A Data Association Framework for General Information Fusion*

Ronnie Johansson, Andreas Horndahl, and Magnus Rosell

***Abstract*— We extend the concept of data association as defined in the context of target tracking to general fusion and call it general data association. The main differences are that general data association encompasses heterogeneous observations (e.g. from social media text, and UAV images), multiple observations from the same data, multiple assignments of observations to entities, and association measures tailored for both observations and environment entities. We propose a framework to encapsulate different aspects of general data association, apply it to an intelligence analysis scenario, and point to some directions for future research. Our general data association framework should serve as a stepping stone for discussion and further exploration of the detected challenges.**

I. INTRODUCTION

In target tracking, data association (DA) is defined as the process of deciding whether an observation should be associated with a specific track. Subsequently, associated observations are used to update existing tracks or initiate new ones. An observation is the result of processing incoming data or sensor measurements and is given as coordinates in space and time. The association is based on the limitations of time and space, so that an observation geometrically close to the latest previous observation for a specific track is likely to be associated to that track.

We are concerned with general fusion for intelligence analysis, encompassing both heterogeneous sources, high-level estimation (e.g., abstract notions such as intentions), and distributed information processing. Here, compared to target tracking, the task of data association is more versatile, and we call this general DA (gDA). With any kind of observation, we cannot in general make the associations based on solely similarity in time and space. The association has to be defined differently for different kinds of observations.

In target tracking, observations are produced by processing of data that are usually well known with regard to their error rate. The processing or calculation, and how this propagates the errors, is also well known. This is implicitly taken into account in the association. However, in general fusion, we deal with much less well specified sources, data, sensors, and calculations. Some of the sources will be similar to those used in target tracking, while others, such as text from social networks and images from UAVs, will be radically different. Texts from social networks will be processed by text analysis algorithms that produce multiple and varied observations. The quality of these observations will depend on the text source and the algorithms. As text in itself is notoriously vague, hard to interpret unambiguously and may change over time, and text analysis algorithms are not necessarily stable or even predictively unstable over these variations, we have to be more careful in the association process. Analogously, processed images from UAVs may also give rise to many varied types of observations. If they, on the one hand, are produced by algorithms, similar issues as for text arise. On the other hand, if they are produced by human analysts the result may vary between analysts and depending on the circumstances the analyst is working under.

In our current work, we are working with supporting situation assessment for intelligence analysts. The analysts use a reasoning-support tool called Impactorium [1] and data is collected and shared in a distributed network architecture called DPIF [2]. We now begin extending the functionalities of Impactorium to explicitly deal with gDA and maintain multiple hypotheses of observation assignments (both those that are mutually consistent and those which require the generation of multiple hypotheses).

In this paper, we broadly discuss gDA as a framework which captures many of its aspects. In contrast to ordinary DA, observations are produced from various heterogeneous data sources in a multitude of ways with different quality and certainty. The observations are further disparate in nature forcing us to handle association in different ways. Our gDA framework is by no means a complete answer to general data association, but serves as a stepping stone for discussion and further exploration of the identified challenges.

In Section II, we present the background to this work. In Section III, we discuss some pertinent properties of gDA. In Section IV, we motivate and detail our framework for gDA. In Section V, we apply our gDA framework to an intelligence analysis example. Finally, in Section VI, we summarize and present future work.

II. BACKGROUND AND MOTIVATION

In this section, we discuss the role of DA in information fusion.

A. *Environment State*

Reasoning perception systems (which our intelligence analysis support system is an example of) are often suggested

* Research supported by the European defense agency (EDA), project no. B 0983 IAP4 GP (IN-4-STARS).

Ronnie Johansson (corresponding author) is with the Swedish defence research agency (FOI), Decision support systems unit, SE-164 90, Stockholm, Sweden (phone: +46-8-5550-3566; fax: +46-8-5550-3700; e-mail: ronnie.johansson@foi.se).

Andreas Horndahl and Magnus Rosell are also with FOI, Decision support systems unit (e-mail: andreas.horndahl@foi.se and magnus.rosell@foi.se).

to be concerned with three fundamental concepts: *prior knowledge*, *environment state*, and the *real world* (inspired by [3]). We consider the prior to hold information (slow-changing long-term memory) which founds a constructive bias to facilitate reasoning including general persistent facts (unlike the environment state which contains momentaneous volatile facts), logical inference rules, ontologies, statistical models, etc., which can deduce additional high-level entities. The prior can also be useful in the DA process.

As a consequence of the concept of *general* DA, we promote a *general* environment state description which represents real object states as well as non-physical relations, intentions, abstractions and aggregates. The environment state (ES) is the system's internal ("mental") representation of *real world* phenomena and consists of entities (representing all perceived things and conceived notions) grounded with some confidence in the real world, containing attributes that characterizes the entity.

Note that although the environment is aimed at reflecting the real world to a relevant and sufficient degree, these two concepts are inherently different. The world itself is naturally indifferent to its interpretation by arbitrary observers and can fundamentally be considered to be a system of elementary particles and rules that govern them. An ES, on the other hand, is a chunk of bits, structured to represent the world to facilitate effective decision making and action by means of task-relevant entities. Furthermore, while the real world is inherently "up-to-date", the ES is a composition of old and new information, annotated with metadata about its uncertainty and pedigree.

Interestingly, when considering multiple assignment hypotheses, as we do in this paper, the ES can even be a "multi-verse" of plausible, but mutually inconsistent states.

B. DA in the Information Fusion Process

In our work, the ultimate purpose of DA is to improve some ES estimation by fusing associated information. Information fusion is, briefly, the process of combining multiple pieces of information to improve the state estimation of some observed world. To enjoy the benefits of fusion, only information that correspond to the same world phenomenon (or entity) should be fused as there might be several plausible but inconsistent options. The DA problem may arise in situations where the origin or cause for the acquired information (or data)[1] is incomplete, in the sense that relating it to entities of the maintained environment representation is ambiguous.

The information fusion process, in a comprehensive description, entails both facilitating pre-processing functions and adaptive post-processing functions (e.g. acquisition of additional information). Preprocessing includes alignment (including data formatting, registration and confidence normalization) and, the topic of this paper, *association* [4, p. 474]. Note that even though normally presented as three distinct processes, they are sometimes for efficiency implemented in an integrated fashion (cf e.g. the JPDA algorithm [5]). The ultimate result of fusion is the updated (i.e. by new relevant information) and improved (i.e. less uncertain) ES.

For general fusion, we choose a very general and inclusive definition of data association.

> *Data association entails the process of relating pieces of information for synergistic processing.*

The concepts introduced in the definition require some further discussion.

- Synergy – the ultimate purpose of data association is to improve system performance. In many systems this is achieved by improving the estimate of the ES. More specifically, the ES is updated or refined by adapting it to "novel"[2] information.
- Information – the information to associate could be of different types, *observations* and environment *entities*. Hence, data association can be used to associate observations to observations, observations to entities, and entities to entities, further discussed in Section II.C.

Data association is a well-established concept in the "target tracking" field (see standard literature such as [6]). In our current work, however, we study the properties of the extended problem, i.e. gDA. In this context, it is common to talk about state estimation by means of high-level fusion, aiming at estimating not only the properties of directly observable objects but also directly inaccessible conceptual entities such as inter-individual coordination (e.g., discovering which players on a soccer field play in the same team) and operative-level events (e.g., intentions of hostile agents). This estimation could, furthermore, be performed in a system of multiple distributed and heterogeneous information sources[3] (e.g. disparate sources such as twitter feeds and UAV images) and multiple distinct environment estimations.

C. Association Types and Assignment Solutions

In target tracking, different types of DA have been described depending on the type of information to associate, measurement and track, e.g. measurement-to-track association. To generalize, we change "measurement" to "observation" and "track" to "entity". We interpret "observation" to mean data that has been generated and enriched (for further processing) and entities to mean more persistent objects maintained in the ES (typically created and updated from underlying observations).

Note that observation-to-observation association assumes that the DA is synchronized by collecting a number of observations (either instantaneously or over some period of time). This formulation of the DA problem is amenable to standard Computer Science algorithms such as *Set partitioning* etc. [7]. In general fusion, with distributed sources and multiple system objectives, observations may

[1] In this paper, we use the terms "data" and "information" interchangeably and don't differentiate between them.

[2] Novel, here, means information that has not previously been integrated with the environment state.

[3] This is in fact our case where the DPIF system distributes data to intelligence analysts using the Impactorium tool.

arrive randomly and synchronization might not be effective. Even though partitioning is sometimes feasible even when observations arrive asynchronously (if old observations are stored and repeatedly re-partitioned when new observations arrive), it appears inefficient for general fusion.

For our work, we primarily consider observation-to-entity DA, which assigns newly acquired observations to entities in the environment (or creates new ones when necessary).

Many types of DA algorithms have been proposed [8, p. 53], [9]. Some of them make a "hard" decision about which assignment to make (e.g., nearest neighbour), only select one preferred hypothesis. Others make a "soft" decision by generating and maintaining multiple hypothetical assignments. Soft decisions, which are based on evaluating and ordering multiple possible assignments, are further subdivided into single-hypothesis reasoning, which only retains one state estimate (an example is the PDA which creates a smoothed estimate over all possible probabilistically-weighted assignments), and multiple-hypothesis reasoning which retains multiple assignments (and corresponding estimates) [10]. The latter has the advantage of deferring the decision of ES until more data has been collected, but requires computer memory, processing power, and suitable algorithms to manage the explosion of alternative hypotheses. This property is especially important for general fusion where observations may be associated with diverse mutually exclusive entities for which a single-hypothesis smoothing would yield a single internally inconsistent ES.

D. Related Work

We are aware of only a few works that address gDA, i.e. that move beyond merely estimating dynamic objects. Steinberg [4] discusses DA for the abstract concepts of "situation" and "threat". The approach breaks down the association process into three steps: hypothesis generation, hypothesis evaluation and hypothesis selection. Our current study differ primarily in our focus on the details of multi hypothesis management.

In [11] abstract information (especially objects and their relations) is represented as graphs, and the result of association is updated graphs.

III. General Data Association

In this section, we introduce and discuss various distinguishing aspects of gDA.

The traditional DA for target tracking is a composite of various fixed properties, which are naturally relaxed for gDA.

1. **Data acquisition -** in target tracking, observations are often associated in synchronized batches. In general fusion, on the other hand, observations may arrive unpredictably and out-of-order.
2. **Multiple observations -** In target tracking, by our definition, data and observations are identical. In general fusion, however, we allow the same data to be processed into multiple observations (cf the extraction of multiple statistics from twitter text or various features extracted from an image). This is convenient as a data chunk (e.g., an image or document) may be composed of information about distinct entities.
3. **a) Decoupling of observations and entities -** In contrast to traditional applications of DA, where there normally is a strong assumption on the type of information (e.g. location coordinates), the prerequisites of a distributed system and source heterogeneity decouples observations from entities and introduces the problem of deciding whether data is relevant for the state estimation of an entity or not.
3. **b) Interpreted observations** – To address the decoupling property of gDA, we allow entities to "interpret" observations by customized transformation to become amenable to fusion and DA.
4. **Multiple assignments -** a single observation may be simultaneously assigned to multiple entities (unlike tracking where at most one track can "claim" an observation), as long as the assignments are not inconsistent. In tracking, there is a natural inconsistency between assignments to two distinct object entities; the observation cannot have originated from both. For instance, a tweet about an armed person can simultaneously update the categorical *weapon type* attribute of a person entity and the Boolean *armed?* attribute of a group entity (where the person belongs to the group).
5. **Association measures –** in target tracking, the measure is typically geometric distance or the probability that an observation belongs to a certain entity. Although geometry remains important, in applications with disparate heterogeneous sources, many different measures may have to be developed (that compare e.g. conceptual similarity for a pair of observation and entity types).
6. **Association constraint management –** association measures are not enough; a mechanism to decide whether observation-to-entity assignments are mutually inconsistent has to be managed.

We provide an overview of the relaxed parameters in Figure 1.

IV. The gDA Framework

To support the exploration of gDA, we propose a framework to capture many of the aspects of gDA that we have identified and listed in Section III (occurrences of these aspects are highlighted with **bold** text).

We suggest a propagation process of acquired information; how to organize that information to facilitate gDA; and finally discuss association measures.

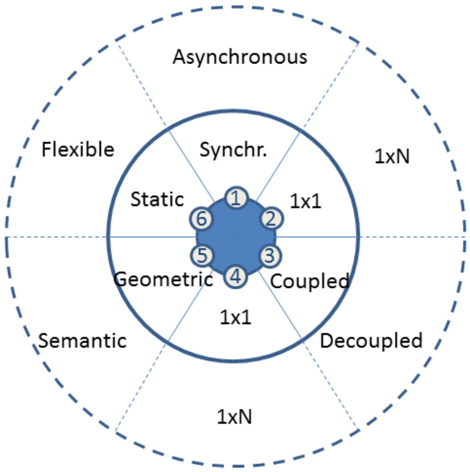

Figure 1. The inner circle represents the typical parameter values of DA (i.e., the DA values for target tracking). The outer circle (with its dashed boundary) illustrate how these parameters are relaxed for gDA.

A. *Propagation of Acquired Information*

World phenomena are measured by heterogeneous information sources such as physical sensors and web crawlers. The **acquired data** are further processed (see Figure 2) into observations by adding metadata (such as timestamps, source ID, source properties, etc.). In the case of some data, such as image and text data, **multiple observations** can be created from the same data, reflecting different features of data, or different ways of extracting the same type of feature (cf ensemble classifiers [12]).

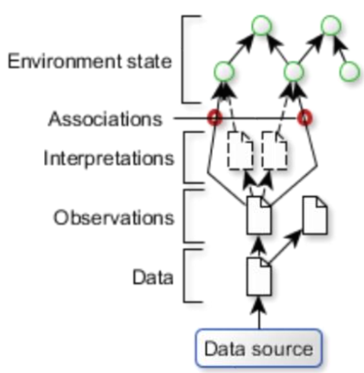

Figure 2. The propagation of information and relation to environment state through observations.

As argued in Section III, observations and entities are **decoupled**, and can in general not be "hard-coded" in advance for association and fusion. Instead we think of observations as "interpreted" (transformed for instance by data alignment, a process that may introduce additional uncertainty) to "match" an entity. Hence, an observation can be **interpreted** in different ways for matching different entities.

Finally, observations are assigned to entities and can contribute to improve the ES.

B. *Association Functions*

To address the properties of gDA, we propose three functions: *observation pool management, assignment management, and possible worlds estimation.*

Figure 3. Three proposed functions for gDA: observation pool management, assignment management, and possible worlds estimation.

1) *Observation Pool Management*

The purpose of the observation pool management is to keep track of all generated observations. We don't consider observations to be "consumed" when used to update the ES, instead we keep them in a "pool" of generated observations. The reasons are twofold: 1) observations may be shared among several entities, and 2) if several observations are assigned to the same entity, the fusion result can easily be corrected if old observations are dropped due to time limit or detected sensor errors.

The manager may, if necessary, get rid of aged and obsolete observations, and ensure the consistency between observations with the same data origin.

2) *Assignment Management*

The meaning of an assignment hypothesis has traditionally been vague, sometimes representing an updated ES based on assigned observations, and sometimes just the actual assignment. In our work, we let assignment hypothesis denote the latter, hence constituting a mapping from observations to entities.

Using one or more **association measures** (further discussed in Section IV.C), candidate observation-to-entity assignments are generated.

We consider that **multiple assignments** (both consistent and inconsistent) of observations to entities is necessary for gDA. As argued in Section III, observations are allowed to be assigned to multiple entities when not mutually inconsistent.

We, furthermore, envision a certain function, **constraint manager**, to be applied here to decide whether two or more possible assignments are possible for an observation. The exact properties of such a manager remains to be determined, but in the case of target tracking, there is typically the simple constraint that each observation can only be assigned to exactly one track (or classified as noise). This is where the traditional DA problem (presented in Section II.C) appears, which could be dealt with by either single-hypothesis or multiple-hypothesis solutions.

We argued that maintaining multiple hypotheses of mutually inconsistent observation-to-entity assignments is of importance, especially for gDA. In target tracking, the MHT [9] algorithm maintains temporal trees of possible assignments.

Given that we deal with a multiple-hypothesis solution, the assignment manager needs to deal both with detecting inconsistencies, creating new hypotheses, reshaping existing ones (if they, e.g., contain obsolete assignments), and pruning the set of hypotheses for computational efficiency.

3) Possible Worlds Estimation

Each assignment hypothesis implies a particular hypothetical ES, i.e. an estimate of a possible world state. The procedure to create an ES instance may consist of, e.g., creating entities based on the information of the world kept in the prior and fusing observations to improve state estimates. In simple cases, the ES is completely specified by the corresponding observation assignment, and can be generated "just in time" (to save memory). However, since the entities to assign to only exist within an ES and are not know in advance, the ES should preferably be readily available at all times. Additionally, in general, repeatedly generating the ES when needed may be inappropriate if the generation process consists of randomized steps yielding an unstable result.

Once available, the ES can be queried for information such as marginal probabilities of pertinent entity properties. Over whole sets of ESs, various statistics and measures, including highest and lowest values can be calculated.

C. Association Measures

The purpose of an association measure (AM), which can be considered as a binary function, formally $a\colon O \times E \rightarrow [0,1]$ from the domains of observations O and entities E, is to reflect the "match degree" between an observation and entity. Tacitly, when used for fusing, the ability to fuse observations with entities (i.e. to perform data alignment) should be reflected in the match score. Consequently, a high match degree should yield a candidate for an observation-to-entity assignment.

We are concerned with general fusion where the task of DA is more complex than in the special case of target tracking. With any kind of observation, we cannot in general make the associations based on similarity in time and space. The AM has to be defined differently for different kinds of observations and entities.

In single target tracking, there is an environment state which keeps the current track estimate. The AM can be described thus: any new observation that is sufficiently close in space and time will be associated with this track. The observation is then used to update the track estimate.

Consistent with previous work on DA, we suggest that the AM for gDA should consider two aspects of the observation, its contents and its contextual metadata. For instance, an observation may consist of temperature measurement content data and metadata about its type (i.e. temperature) and the location and time where and when the measurement was made. A candidate observation-to-entity assignment may be constructed based on only metadata, e.g. if an entity requests all temperature measurements in a certain area (e.g. a greenhouse). Some AMs may also involve the inspection of the contents of the data. For instance, the entity may not be interested in too low temperatures of too uncertain measurements. There might also be an interest in inspecting the results of fusing an observation with an entity to determine its AM. One might for instance want to check whether the observation is too contradictory to the entity's current state; a similar approach is applied by [13] (which has an observation-to-observation solution).

AMs also need to consider the entities. There might not be a perfect match between observation and entity, e.g. the type of an entity might be a car but an observation is of the type vehicle. By allowing the AM to interpret the vehicle as a car (using an ontology stored in the prior) the assignment is possible and useful attribute information of the observation can be utilized.

The matching of metadata between observations and entities mentioned above may need further explanation. It could be straight forward, only matching "literally" or using simple "translation tables". There is probably some kind of difference between continuous and real valued observations and those taking on discrete values, etc. More complex "translations" of observations, including combinations of several observations (e.g. complex event processing), are left to the entities and inference models. They could specify the need for any kind of observation in their metadata and store the values until they have enough information to calculate what they aim for.

V. Simple Example

We now present a simplified intelligence analysis scenario to demonstrate how the proposed gDA framework can be used (Section V.A) and how assignment management can be realized (Section V.B). Keywords from Section III and Section IV are **bold**.

A. Scenario

One key task in the intelligence analysis process of achieving situation awareness is to associate observations with threat indicators (i.e. entities which reveal threats). The following scenario illustrates some association issues, in terms of our gDA framework.

Geographical region A, comprising three towns, is known as a restless area with a history of riots and conflicts. The unease originates from inter-group conflicts and an unstable and corrupt government. A peace force is present in the area to prevent the situation from escalating. In order to prevent and get early warnings of possible future events, sensors have been deployed at strategic locations in the towns. The sensors detect persons, groups of person, and vehicles. In addition to the sensors, a social media monitoring system (SMMS) is also used to monitor the situation in the region. The SMMS, the sensors and direct observations represent heterogeneous sources with different characteristics.

In this scenario, we consider two different kinds of threat entities: *riot* (due to law enforcement abuse) and *fights* between the local groups. The plausibility of these threats are assessed by analyzing observable indicators that may have an indirect or direct causal link or correlation with the threats. The indicators represent entities in the ES.

Examples of indicators that can be used to assess the plausibility of a riot are: *harsh sentiments*, *seditious statements* and *observations of armed groups of people*. Indicators that can be used to assess the plausibility of a fight are: *Observation of distinct conflicting groups of people*, *Observation of armed groups of people* and *tensions between* *conflicting groups*. The fusion mechanism used for combining the factors and indicators can be based on Bayesian belief networks or another adequate fusion technique.

Subsequently, the plausibility assessment is updated after each new association is made. The actual process of fusing the assigned observations for each indicator is however beyond the scope of this paper.

In the following section, we will illustrate a set of gDA issues that arises based on the scenario and a series of events. The events are described in Table 1.

Table 1. Scenario events

Event id	Time	Event description
1	T0	The sensors detect a group of people at the main road in the town X
2	T0 + 10 m	The SMMS captures a tweet saying that there is an armed group of people in the town square of town X.
3	T0 + 20 m	The SMMS detects a set of tweets with seditious content directed towards the local authorities. Region A is mentioned in the tweets but not any of the three towns.
4	T0 + 30 min	The result of analyzing a UAV image, depicting a group and an armed vehicle, becomes available.

The first event yields an observation of the presence of a group. Even though the observation doesn't specify if the group carries any weapons, it can potentially be associated with both threat models via the *armed group of people* indicator. Even though it is not a perfect match, we would like to make use of this relevant observation. This can be achieved by maintaining **multiple assignment** hypotheses. In this case, this means that we generate two assignment hypotheses for the *armed group of people* indicator, one where the event observation is **interpreted** as an armed group and taken into account and one where it is ignored.

The second event yields an observation that can be understood as if the group of people entity detected by the sensors in the first event has moved to the town square. Alternatively, there are in fact two distinct groups. In the second alternative, the **interpreted** observation can potentially be associated with the *Observation of distinct conflicting groups of people* indicator and consequently increase the probability that there is a risk for a serious fight between the two groups.

This illustrates that, depending on how the events are interpreted and associated with entities in a threat model, the situation can be assessed differently. In this case, there are two assignment hypotheses, each based on the two different **interpretations**.

The hypotheses that there are two distinct groups and the hypothesis that there is one single group are mutually inconsistent according to the **association constraints** specified for our **assignment manager**. At some point in time, one of the hypotheses may have considerably stronger support than the other. When this situation occurs it is reasonable for the hypothesis manager to prune the weaker incompatible hypothesis.

The reliability and credibility of the source of the second event needs to be questioned. If it is unknown if a source can be trusted, the multiple world strategy can be used. In this case we can have an optimistic world where information from the source is included and pessimistic world where the information is ignored.

The third event may support the threat hypothesis that there is a riot underway in the area of interest. However, it is unknown if the tweet concerns town X. It is possible that it is a local issue associated with one of the other towns in the region which makes an association to the hypothesis of a riot in town X questionable. In this case, three mutually inconsistent **assignment** hypotheses could be generated, one for each town in Region A.

The fourth event is illustrates the case when data is processed into **multiple observation**. In this case, data yields two observations. The first is that there is a group at the main road whereas the second observation tells us that there is an armed vehicle nearby. The group may be associated to the armed group entity. Even though we lack information whether the group is armed, the concept of an armed group can be viewed as a sub concept to the concept of a group and can thus be interpreted as close enough given a generous **association measure**. Keep in mind that assigning an observation to an entity means that the observation is relevant to the entity but does not necessarily imply that the event state must be updated. Even though the result was available at T+30 min, the picture was taken at T0. This illustrates that the observation can be available in an unordered fashion.

B. *Implementation Discussion*

Assignment management is a central part of gDA. In the previous section, we speak loosely of its application, but not how it could be implemented. One approach is to allow entities to specify patterns of observation features of interest. Once a sought pattern has been detected for an observation, an assignment is created.

In [14], we implemented this assignment management approach for heterogeneous information fusion in a port security scenario, where information from sensors was fused with intelligence reports. The data and metadata used in [14] was aligned for high level fusion purposes by expressing the data as a collection of RDF-statements[4] using terms defined in a domain ontology. Conceptually, a collection of RDF-

[4] RDF stands for Resource Description Framework.

statements represent a labeled directed graph (i.e. a semantic graph). The association patterns, which can be seen as graph matching patterns, was also expressed using the terms defined in the domain ontology using SPARQL.

VI. Summary and Discussion

Briefly, in this paper, we point out fixed features of data association (DA) for target tracking which are relaxed when addressing general data association (gDA). We then propose a framework for dealing with these properties and especially multiple hypothesis solutions.

Practically, our work on DA for general fusion, gDA, focusing on observation-to-entity assignments, forms a first step to extend the functionality of the Impactorium tool to facilitate multiple observation-to-entity assignments, both consistent and inconsistent ones. The primary application is intelligence analysis and the acquisition of heterogeneous data is supported by the DPIF network architecture.

Admittedly, we have only begun to unravel the aspects of gDA. During the course of the work, we have encountered a few topics for future work that either remains to be characterized or are seemingly challenging:

1. **Constraint management** – How does the gDA decide whether an assignment hypothesis should be created or not?
2. **Hypothesis management** – Hypotheses should be stored and pruned efficiently.
3. **Possible world estimation** – How to achieve an efficient representation of multiple possible worlds? The possible worlds will have subsets of entities in common. An analogy exists with uncertain databases [15] where uncertainty in data results in multiple alternative certain databases.
4. **Performance metrics** - Metrics for gDA performance are not considered here, but metrics for traditional DA are covered in [16].
5. **Mixed initiative** – There is a need for simple yet powerful interfaces that present the current world state, including competing hypotheses. However, even with an effective pruning method in place, the number of competing hypotheses may become overwhelming. We intend to look into the problem of how to create good user interfaces as well as pruning methods in the future.

References

[1] R. Forsgren, L. Kaati, C. Mårtenson, P. Svenson and E. Tjörnhammar, "An overview of the Impactorium tools 2008," in *Proceedings of the Second Skövde Workshop on Information Fusion Topics (SWIFT 2008),* Skövde, Sweden, 2008.

[2] A. Penders, G. Pavlin and M. Kamermans, "A collaborative approach to construction of large scale distributed reasoning systems," *Intl. Journal on AI tools,* vol. 20, no. 6, pp. 1083-1106, 2011.

[3] M. Baum, I. Gheta, A. Belkin, J. Beyerer and U. D. Hanebeck, "Data association in world model for autonomous systems," in *2010 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, Utah, US, 2010.

[4] A. N. Steinberg, "Foundation of Situation and Threat Assessment," in *Handbook of multisensor data fusion*, 2nd ed., M. E. Liggins, D. L. Hall and J. Llinas, Eds., 2009.

[5] T. Kirubarajan and Y. Bar-Shalom, "Target Tracking Using Probabilistic Data Association-Based Techniques with Applications to Sonar, Radar, and EO Sensors," in *Handbook of multisensor data fusion*, 2nd ed., CRC Press, 2009.

[6] S. Blackman and R. Popoli, "Basic methods for data association," in *Design and analysis of modern tracking systems*, Artech House, 1999, pp. 365-482.

[7] R. T. Collins, "Multitarget data association with higher-order motion models," in *IEEE Conference on Computer Vision and Pattern Recognition*, Rhode Island, 2012.

[8] D. L. Hall, Mathematical techniques in multisensor data fusion, Artech House, 1992.

[9] F. Castanedo, "A Review of Data Fusion Techniques," *The Scientific World Journal,* vol. 2013, 2013.

[10] R. T. Antony, Principles of Data Fusion Automation, Artech House, 1995.

[11] G. A. Gross, R. Nagi, K. Sambhoos, D. R. Schlegel, S. C. Shapiro and G. Tauer, "Towards Hard+Soft Data Fusion: Processing Architecture and Implementation for the Joint Fusion and Analysis of Hard and Soft Intelligence Data," in *Proc. 12th intl. conf. on Information Fusion*, Singapore.

[12] D. Opitz and R. Maclin, "Popular ensemble methods: An empirical study," *Journal of Artificial Intelligence Research*, vol. 11, pp. 169-198, 1999.

[13] A. Ayoun and P. Smets, "Data association in multi-target detection using the transferable belief model," *Intl. Journal of Intellgent Systems,* vol. 16, no. 10, pp. 1167-1182, September 2001.

[14] A. Fensel, M. Rogger, T. Gustavi, A. Horndahl and C. Mårtenson, "Semantic data management: Sensor-based port security use case," in *European Intelligence and Security Informatics Conference (EISIC)*, Uppsala, Sweden, 2013.

[15] D. Suciu, D. Olteanu, C. Ré and C. Koch, Probabilistic databases, Morgan&Claypool, 2011.

[16] M. E. Silbert and C. S. Agate, "New metrics for quantifying data association performance," in *Proc. of the Intl. Conf. on Information fusion*, Salamanca, Spain, 2014.