Towards an Extendable Software System for Information Integration

Paul Whitney
Pacific Northwest National Lab
Richland, WA 99352
01-509-375-6737
Paul.Whitney@pnl.gov

Christian Posse
Pacific Northwest National Lab
Richland, WA 99352
01-206-782-1307
Christian.Posse@pnl.gov

Xingye Lei*
Pacific Northwest National Lab
Richland, WA 99352
01-509-372-4968
Xingye.Lei@pnl.gov

ABSTRACT

Data fusion (or information integration) from different sources is required for many technical fields, including epidemiology, medicine, biology, business, and military applications such as intelligence. Data fusion may involve integration of spectral content with imaging, text, and many other observations or with sensor data. Therefore, any software system for information integration should be extendable from one application to another. This paper describes some specific instances of data fusion, and shows how the fusion process can be viewed as multi-path and multi-stage mappings. These mappings can be organized from a software perspective into parallel development and management of four tools libraries: structure tools for data-fusion structure specifications, analytic tools for mapping methodologies, visualization tools for human and machine interface, and evaluation tools for evaluating fusion outcomes. Such software architecture will ensure the extendibility of the information integration system.

Keywords

Mappings, multipath and multistage, implementation, extendable software.

1. INTRODUCTION

Data, or information, integration exists in many fields of study, such as epidemiology, medicine, biology, business, military applications including intelligence, and non-destructive testing. For example, in epidemiology, information is often obtained based on many studies conducted by different researchers at different regions with different protocols. In medicine, the diagnosis of a disease is often based on imaging (MRI, x-ray, and CT), clinical examination, patients' description of symptoms, and lab results. In biology, information is obtained based on studies conducted on many different species [1] using many different tools, e.g., electrophoresis or spectrometry. In business, financial as well as political information is gathered and analyzed. In intelligence and military fields, data can originate from radar sensors, text messages, chemical/biological sensors, acoustic sensors, optical warnings, and many other sources [2-4]. In non-destructive testing, visual examination, eddy current testing, and other kinds of tests (e.g., ultrasonic test results) are integrated to detect the flaw and its depth and length [5].

"Data fusion," "information fusion," "data integration," and "information integration" are all used synonymously. In this paper, we use data fusion, information fusion, and information integration interchangeably, and refer to it as the results of any activities that analyze data of different sources using a variety of methodologies to understand the problem at hand.

There are several versions of the generalization of data fusion. Hall and Llinas [6, 7] point out three processing architectures for data fusion: direct fusion of sensor data, fusion of extracted features data from sensor data, and fusion of decision data formed from individual sensors. This paradigm can be generalized to all types of data, including sensor data fusion, as three types of architectures: raw data fusion, feature data fusion, and decision data fusion. Figure 1 depicts this classification. Gros [5] had already adapted this integration paradigm to the non-destructive testing situation, as data fusion at the signal level (data), evidence level (features), and dynamics level (decision).

This classification does not capture the interactions among different parts of the data-fusion process. For instance, there are situations in which featured data may need to be combined with raw data or decision level data to reveal the correlations among different attributes of variables. In this case, the process is a hybrid of data fusion and feature fusion, or feature fusion and decision fusion.

The Joint Directors of Laboratories (JDL) Fusion Working Group has defined four different levels for the data-fusion process, targeted mainly at military applications [8]: level 1 involves object assessment from the raw sensor data; level 2 takes the results from level 1 to perform situation assessment; level 3 takes results from level 2 for impact assessment; and level 4 provides resource management and feedback for refinement of the previous three levels. This definition assumes a sequential flow of information. However, there might be situations for which results from both levels 1 and 2 need to be considered to assess impacts. Most recently, a level 5 process was proposed for this model [9], which captures the human interaction with computers in the process of data fusion. JDL's classification is a good overview of the data fusion process but does not indicate how different levels of the process interact. Figure 2 illustrates JDL's four-level fusion process.

---

* Corresponding author

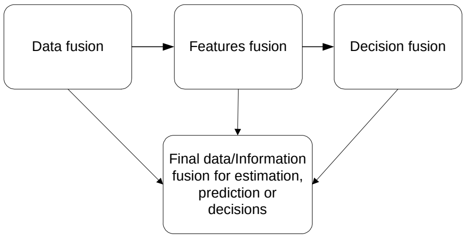

Figure 1. Data fusion classified into data fusion, features fusion, and decision fusion (adopted from Hall and Llinas [9] and Gros [5])

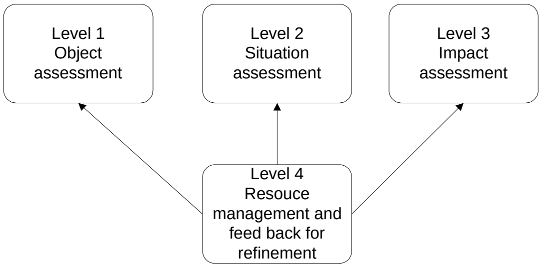

Figure 2. Joint Directors of Laboratories (JDL) process model for data fusion targeted at military applications (adopted from Steinberg and Bowman [6])

Kokar [10] proposed using category theory from mathematics to classify information fusion systems. This classification, which consists of algebraic specifications and morphisms among the specifications, classifies information fusion into three classes: data fusion, decision fusion, and data association. They claim that decision fusion and data association can be framed as a special case of data fusion. Their classification can potentially provide a framework for computers to automatically specify algorithms and to synthesize and analyze a fusion system.

Effective development and assessment of methodologies are vital to integrating the data, whether it is raw data, feature-extracted data, or individual decision data. Depending on the kind of data to be combined, both methodology and evaluation of the effectiveness of the method can vary. For raw data or extracted features, methods include parametric templating, cluster analysis, adaptive neural networks, physical models, knowledge-based methods, and others. For the decision-level data, methods include classical statistical inference, Bayesian inference, Dempster-Shafer's method, and other heuristic methods [7, 9].

The framework or classifications of data/information fusion are still evolving. However, in all the classifications mentioned above, the fusion process depends on the types of data (e.g., raw data, extracted-feature data, or decision-type data) and the methodologies used.

It is clear that, no matter how the process is classified or framed, information fusion contains many steps of data processes that entail mapping data from one domain to another with varying methodologies. In the data fusion models described, software systems to implement the model will be re-engineered each time a new component is needed, or a new application is desired, and therefore lack of the easy extendibility. In this paper, we show that data fusion can be viewed as multipath and multistage mappings. Specific data fusion activities can be performed by retrieving elements from four independently managed libraries as illustrated in Figure 3: structure tools for data-fusion structure specifications, analytic tools for mapping methodologies, visualization tools for human and machine interface, and evaluation tools for evaluating fusion outcomes. This approach provides a general architecture for a fusion system extendable. Components are added instead of re-engineered. Three examples are included to describe the overall software architecture and those four components involved.

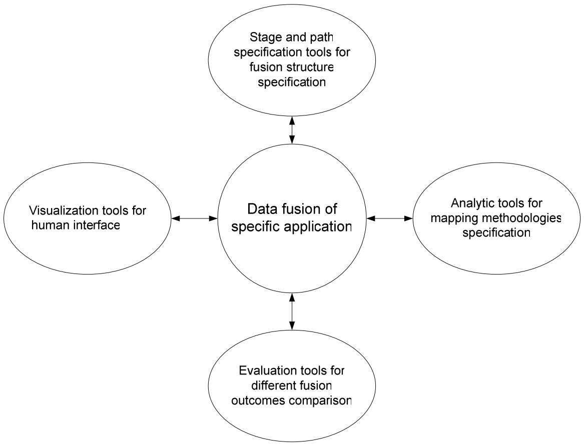

Figure 3. Schematic of unified software architecture

2. A UNIFIED APPROACH FOR EXTEDABILITY: MULTI-PATH AND MULTI-STAGE MAPPINGS

Any data fusion application can be divided into three parts: the data fusion process, data fusion methodologies, and data fusion system development. In the unified approach for extendibility and minimal software reengineering, the data fusion process is managed through multi-path and multi-stage mapping specification tools, mapping methodologies are collected in the analytic tools library and accessed through user-interface. The assembly of the final data fusion system is accomplished by accessing different components in the four tools library through the user-interface. New components specific to different tools libraries can be simply added without affecting other tools libraries.

2.1. Management of data-fusion process

As shown above, the information-fusion process can be categorized in different ways: from the perspectives of architecture, pure data flow, or mathematical category theory. These classifications each try to put a structure to the data-fusion process. However, those structures depend on the specific applications and cannot be described in one uniform classification. From an examination of the existing fusion methodologies at all stages of data fusion, it can be seen that the key component is the transformation/mapping of a set of measurements/observations from one domain to another. For example, in the case of a radar sensor used for target tracking (e.g., Stone [18]), observations in measurement space (radar) are mapped to the space of location and time period. Different sensors have different measurement spaces. To combine data from all sensors, data from measurement spaces for target tracking often need to be mapped into the same location and time space. The mapping can take many different forms, depending on the domains and the objectives of the mapping. For example, with radar sensor data, spectral data are mapped to the location and time space, e.g., in Kalman filtering, neural networks, and splines [9, 11-13]. If the purpose of the radar data is to identify a target, then mapping will be pattern recognition, which can take the form of templating, clustering analysis, an adaptive neural network, or a knowledge-based technique. If we use X to denote the original measurement space of a sensor, Z the (location, time) space, and M_X the mapping (be it Kalman filter, neural network, splines, clustering analysis, or Bayesian analysis, etc.), then we can write: $X \xrightarrow{M_X} Z$ and use $(X, Z, M_x\text{:}\ X \xrightarrow{M_X} Z)$ to represent the process of the mapping of space containing X to space containing Z, with method of mapping M_X. The elements in either X or Z can take continuous values or a discrete value, such as 0 or 1.

With the notation above, data fusion can be described as a one-time data-mapping process, or a combination of two or more mappings. Therefore, it can be a multi-path and multi-stage construction of mappings, depending on the specific data-fusion application.

The data fusion process can be managed through a mapping tools library, which will be the central point for linking input and output data/feature spaces for data fusion activities. The access to the mapping library is through the human-machine interface component.

2.2. Management of fusion methodologies

Methodologies or mappings for data fusion include pixel-level fusion, Bayesian theory, the Dempster-Shafer theory of evidence, neural networks, the Newman-Pearson criteria, fuzzy logic, knowledge-based systems, or Markov random fields [5, 9].

Data can be generated from radar or biological sensors, microarrays, or other spectral data; can be sets of new data, often referred to as kinematic and attribute estimation (extracted features) in target tracking or as biological signature generation in biological biomarker discovery; and can involve mapping methods such as a Kalman-filter, Alpha-Beta filter, least squares estimation, and principle component analysis [9, 14]. From either a set of raw or extracted features to a set of identities, the methods used include clustering, physical templating, pattern recognition, and knowledge-based database matching [9, 14-16]. For mapping a set of identities to another set of identities, methods used include Bayesian, Dempster-Shafer, and fuzzy logic [9, 17-19].

Data fusion is not a simple combination of data from all sources; it includes consolidation, re-organization, and abstraction of data [20]. The purpose of fusion is to optimize the total information content from multiple sources. Antony [19] pointed out that total information content can be enhanced in at least four approaches for the case of multiple sensors fusion:

1. New sensors can be used to provide more data, and old sensors can be improved.
2. Similar sensors can be added to provide more coverage or more confidence for observed data.
3. Dissimilar sensors can be used to complement the other sensors.
4. Domain knowledge can be used to constrain the decision process.

These four approaches can be extended to integrating information other than from sensors. Specifically, these approaches can be used to create composite knowledge signatures from multiple sources. Suppose that multiple signatures have been created from each individual source (e.g., remote imaging, text documents, spectral analysis, etc.). The composite knowledge signatures can be formed by 1) creating new signatures and improving the existing ones from raw data; 2) adding additional signatures to the existing ones to increase coverage; 3) studying the dissimilarity among signatures, and creating signatures that complement each other; or 4) using expert knowledge to facilitate the above three paths to fusion.

Mapping methodologies are managed by an analytic tools library. From a software perspective, model management is needed to simplify the programming and make the tools portable [21, 22]. For example, model management operators such as COMPOSE, MERGE, and DIFF can be linked with the first three of the above-mentioned ways (i.e., all but expert knowledge) of creating composite knowledge signatures.

The selection of a mapping methodology depends on the effectiveness measures. There is no uniform standard for choosing one mapping over another. In the case of the Bayesian approach, the evaluation is done through minimizing cost functions or maximizing posterior distribution functions [9]. The development of methodologies is coupled with the development of effectiveness assessments of methods and data-fusion outcomes. Research into the effectiveness of assessments/criteria and of methodologies is needed to advance data fusion to the next level. Since the performance evaluation methods are still evolving, a separate tools library is desirable to manage the evaluation tools and track the progress of development.

2.3. Data-fusion system development

Figure 3 illustrates the software architecture for the proposed data fusion system. At the center of the system is the user-machine interface (the data fusion of a specific application). The user interface can launch multiple windows. Through this center, libraries of mapping, visualization, analytic, and evaluation tools are accessed, and interact with each other. Current data-fusion systems are methodology-specific and need to be reworked when new methodologies are implemented. With the framework proposed here, the fusion-process structures and mapping methodologies can be developed independently. A collection of methodologies can be managed independently in an analytical tools library using model management and can be retrieved when needed. To further facilitate the fusion process, collections of visualization and evaluation tools for comparing different fusion outcomes are needed. With those four components independently managed, specific data-fusion activities can be performed by retrieving elements from those four components.

3. INFORMATION FUSION MODELS

Every time a decision is made based on analyzing data, a data-fusion process is performed. Three examples are considered below to illustrate the approach proposed in this paper: 1) the discovery of the preventive effect of fluoridated water on tooth decay; 2) the cross-referencing of information among studies on the effect of difference diesels conducted on both animals and humans; and 3) the command and control management of battlefields. In each situation, the data-fusion process, data-fusion methodology, and possible system development are discussed.

3.1 Fluoridated water and tooth decay

In 1908 in Colorado, there was a brown stain (enamel fluorosis) endemic to children ages 12 to 14 [23]. Subsequently, from 1908 to 1942, several historical studies and observations were conducted to study the cause of the brown stain. Clinical dental and health examinations were recorded that included the color of teeth, x-ray estimations of bone maturation, blood counts, eye and ear tests, tests for excretion of albumin and for red blood cells, and dental casts. It was found that the stain was caused by fluoride in the water. Meanwhile, it was observed that those with brown stain had fewer dental caries. The more brown stains, the fewer dental caries. This led to the hypothesis that fluoridated water prevents dental caries. In 1945, a prospective study, the Newburgh-Kingston caries-fluorine study, confirmed that the more fluorine in the water given to a study population, the fewer dental caries [24-29]. This study led to new public health policies recommending that fluorine be added in drinking water or that toothpastes and fluorine tables or solution be used in the prevention of cavities.

3.1.1. Data-fusion process

In this case, the data fusion process encompassed two components: the hypothesis generated and the hypothesis supported. Historical observations on brown stains and fluorine contents in drinking water and tooth decay were sequentially combined to infer correlations between brown stain and fluorine and brown stain and cavities. From those correlations, a hypothesis was postulated and supported by prospective observations of cavities and fluorine contents. Figure 4 depicts the process.

Let X_1 be the collection of observations of the severity of the brown stains and the severity of cavities, Y_1 the correlation coefficient with its uncertainty assessment, and M_1 the regression model. Then, (X_1, Y_1, M_1 : $X_1 \xrightarrow{M_1} Y_1$) represents the mapping between the severity of the brown stains and the severity of the cavities and their correlation assessment. Similarly, (X_2, Y_2, M_2 : $X_2 \xrightarrow{M_2} Y_2$) represents the mapping between the severity of the brown stains and the fluorine content in the drinking water and their correlation assessment. For Y_1 and Y_2 combined, noted as (Y_1, Y_2), and letting Z denote the decision space as 0 or 1, with 1 = the drinking water reducing cavities and 0 otherwise, and M_Y be the method of hypothesis, then, ((Y_1, Y_2), Z, M_Y: $(Y_1, Y_2) \xrightarrow{My} Z$) represents the second stage of the mapping process of the forming of the fluorine-cavity hypothesis. In this example, with observations from the prospective study denoted as W, Z as above and M_W the method of hypothesis testing, the last stage of the data fusion is $W \xrightarrow{M_W} Z$, the mapping of W to Z. There are a total of three stages of mappings to arrive to a final conclusion. Using the multi-stage mapping approach, we have the schematic shown in Figure 5, a symbolic version of Figure 4.

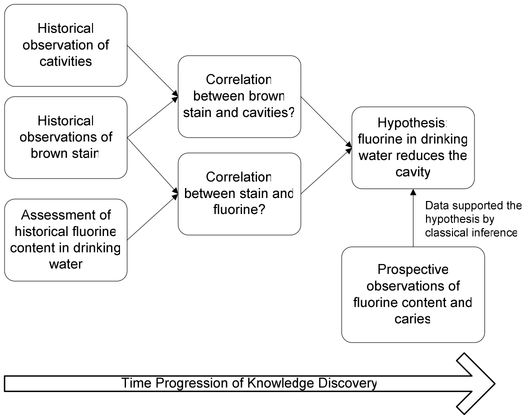

Figure 4. Data-fusion process for the fluorine and tooth decay situation

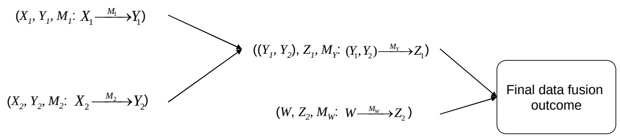

Figure 5. Schematic of multi-stage mapping of fluorine-cavity situation

3.1.2. Data-fusion methodology

Exploratory analyses with bar charts, scatter plots, frequency tables were used to model the data. Variables of interest were also identified, such as fluorine content in ppm and the number of decayed, missing, or filled (DMF) teeth per 100 erupted permanent teeth. Regression analysis was used to identify the correlations between brown stain and fluorine, brown stain and dental caries, and dental caries and fluorine. For example, pairs of the index of dental fluorosis and fluoride content in water in ppm were extracted from different cities and states and scatter-plotted to generate relationships between fluorosis and fluorine content [26]. Regression analysis of these pairs of data resulted:

$$\text{Fluorosis index} = 0.237 + 2.275 \times \text{Fluorine content}$$

with a 95% confidence interval for the coefficient of fluorine content between 1.589 and 2.961, showing the statistical significance of the fluorine content to the fluorosis content. This same regression analysis and a classical testing of the hypothesis were used to deduce a strong effect of fluorine content on dental caries.

3.1.3. *System development*

A software system for this application involves the user interface that can link the mapping tools library to create Figure 5 with executable components, say, when X_1 can be opened to examine the values and to access a visualization tools library and M_1 can be opened to access the analytic and evaluation tools libraries. Multiple windows should be allowed to correlate information from different tools libraries; the visualization tools library should contain bar charts and scatter plots; and the analytic tools library should include analysis of contingency tables, regression analysis, and the classical test of hypotheses.

3.2 Effect of chemicals on humans and animals

Cancer risk assessment of the effects of chemicals on humans is an important component in government agencies' decisions about chemical regulations. Chemical cancer risk is often studied *in vitro* or in animals. Sometimes, observations of the effect of similar chemicals on particular organs of humans can be observed in occupational studies. In the second example, 37 risk assessments of 10 chemicals were conducted on five cell lines, on animal systems, or on humans. Among those 37 studies, only four observed human lung cancers. Combining biological information and similarities among the structures of the chemicals, Bayesian approaches were employed to help extrapolate cancer risk of the chemicals to those species on which no studies were conducted [1]. Similar approaches were also applied to assess the effect of plutonium on human bone cancer[30].

3.2.1. *Data-fusion process*

The data-fusion process in this case encompasses two stages. The first stage is the estimations of the dose-response slopes and their coefficient of variations for all 37 experiments/studies separately with their respective dose-response models. The second stage is the estimation/extrapolation of the dose-response slopes and their coefficient of variations for all chemical/species combinations using hierarchical Bayesian models. Figure 6 depicts the process.

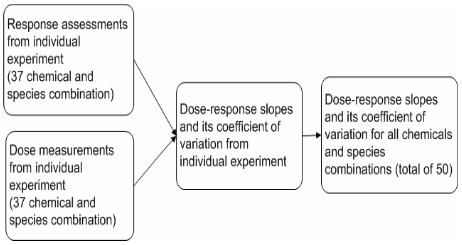

Figure 6. Data-fusion process for the effect of chemicals on humans and animals

In the case of the effect of diesel emissions on humans and animals, each individual experiment (i^th^ experiment) has mappings of the dose-response slope and coefficient of variation (Y_{i1}) to the dose (X_{i1}) through the dose-response model (M_{i1}). Letting $(Y_{11}, Y_{21}, \ldots, Y_{37,1})$ be the collection of all the (Y_{i1}), Z the collection of all the dose-response slopes and their coefficients of variation for all the species and chemicals (thus, there are 50 components in Z), and M the Bayesian methodology, then $((Y_{11}, Y_{21}, \ldots, Y_{37,1}), Z, M)$ represents the final mapping of the data-fusion process. So, in this case, there are two constructions of mappings. See Figure 7 for a schematic.

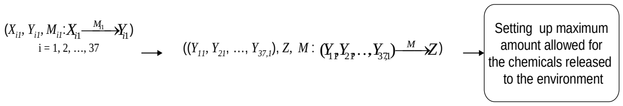

Figure 7. Schematic of multi-stage mapping of the effect of diesel emissions on humans and animals

3.2.2. *Data-fusion methodology*

A hierarchical Bayesian approach, combined with regression analysis, was used to combine the results from the studies conducted on 37 chemical and cell line/animal/human combinations and to extrapolate to all 50 chemical and cell line/animal/human combinations. This modeling was applied to the dose-response summaries obtained from the individual studies. Specifically, suppose the log of the response slope can be decomposed into three components (mean effect, species-specific effect, and agent-specific effect). Then, letting Y denote the collection of all observed/estimated log slopes; β the collection of mean effect, species-specific effect, and agent-specific effect; X a design matrix; δ the random effect from the combination of different species and agents; and ε the overall dose slope estimation error:

$$Y = X\beta + \delta + \varepsilon.$$

Bayesian hierarchical modeling assumes that, we can specify the following distributions in a hierarchical order:

$\sigma \sim$ (distributed as) $p(\sigma)$,

$(\beta. | \sigma) \sim N(b, V)$, N represents normal distribution,

$(\theta. | \beta, \sigma) \sim N(X\beta, \sigma^2 I)$, where $\theta = X\beta + \delta$,

$(Y. | \theta, \beta, \sigma) \sim N(\theta, C)$.

The estimate of β was obtained by maximizing the posterior distribution of β given the data Y, $P(\beta|Y)$, which is a mixture of multivariate normal distributions.

3.2.3. *System development*

Again, a software system for this application involves the user interface. This user interface should provide the link for a mapping tools library to create Figure 7 with executable components. A visualization tools library should contain bar charts and scatter plots. An analytic tools library should include dose-response analysis, hierarchical Bayesian method with flexibility for variety of prior distributions specifications.

3.3 Battlefield management

Development in data fusion, especially in multisensor data fusion, provides a platform for automated battlefield management. Several command and control systems for battlefield management have been developed [2, 31]. Such a system should be able to manage data from multiple sources, to correlate and evaluate the data, and to provide consistent and coherent tactical support to the commander. Data sources for such systems can include a chemical/biological sensor, radar sensor, acoustic sensor, nuclear detector, free text message, and many others. Developments of methodologies in target detection and tactical situation assessment [3, 32] are continuing. A fully integrated command and control system for real-time use of battlefield management is not too far from reality.

3.3.1. Data-fusion process

Data from each individual sensor were processed separately to provide state determination (state signature), which includes the identification of objects/targets and their position, velocity, and other physical parameters. Each individual target is further differentiated by a probability assessment and its physical state from all sensors having potential information about it. Target information is further fused/analyzed to provide a tactical assessment (e.g., friend-foe-neutral, attach, or watch) to the command team. Figure 8 illustrates the data-fusion process for battlefield management.

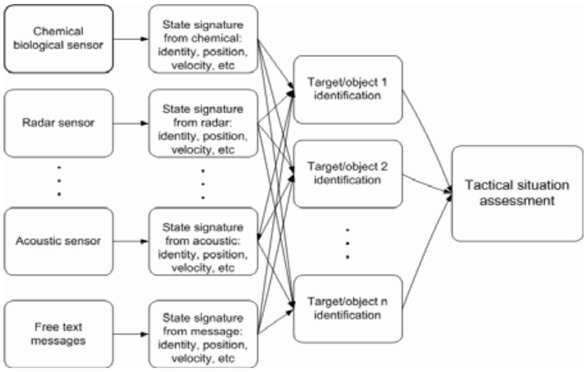

Figure 8. Data-fusion process for the battlefield management situation

Letting X_C be the collection of measurements from a chemical/biological sensor for target identification and Y_C the collection of values resulted from transforming/estimating X_C with mapping method M_C, then, (X_C, Y_C, M_C : $X_C \xrightarrow{M_C} Y_C$) represents the process. Similarly, we can have (X_R, Y_R, M_R : $X_R \xrightarrow{M_R} Y_R$) for the radar sensor, (X_A, Y_A, M_A : $X_A \xrightarrow{M_A} Y_A$) for the acoustic sensor, (X_T, Y_T, M_T : $X_T \xrightarrow{M_T} Y_T$) for the text message, etc. The next natural stage is to combine sources of information obtained to identify potential targets, 1, 2, …, n. This process can be *n* separate processes for those *n* target assessments, or it can be one process. Figure 9 depicts *n* separate processes/mappings. The last stage of the data fusion is the tactical situation assessment for commanders to take necessary actions.

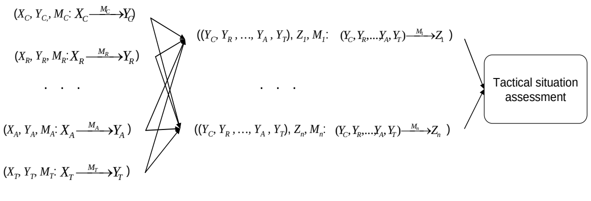

Figure 9. Schematic of multi-stage mapping of battlefield management

3.3.2. Data-fusion methodology

Physical information and templating are used by each sensor to detect targets [9] and estimate/extract qualities of the physical state (position, velocity, etc.). Weighted, least-squares splines models can be used in tracking targets [11]. A knowledge-based system with lookup tables is used to correlate results from all sensors for event detection and situation assessment. Both Bayesian probabilistic and Dempster-Shafer's evidential reasoning are used to further assess tactical situations [17, 19].

In the Bayesian approach to combining evidence of multiple sources for target identification, the probability of target identification (e.g., friendly, foe, or neutral) is computed from posterior distributions. The prior specifications of the probabilities of events are required to be mutually exclusive and consistent. However, inconsistent evidence is more realistic. Because of the uncertainties in sensors as well as situations, different sensors may provide different evidence for the same or similar events and conflicting evidence for the same situation or events may occur. In the Dempster-Shafer approach, two beliefs are computed: a) the belief of the event given by the data and b) 1–belief of the complement of the event. Because the prior specification need not be mutually exclusive, inconsistent evidence for specifying priors is allowed. However, more efficient analytic tools for implementing the Dempster-Shafer method are still needed [9, 17].

3.3.3. System development

Caito and Simmen [2] developed a prototype of a vehicle-integrated defense system. Julier et al. also proposed a software architecture for real-time battlefield visualization [33]. Such a system needs to combine both hardware and software. The hardware includes sensors, such as optical warning sensors, laser detection systems, passive missile detectors, nuclear detectors, and millimeter wave radar detectors. The software system for this application involves the user interface that can link the mapping tools library to create Figure 9 with executable components. A visualization tools library should contain bar charts, scatter plots, and time series plots. An analytic tools library should include Bayesian methods, Dempster-Shafer, physical templating, and knowledge-based intelligent models.

4. Conclusions and Future Work

Data fusion (information integration) takes many forms, from simple exploratory data-summarizing to sophisticated expert systems with evidential reasoning. Developments in both methodology and system implementation are still evolving. This paper presents three specific examples in which data fusion activities can be viewed as interactions of four tools libraries accessed through user interfaces. These four tools libraries are multi-path and multi-stage mappings, visualizations, analytic methods, and evaluations. Because methodologies are often specific to a scientific discipline and dependent on the evaluation criteria selected, new problems and new criteria will trigger the development of new methodologies. The unified approaches proposed here allow the collection of methodologies and evaluation criteria to expand and evolve independently of the fusion process and therefore providing the extendibility of the software to any applications and any addition of components to the system. The future work will be a demonstration system with the unified software architecture applied to those three specific applications with extendable architecture demonstrated.

5. Acknowledgement

We would like to thank Dr. Patrick Wong for reviewing this article and for providing valuable comments, James Weber and Brent Pulsipher for technical editing. This work is supported under the US Department of Energy contract DE-AC05-76RL01830.

6. REFERENCES

[1] DuMouchel, W.H. and P.G. Groer, *Bayes methods for combining the results of cancer studies in humans and other pecies (with discussion).* J Am Dent Assoc, 1983. **78**(382): p. 293-315.

[2] Cato, A.A. and R.L. Simmen, *Development of the vehicle integrated defense system feasibility demonstration.* Proceedings of the 1987 Tri-Service Data Fusion Symposium (DFS-87), 1987. **June** p. 402-414.

[3] Pemberton, W.G., M.S. Dotterweich, and L.B. Hawkins, *An overview of ATR (automatic target recognition) fusion techniques.* Proceedings of the 1987 Tri-Service Data Fusion Symposium (DFS-87), 1987. **June** p. 115-123.

[4] Simpson, W.R. and B.A. Kelley, *Multidimensional context representation of knowledge-base information.* Proceedings of the 1987 Tri-Service Data Fusion Symposium (DFS-87), 1987. **June** p. 239-246.

[5] Gros, X., *NDT Data Fusion*. 1997: Elsevier.

[6] Hall, D.L., *Mathematical techniques in multisensor data fusion*, in *Artech House microwave library*. 1992, Artech House: Boston. p. xi, 301 p.

[7] Hall, D.L. and J. Llinas, *Multisensor Data Fusion*, in *Handbook of multisensor data fusion*, D.L. Hall and J. Llinas, Editors. 2001, CRC Press: New York. p. 1-1 to 1-10.

[8] Steinberg, A.N. and C.L. Bowman, *Revision to the JDL data fusion model*, in *Handbook of multisensor data fusion*, D.L. Hall and J. Llinas, Editors. 2001, CRC Press: New York. p. 2-1 to 2-18.

[9] Hall, D.L. and S.A.H. McMullen, *Mathematical techniques in multi-sensor data fusion*. 2nd ed. Artech House information warfare library. 2004, Boston: Artech House. xiii, 449 p.

[10] Kokar, M., J. Tomasik, and J. Weyman, *Formalizing classes of information fusion systems.* Information Fusion, 2004. **5**: p. 189-202.

[11] Campbell, J. and J. Samaan, *Algorithm for decisions on merging and linking target tracks.* Proceedings of the 1988 Tri-Service Data Fusion Symposium (DFS-88), 1988. **1**: p. 414-424.

[12] Kay, J. and D. Titterington, eds. *Statistica and neural netwokrs: advances at the interface*. 1999, Oxford University Press.

[13] Wahba, G., *Spline models for observational data*. 1990, Philadelphia: US: Society for industrial and applied mathematics.

[14] Hilario, M., et al., *Machine learning approaches to lung cancer prediction from mass spectra.* Proteomics, 2003. **3**(9): p. 1716-9.

[15] Gasparini, G. and D. Hayes, *Biomarkers in breast cancer : molecular diagnostics for predicting and monitoring therapeutic effect*. Cancer drug discovery and development. 2005, Totowa, N.J.: Humana Press.

[16] Johnson, R.S., et al., *Informatics for protein identification by mass spectrometry.* Methods, 2005. **35**(3): p. 223-36.

[17] Carl, J.W., *Contrasting approaches to combine evidence*, in *Handbook of multisensor data fusion*, D.L. Hall and J. Llinas, Editors. 2001, CRC Press: New York. p. 7-1 to 7-31.

[18] Mahler, R., *Random set theory for target tracking and identification*, in *Handbook of multisensor data fusion*, D.L. Hall and J. Llinas, Editors. 2001, CRC Press: New York. p. 10-1 to 10-30.

[19] Stone, L.D., *A Bayesian approach to multiple-target tracking*, in *Handbook of multisensor data fusion*, D.L. Hall and J. Llinas, Editors. 2001, CRC Press: New York. p. 10-1 to 10-30.

[20] Antony, R., *Data fusion automation: a top-down perspective*, in *Handbook of multisensor data fusion*, D.L. Hall and J. Llinas, Editors. 2001, CRC Press: New York. p. 6-1 to 6-25.

[21] Melnik, S., et al. *Supporting Executable Mappings in Model Management*. in *SIGMOD 2005*. 2005. Baltimore, Maryland: ACM.

[22] Melnik, S., E. Rahm, and P.A. Bernstein. *Rondo: A Programming Platform for Generic Model Management*. in *SIGMOD 2003*. 2003. San Diego, CA: ACM.

[23] Hilleboe, H.E., *History of the Newburgh-Kingston caries-fluorine study.* J Am Dent Assoc, 1956. **52**(3): p. 291-5.

[24] Ast, D.B., et al., *Newburgh-Kingston caries-fluorine study. XIV. Combined clinical and roentgenographic dental findings after ten years of fluoride experience.* J Am Dent Assoc, 1956. **52**(3): p. 314-25.

[25] Dean, H.T., *Some reflections on the epidemiology of fluorine and dental health.* Am J Public Health, 1953. **43**(6:1): p. 704-9.

[26] Dean, H.T., *Fluorine in the control of dental caries.* J Am Dent Assoc, 1956. **52**(1): p. 1-8.

[27] Schlesinger, E.R., D.E. Overton, and H.C. Chase, *Newburgh-Kingston Caries-Fluorine Study; pediatric aspects--preliminary report.* Am J Public Health, 1950. **40**(6): p. 725-7.

[28] Schlesinger, E.R., D.E. Overton, and H.C. Chase, *Newburgh-Kingston caries fluorine study. V. Pediatric aspects; continuation report.* Am J Public Health, 1953. **43**(8): p. 1011-15.

[29] Schlesinger, E.R., et al., *Newburgh-Kingston caries-fluorine study. XIII. Pediatric findings after ten years.* J Am Dent Assoc, 1956. **52**(3): p. 296-306.

[30] DuMouchel, W. and P.G. Groer, *A Bayesian methodology for scaling radiation studies from animals to man.* Health Phys, 1989. **57 Suppl 1**: p. 411-8.

[31] Hafner, A. and B. Thompson, *Naval command and control system - AFLOAT NCCS(A), the composite warfare commander's tactical decision support system.* Proceedings of the 1988 Tri-Service Data Fusion Symposium (DFS-88), 1988: p. 80-99.

[32] Waltz, E., *data fusion functions required for tactical situation assessment.* Proceedings of the 1987 Tri-

Service Data Fusion Symposium (DFS-87), 1987. **June** p. 176-186.

[33] Julier, S. and J.K. Uhlmann, *Data fusion in nonlinear systems*, in *Handbook of multisensor data fusion*, D.L. Hall and J. Llinas, Editors. 2001, CRC Press: New York. p. 13-1 to 13-21.