

Data Fusion and Bayes Estimation Algorithm Research

Yulian Gai

Modern Educational Technology Center
Xi'an International University
Xi'an, Shaanxi, China, 710077
adam_hero@sina.com

Yaping Wang

Modern Educational Technology Center
Xi'an International University
Xi'an, Shaanxi, China, 710077
xdwyp@163.com;

Abstract-This paper starts from the prospective of data processing and use of information, analysis the meaning and realistic background of processing integrated data by using Data Fusion technology. On the basis of a clear basic idea and principle theory of Data Fusion, studies and discusses its hierarchical levels from three aspects. A relatively comprehensive description of Data Fusion process is given in the paper. Incorporate with the description of the basic principles and ideas of Bayes estimation algorithm, identifies the limitation of Bayes estimation algorithm. The practical significance of Data Fusion technology in dealing with information uncertainty and incompleteness are summarized.

Keywords- Data Fusion, Bayes Estimation Algorithm, Multi-Source Data, Data Processing, Likelihood Function

I. INTRODUCTION

Along with the development of electronic information technology, in addition of the large amount of natural (physical) data, a huge amount, multi-dimension and a wide range of data have also been produced through sports, visual, tactile, large-scale Chinese-language material, knowledge of languages and the interpretation of natural language. At the same time, according to users' requirements, these data should be able to make timely changes to adapt to the objective environment corresponding to the spatial, data types changes, and make fast, accurate and timely reaction, as well as find the useful and implied information to support some kind of decisions, which greatly exceeds the capacity of traditional data processing methods. Data Fusion technology as a new type of data processing method is of particular importance.

Realization and development of Data Fusion technology is base on the principles, methods and techniques of electronics, a variety of data such as sound, electricity, light, movement, vision and language are collected by multi-sensor systems, by using fractal, estimate reasoning, artificial neural networks and other mathematical and physical theories and methods to process data comprehensively, avoiding the phenomenon of rich in data and poor in information[1],it is an important research approach in fields of electronic information, automation and computer science, it is with broad development prospects.

II. DATA FUSION OVERVIEW

A. Definition of Data Fusion

Data Fusion is an attribute fusion, it is a new data processing research discipline used to intelligently combine multi-source data from diverse and multiple sensor systems [2]. Data Fusion technology is a general name for the variety methods used to optimize the multi-source data processing. Multi-sensor data from an integrated optimized processing procedure will make an expression form and inferences about certain activities, events and situations, data after the fusion procedure are able to reflect the environment more thoroughly and accurately. These data shows features such as information redundancy, information complementary, real time information, low cost to access, etc.

There has not been a unified definition for Data Fusion, with the development of computer science technology, the expansion in its application fields and the deepening in relevant research, Data Fusion could be defined as: using computer technology, it is a data processing procedure allows the system to obtain more sufficient information than its individual constituent part, within the procedure, data with time series obtained from observing multi-sensor systems[3], will be automatically analyzed, synthesized, dominated and used under certain rules, in order to complete corresponding decision-making and estimating tasks. Thus, Data Fusion technology is an automated and integrated data processing technology, its essence lies in: by using the complementary strengths from multi-source data association, computer's high-speed computing and intelligent, so that the data observation and recognition process become relatively simple and are with improved quality. Compared with other data-processing techniques, it is able to provide comprehensive conclusion of certain environmental status, more rapidly, accurately, credibly, continuously and more comprehensively

B. Basic Data Fusion Philosophy

The ability of processing data with Data Fusion technology is a ubiquitous basic feature of human life and other biological systems. Compare with the human analysis of multi-source information processing procedures: human sensory organs equivalent of multi-sensor, human brain is equivalence to microprocessor, and human body functions as the actuator. While further analyzing the correspondence

relationship between human sensory organs and single sensors: eyes - light sensors, ears - sound sensor, nose - gas sensor, tongue - taste sensor, and skin - pressure, sensitive, thermal and humidity sensors. Human sensory organs will continually store and blend multi-source information from the natural world to give the perception of information, as the human instinct is with the ability to perceive information and integrate prior knowledge, therefore, it is able to make full sense of the current environment characteristics and the occurrence of an event from statistic, inference and estimating, at the same time, transforming the variety of information into environmentally reasonable explanation. This process is a process that the human brain using the basic theory of multi-sensor data fusion to naturally process complete and integrated problems. And this is a human adaptive process.

The basic principles of Data Fusion are similar with the processes while human brain dealing with integrated information, taking full advantages of multi-sensor resources[4], making reasoning control and use of the observational information, the complementary and redundant information from various sensor timely and spatially will be organized according to some algorithms and optimization criteria, then to generate a consistent explanation and description from the environmental observations, and finally to export the more effective information taking advantage of the joint operational effort of multiple sensors. To some extent, Data Fusion is a functional mimic of the human brain to synthetically process information.

Data Fusion technology including collection, transmission, synthesis, extraction, association and composition of all kinds of data obtained from variety information collection techniques, its main purpose is to extract useful, accurate information from a wide range of mass data, the most fundamental purpose is to enhance the capacity of human understanding of the complex unknown environment and things.

C. Data Fusion Levels

Data Fusion can effectively process homogeneous and heterogeneous multi-source data in different levels and hierarchies, which is the essential difference comparing with other data processing techniques. According to the abstraction level, data flow form and data transmission method, Data Fusion is divided into three levels: low-level (pixel level fusion), middle level (feature level fusion), and senior level (decision level fusion).

1) Pixel Level Fusion

In the pixel level fusion, pixels are used as the basic units, to process the data generated in the original data from sensors and various preprocessing stages respectively, it is the data fusion based on the original collected data. The main advantage of this fusion is to maintain the original data as much as possible, and to provide subtle data the other two levels are not able to provide, it is with high accuracy. Meanwhile, it shows obvious limitations, on the one hand, pixel level fusion requires data having accurate pixel registration, if the data sources come from different types of

sensor, it will be confronted with a large amount and complicated registration problems while processing them. On the other hand, due to the inaccuracy, non-integrity and instability of the original data, it requires a high error recovery ability while implementing the fusion.

2) Feature Level Fusion

Feature Level Fusion refers to the fusion which extracting features from original data of respective sensors, and making multi-source data classification, aggregation and integration to obtain a feature vector, and integrate the vectors. In this level, it also requires the extracted feature vectors are the full representation or sufficient statistics of data. According to its application fields, feature level fusion could be classified into two categories, target state data fusion and target feature data fusion.

3) Decision Level Fusion

Decision level fusion directly targets the specific decision-making process, the fusion results will directly reflect the decision making level. Therefore, it should start from the needs of the specific decision problems, make full use of all extracted data from feature level fusion, and apply appropriate fusion technique to complete. In this level fusion, different types of sensor will observe one certain target, each part of the local sensor will get the initial conclusion of the target from their data preprocessing, feature extraction, decision making or classification. And finally making the joint inference conclusion through related processing and decision level fusion judgment, this will then be used to provide basis for decision-making.

D. Data Fusion General Process

Data fusion process including signal acquisition, data preprocessing, feature extraction, fusion calculation, result output and some other steps, the process of data fusion is shown in figure 1.

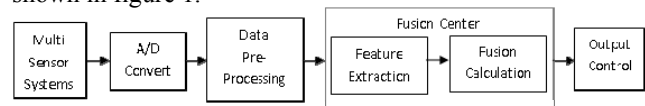


Figure 1. Data Fusion Process

1) Signal Acquisition

There are many ways to obtain signals, the signal of the target subjects can be obtained under specific environmental conditions. Due to the testing objects (such as temperature, humidity, pressure, color and grayscale, etc.) always have non-electrical quantities with different characteristics. So, first of all, those signals need to be converted into electrical signals, and then, with A/D convert, turn to the computer processed data flow.

2) Data Preprocessing

Due to the influence of random factors such as environment, so that within the digitized signal, there is an inevitable existence of some noise and inference signals, by preprocessing process, data will be filtered and distributed, in order to get rid of these noise and inference as much as possible to improve the Signal to Noise Ratio (SNR) and obtain useful signal at last.

3) *Feature Extraction*

To extract features from the useful signals after preprocessing, in order to reduce computational burden of the fusion center.

4) *Fusion Calculation*

To do the fusion calculating of the feature vector and finally output the fusion result. There are many methods for data fusion calculation, the most frequently used data representation and processing methods are generally come from related technologies, the estimation principles and recognition technology.

$$P(A|\theta) = \frac{P(A, \theta)}{P(\theta)} \tag{1}$$

By using the Bayes theorem

$$P(\theta|A) = \frac{P(A|\theta)P(\theta)}{P(A)} \tag{2}$$

If θ is considered as the unknown parameters while discussing parameter estimation problems under the condition of random variables, if it is possible to provide some additional information in advance, then the estimation of parameter θ is beneficial.

Assuming θ is the parameter of overall distribution $\rho(\chi|\theta)$, for estimating this parameter, a sample $X = (\chi_1, \chi_2, \dots, \chi_m)$ can be randomly selected from this overall distribution, meanwhile, based on the prior information of parameter θ , to select a prior distribution $\pi(\theta)$, then use the Bayes function to calculate posterior distribution $\pi(\theta)$, as the possible estimate of θ from certain feature position of $\pi(\theta/X)$, so it is estimated the most simplified inferred form of using posterior distribution [8]. This is the basic idea of the Bayes estimation algorithm.

III. BAYES ESTIMATION ALGORITHM DESCRIPTION

A. *Basic Concept*

Bayes estimation algorithm provides a measure for data fusion, it is a common method used for multi-sensor data fusion in a static environment, it is with a strict and complete theoretical basis [5]. The description of information is a probability distributions, it is suitable for the uncertainty data processing with an addable Gaussian noise (the probability density function a Gaussian distribution).

In a multi-sensor data fusion system, due to the damage of some part of data from noise, and can not be extracted in the future to restore the original data, the signal itself can not contain full and complete observation data, all these lead to the incomplete, inaccurate, ambiguous and even contradictory data from multi-sensors, there exist a high level of uncertainty. All these factors make the data fusion system not be able to obtain complete and accurate data, and the pure mathematics method is applied to analyze and solve the problems, it is also not be able to increase the data amount from the post-processing of data. Therefore, the data fusion center has to inference the data based on the uncertainties, in order to achieve the purposes of target identification and attribute judgment [6]. Bayes estimation algorithm can be effectively used for uncertainty multi-source data fusion.

B. *Basic Ideas*

With Bayes estimation algorithm, each sensor can be seen as a Bayes estimator, based on the probability principle, each associated probability distribution of a observing target can be integrated into a joint posterior probability distribution function, with the advent of the observations, and constantly update the likelihood function of the assumptions, and then provide the final multi-sensor fusion data by using the likelihood function of the joint distribution function. Within the whole procedure, likelihood function plays an important role. As we all know, likelihood is used when a certain observation target has a result, and estimate the relevant parameters of natural things. In this sense, the likelihood function can be understood as the reverse of the conditional probability.

The parameter known as θ , the probability of the event A denoted as $P(A|\theta)$

C. *The Limitation of Bayes Estimation Algorithm*

Although the Bayes estimation algorithm is able to solve the shortcomings of the traditional estimation methods, but also with a relatively great limitations, this will be include into the following three aspects.

1) *It is very difficult to define the prior likelihood probability*

Bayes estimation algorithm is able to directly determine the probability of the ones assume to be true, in case of given parameters, and allows assume the prior knowledge of a determined likelihood, it also allows the use of subjective probability as assuming prior probability. But they need to provide the distribution type and prior likelihood probability of observing target from different types of sensor in advance. It is with more difficult to determine the distribution type or it is not precisely, the prior likelihood probability is based on the large amount of statistics. It requires a very large workload statistics while dealing with a more complex problem, it often requires to solve a likelihood function, but there are also problems such as lack of precision, this makes it very difficult to define prior likelihood probability. Therefore, in many practical problems, a prior likelihood probability is generally obtained from experience it is with high level of subjective factors.

2) *It is not able to deal with the generalized uncertainty problems*

Bayes estimation algorithm requires incompatible or mutually independence among various assumption events. While there are more than one potential and conditions related events, it requires mutually exclusive among the assumptions, while makes the rapid increase in calculation

complexity, also reflects the lack of the ability to allocate the total uncertainty of Bayes estimation algorithm.

3) *Bayes estimation algorithm can not be directly used for data fusion*

When the sensor group get the consistent observing coordinate, it could directly do the data fusion from sensors, sensor can be seen as an objective description of the different coordinate frame, at this time, the observing data will use Bayes estimation algorithm for data fusion. The indirect way to solve the problem is obtained consistent group reading rotation matrix and translation vector from the sensor group.

IV. CONCLUSION

As can be seen from the above analysis study, data fusion is a new topic in the field of data processing, it is a technique to obtain and express a variety of information and process and optimize their intrinsic link. Data fusion technology processes and integrates the multi information visually to obtain the intrinsic link and regular pattern of various information, effectively solve the uncertainty and incompleteness of information, thus eliminate the useless and incorrect information, and retain the correct and useful information, ultimately achieve the data processing target of proving reliable basis for decision making.

Currently, there are many methods for data fusion, but mostly based on the stationary random process of linear Gaussian distribution. Therefore, further research and development in new data fusion algorithm, continuously improve the system performance of data fusion to solve the actual data fusion problem of non-linear stationary Gaussian distribution is needed, I believe that there are limitations for single data fusion algorithm, it will be a future research focus of integrating advantages from variety data fusion algorithms. In addition, data fusion requires the system to be

able to adapt to changes in the external world, which have a corresponding database principles and structure of the data fusion process to provide a guarantee for integration of data changes over time, space. The idea of data fusion, space, time, the database will become an important research direction in computer science.

ACKNOWLEDGMENT

The financial of this work is supports from a project with special fund for 2012 soft science projects (2012KRM65) in Shanxi Province.

REFERENCES

- [1] J. Li, L. M. Jia. Data Fusion Overview [J]. Beijing: Communications Standardization, 9th 2007.
- [2] T. M. Liu, Z. X. Xia, H. C. Xie. Data Fusion Technology and Applications [M]. Beijing: National Defence Industry Press, October, 2000.
- [3] X. Gao, Y. Wang. Data Fusion Technology Overview [J]. Beijing: Computer Measurement & Control, 2002. 10 (11).
- [4] H. C. Yan, X. H. Huang, M. Wang. Multi-sensor Data Fusion Technology and its Applications [J]. Beijing: Sensor Technology, 2005. 24 (10).
- [5] X. H. Qu, G. An. Data Fusion Method Overview and forecasting [J]. Hubei: Ship Electronic Engineering, 2003: 2.
- [6] Y. He, G. H. Wang, D. X. Lu. Multi-sensor Data Fusion and its applicaiton [M]. Beijing: Electronic Industry Press, 2011.
- [7] Department of Mathematics and Mechanics, Zhongshan University. Probability Theory and Mathematical Statistics [M]. Beijing: Higher Education Press, 1980.
- [8] J. Q. Wang, H. Y. Zhou, Y. Wu. Zhongshan University. Data Fusion Theory based on optimal estimation [M]. Hubei: Applied Mathematics, 2007. 20 (2).