware of the situation to take decision else it might lead to some inappropriate decision. Context-aware decision discovery tries to aim at this, by taking into account information, such as location, availability etc., when presenting the services found to the user/application. This can be done by applying a score value, $S$, to each services, according to equation (6).

$$S = \sum_{i=1}^{N} w_i f_i (x_i, x_{i, ref})$$

(6)

Where, $x_i$ is the ith context value, $x_{i, ref}$ is the ith reference or desirable context value, $f_i$ is the ith context mapping function and $w$ is the ith context weight value. Using both the context values itself, and a reference value, i.e. the desired value, enables the system to identify how close the context value is to the desired value. To determine this, a set of functions is used to achieve the desired functionality of calculating scores.

Figure 3 shows four different functions, that allow selecting values above, below, or to select values between or even avoid values.

These functions are determined by two or four parameters, i.e. $\alpha$ and $\beta$. For the below and above function only one set of parameters are used, while the between and avoid needs additional two sets. These parameters are specific, not only to the service itself, but also to the user/application.

### 7. Approach

From the literature survey, it is clear that all existing techniques of information fusion are based on filters, Bayesian and Dempster-Shafer inference, aggregation functions, interval combination functions, and classification methods. It is important to note that there is a need of intelligent information fusion method which will be lightweight and reliable so that inferences are accurate in resource scary environment. In the hostile environment, un-supervised decision making is very important and the proposed algorithms need to be incorporated with cognitive
computing methods. Solution or application for fusing the information fusion from multi-sensor data is difficult to design with traditional software development approach which uses major software packages and traditional data exchange. The proposed algorithms for information fusion will be implemented using the concept of geospatial managed objects. More focus will be first given to prototyping. Then the realistic implementation will be carried out concerning applicability. Basic computing facilities like wireless communication standards, sensor nodes and ZigBee modules and other sensing modules are available at institute. There is also alternative way to do this using Raspberry Pi. The Raspberry Pi is an amazing device that packs a fully-functional Linux computer into a single chip showcasing low power and cooling requirements. It has open source hardware platform can be connected to a local area network through Ethernet cable or USB Wi-Fi adapter. Raspberry Pi is compact, low-power, credit-card-sized computer where in single microcomputer we can integrate gateway node of wireless sensor network, database server and web server.

8. Methodology

The methodology of mechanism for information fusion of multi-sensor data involves following steps:

1. **Data Preparation**: This is the phase where the data is acquired from various sources.
2. **Information fusion**: Investigating lightweight algorithm for fusing the information from various sensor sources and integrating it into the single container.
3. **Data cleaning**: Investigate reliable data cleaning algorithm so that the transformation is applied to ensure that data is clean and structured. The data formats may be text, csv, audio, images, html etc.
4. **Analysis**: At this phase the data science team studies the data and its points of distribution using tools like R, SaS, Matlab etc. We then apply clustering for data points and try to fit the best possible model for the data.

5. Devising the un-supervised approach for steps 1-4 using cognitive computing learning methods.
6. Integration of all the modules into single distributed knowledge ecosystem architecture using cognitive computing approach.
7. The output of this phase is compared with results from historical data and also matched with similar clusters to identify any anomalies, trends and behavior of the data points.
8. Once the intended objective is satisfied with the results they are then published to the end stakeholders via visualizations, reports etc.

High level view of methodology is depicted in the Figure 4.

9. Results and Discussion

For the simulation, we have used the raspberry pi 2 Module. This Pin consists of 40 pins in which 27 pins are GPIO pins and remaining pins are used to manipulate the voltage. This complete circuit operates on 5v. GPIO pins are used to connect devices. Raspbian Wheezy operating system manages operations of the raspberry.

There are three types of jumper wires 1) Male to Male (M to M) 2) Male to Female (M to F) 3) Female to Female (F to F).

M to F jumper wire is required to connect devices to the raspberry. The LEDs are mounted on the breadboard. The cathode pins of LEDs are connected to the Ground pin of the raspberry via registers and anode pin is connected to the output pin of the raspberry so that we can perform actions like on or off of LED through raspberry. The whole code is written in Python.

The following Figure 5 shows the implementation of the proposed work. LED is connected to the raspberry- pi micro controller which acts as middleware. At the same time, it is working as data collection node. If we connect sensor nodes to raspberry- pi module the data is aggregated athe Raspberry Module.

The aggregated data is pre-processed and filtered at raspberry module with the help of raspberry module; we can access and monitor the devices at the remote place. At the other end IMAP server has been configured and provision is also made in the server to access the devices via raspberry module.
IMAP server deployed over cloud for accessing devices ubiquitously. Cloud also helps to manage load distribution among all the instances of the server. When LED is on and user wants to make it off provision has been made to switch off the LED. Raspberry module will send notification to the user on his smartphone as well as email account.

These LED can be replaced by any electronic appliances. The devices can be accessed by only authenticated users who are owner of these devices. This will become a generic communication framework.

In our experimentation, information is sensed from one sensor as well as three sensors using the setup presented in figure 5. Time window is varied from 1 second to 5 seconds in the interval of 1 second. Access time for the sensed information is measured in milliseconds. Figure 6 shows the result of time window versus access time for one sensor.

Comparison of access time for one and three sensor presented in Figure 6 and 7 shows that there is marginal difference in access time for the increasing number of sensors. This is significant contribution towards field of information fusion and there is a need of lightweight and reliable information fusion algorithm for different use cases presented earlier.

10. Conclusion and Future Outlook

This paper presented an overview of information fusion and its mathematical model. This information fusion has great potential to provide smart decision making in all Internet of Things use cases. Information fusion from multi-sensory data is challenging in applications like space and wireless domain. The proposed data science approach using cognitive computing surely helps to design lightweight and scalable algorithm for information fusion. Experimental results also show that there is marginal difference in the access time with the increasing number of sensors. This is indeed a significant contribution in the field and there is a need of smart information fusion algorithm.

Future outlook is to investigate and demonstrate the usefulness of information fusion in remote sensing and GIS applications equipped with visual and non-visual sensors. Next step is to model an intelligent remote sensing environment as a knowledge ecosystem to perform multi-sensor data fusion. Another interesting future work will be to propose lightweight and reliable algorithms for information fusion and to evaluate the reliability of the proposed algorithms based on the decision making ability in comparison with the alternative approaches.

References

1. Identity Management for Internet of Things, River Publishers
3. Data Structures and Algorithms – Cengage Publications
4. Theory of Computation, Gigatech Publications
5. Fundamentals of Programming Languages – I, Gigatech Publications

He is also the recipient of “Best Faculty Award” by STES and Cognizant Technologies Solutions. He has also delivered invited talk on “Identity Management in IoT” to Symantec Research Lab, Mountain View, California. Currently he is working as Professor and Head in Department of Computer Engineering at STES’s Smt. Kashibai Navale College of Engineering, Pune, India. He has guided more than 100 plus undergraduate students and 20 plus post-graduate students for projects. His recent research interests include Algorithms, Internet of Things, Identity Management and Security.

He has visited few countries like:

- Denmark
- Sweden
- Germany
- Austria
- Norway
- China
- Switzerland

Prof. Saudagar S. Barde received the Master in Computer Engineering (Software System) from BITS Pilani Maharashtra, India in the year 2011. He is currently working as an Assistant Professor in Department of Computer Engineering, STES’s Smt. Kashibai Navale College of Engineering, Pune, India. He has published 12 plus papers at international journals and conferences. He has guided more than 30 plus undergraduate students and 6 plus postgraduate students for projects. His research interests are Discrete Mathematics, Artificial Intelligence & Data Mining.

Prof. Poonam N. Railkar received the Master in Computer Engineering (Computer Networks) from Pune University Maharashtra, India in the year 2013. From September 2012, she is currently working as an Assistant Professor in Department of Computer Engineering, STES’s Smt. Kashibai Navale College of Engineering, Pune, India. She has published 15 plus papers at national and international journals and conferences. She has guided more than 10 plus undergraduate students and 3 plus postgraduate students for projects. Her research interests are Mobile Computing, Identity Management, Security and Database Management System Applications.

Prof. Pankaj R. Chandre received the Master in Computer Engineering (Computer Engineering) from Mumbai1 University Maharashtra, India in the year 2011. He is currently working as an Assistant Professor in Department of Computer Engineering, STES’s Smt. Kashibai Navale College of Engineering, Pune, India. He has published 45 plus papers at international journals and conferences. He has guided more than 30 plus undergraduate students and 20 plus postgraduate students for projects. His research interests are Network Security, Operating System Design & Information Security.