

Challenges in Intelligence Fusion

Ulla Uebler, Hans-Joachim Kolb

MEDAV GmbH

Uttenreuth, Germany

E-mail: [Ulla.Uebler](mailto:Ulla.Uebler@medav.de) | [Hans.Kolb](mailto:Hans.Kolb@medav.de) @medav.de

ABSTRACT

Abstract – Not only the amount of incoming information is increasing but also the different sources and types of information that become available. Problems nowadays are to cope with the large amount of data to be processed manually – here some software products have been developed to support analysers to faster sort out relevant from unimportant information. Another problem is to fuse information from different sources / sensors / types to obtain a good view of the complete image. Last but not least, it is important to obtain all the relevant information the quickest possible.

This paper addresses the challenges of this process of information fusion and describes the solution MEDAV has developed.

Keywords: Information fusion, intelligence, automatic processing, data management.

INTRODUCTION

In the first part of the paper, we describe present available approaches to fuse data from different sensors and how we can provide a common basis for the analysis. The second part addresses the approach we have developed for information fusion. We describe in detail, how data are collected from different sensors and how they are processed finally to be stored in the same data management system. All data in the data management system are treated in the same manner.

The following section *Sensors* presents, how different sources and sensors can be. Section *Documents* specifies how different data types can be collected and treated with respect to their type. Section *Approaches* shortly depicts possible approaches to information fusion. The main section *Our chosen approach* introduces the approach we have chosen to fuse data from all possible sources.

SENSORS

Using modern technology, more worldwide communication is possible, and therefore a large variety of communication means becomes available. All these means of communication can theoretically be used as sensors. Some examples of the COMINT / SIGINT / MASINT / OSINT / IMINT scenario are:

- Radio monitoring: data are communicated by air with different technologies and coding/encryption algorithms; the results are wideband signals that may be further processed until information can be extracted as audio signals, Morse signals etc.
- Telephony: even at the traditional fixed line network not only telephone calls, but also data can be transmitted via phone. Mobile phones and mobile stations also communicate with audio signals or digitally encoded information.

- Internet: communication in the internet can take place in manifold ways: homepages, e-mails, voice-over-IP telephony. Types of documents can be audio signals, text signals, images and other decrypted information.

Obviously, the sources of information differ in their availability, technical access. People, however, may communicate using any of these means, even at the same time. Of course, it would be helpful to extract information from the different sensors in a similar manner.

DOCUMENTS

A philosophy is to treat any bit of incoming information as a document. Each document is described by its type and thus shows special characteristics and needs special ways to be treated (even automatically). Types of documents are for example:

- Wideband data are defined by their frequency range. Parts of the wideband data can be further analysed depending on the type of coding.
- Images can be obtained e.g. from the internet. They can be further processed by means of image analysis, e.g. to detect people in the image. A special case of image processing is OCR, where text in images is recognized and obtained as processing result.
- Audio files are usually obtained from telephony. Further processing can provide speech parts, the spoken language. A special case is the transcription of audio files, where text files are the result.
- Text files are obtained from the internet or as results from the file types described above.

Text files are the basis for all further information processing and condensing. After the analysis of incoming information, analysers process the information (manually or automatically). The produced reports are written as text files, possibly enriched with images.

APPROACHES

Different approaches of processing and analysis are possible until reports, conclusions or warnings can be generated.

- Processing of each sensor type on its own. For each source, the information is evaluated. After the conclusion is generated, the report is passed to other sources.
- Processing of each document according to its type. Each type (audio, image, text etc.) is processed in the best suited manner. This means, that the same document type – even when coming from a different source – can be treated in the same way. This of course means, that data fusion is recommended to take place before the analysis is done.

Having M sensors at different places and N types of data which have to be processed, we have to process

- with the first approach: $M \times N$ algorithms at the different places
- with the second approach N algorithms at the central station.

With respect to the fact that the algorithms need training and operator experts, it is obvious, that the second approach is easier to use.

OUR CHOSEN APPROACH

The approach presented in this paper is a hybrid approach of the possible approaches above. The process of information fusion and processing is performed in three steps. In this way we are able to combine the advantages of both approaches.

Step 1: Close to the sensor

The information is collected close to the sensor. A first processing of the incoming mass data is performed. The goal of this first step is to reduce the data amount by discarding definitively irrelevant information. We call this step pre-processing.

Using manual or semi-automatic or even automatic algorithms, additional information is attached to each collected data piece. We call this additional information *meta-information*. Meta-information is structured additional information, which can be added to the raw data normally present in an unstructured way. In a simple example, it may consist of the telephone number or geographic information, e.g.

telephone number	+49-9131-583-0
date	September 29, 2009
type	audio signal

Using this meta-information, a pre-selection of the information can be already performed. This possibly reduced amount of data is then processed at a central place, as shown in Figure 1 from the left part (close to the sensors) to the right part of the system, guiding us to step 2.

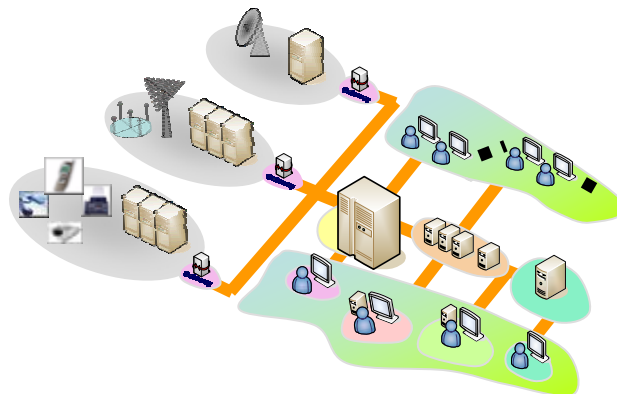


Figure 1. IFS system architecture

Step 2: Analysis at a central place

Data are now gathered from the different sensors at the central place and stored in the data management system. They keep their source-specific meta-information, and eventually obtain new meta-data, e.g.

source	HF antenna Uttenreuth
--------	-----------------------

Now, the data can be processed manually by the analysers, or automatically by special algorithms. All of the automatic algorithms result in a new entry in the meta-information. With the above example, an algorithm for audio analysis could be applied, leading to new entries

audio.type	Speech
audio.lid	German
audio.topic	Sports
audio.keyword	President, Germany

Most of the algorithms for further (automatic) processing need some type of information before they can be used, e.g.

- Basic knowledge data base (mainly for operators, where geographical information is used).
- Maps and other basic information for visualisation.
- For automatic algorithms, training data must be available before these algorithms can be used. Example: for the identification of languages in audio or text documents, documents must have been collected beforehand for any language of interest in order to perform training.

All these data can be stored also in the central data storage. Real data coming from sources can be used as training material when an update of the automatic algorithms is necessary.

Step 3: Using the gathered information

Let's now have a look at the data storage, which we is the central instance in our system. Data documents from all sensors of the complete system have been processed and are saved in the data storage system.

The data documents are not only saved as they are, but they have attached all available meta-information. This information has been obtained at step 1 or step 2 and is attached to the original document. For the analysis operators, different ways of working are possible:

- Analysing where the information came in:
 - Close to the sensor
 - At the central place
- Researching in the data storage:
 - Fast search in the meta-information
 - It can be searched through all meta-information types, also combined.
- Starting an automatic evaluation procedure, that regularly checks new inputs for their relevance and may notify analysers e.g. if thresholds are exceeded. Rules for this type of alarm can be set up by analysers.

The first of these points is conform to “traditional” approaches. The second point treats the information from different sensors in the same manner and thus can pay respect to all incoming information in the same way. For example, the analyzer can search for keywords that occur in both speech and text documents, eventually also in images with a search in the attributes of the same type.

audio. keyword
text. keyword
image. keyword

A search be combined with different criteria, e.g.

date	September 1, 2009 - September 30, 2009
*.language	German
*.keyword	Merkel
*.topic	economics

would search in all documents that have been obtained in September 2009 in German language and in the topic *economics* where the name of the German chancellor *Merkel* appears, shortly all relevant documents where the chancellor is mentioned related to economic matters. The type of document is not prescribed. With this search command, audio, image and text documents can be searched at the same time and may be presented as results.

Of course, it is also possible to perform research and evaluation only on the documents of one sensor if this is of special interest.

source	HF antenna Uttenreuth
date	September 1, 2009 - September 30, 2009

The third point is that, an automatic evaluation procedure can be performed when rules are stated by analysers, also including statistical matters.

Example:

- Notify whenever the name *Bin Laden* appears in any text file (the search can be performed in the meta-information or in original text files at the same time)
- Notify whenever, if more than 20 % of all audio files are in German (if they usually would be 10 % in a certain scenario).

Results of the analysis by human operators can also be stored in the data storage. They can be researched just in the same way as the other meta-information.

Comment.XY	This document is the same as document pf09ko.001; the information of this document is contradictory to the information in pf09ko.002.
------------	---

Since many of the algorithms for the extraction of the fused documents are done automatically, they provide a certain percentage of errors – this cannot be avoided, even not if all would have been processed by humans.

ESTABLISHING AN IFS SYSTEM

This section describes the practical aspects of how such a system for information fusion and analysis can be used. Although practical experience is given, this example is fictive.

In order to establish a system, a sequence of steps must be processed. The user must define the type and amount of incoming data, the type and amount of processes to be undertaken (automated or manual). The storage of data must be defined as well as the role of human experts in the system.

Definition of the sensors

According to current and future missions, a set of sensors must be included in an IFS system. These sources may consist of antenna, satellite, and internet input.

Not all data captured by different sources will be used for analysis, but only those relevant for the current mission. Still, the system will be flexible to process more/other input data for more missions than the current one.

Definition of the involved classification algorithms

Automated classification algorithms are important for the analysis of mass data, because the amount of data that must be analysed manually is dramatically reduced.

Depending on the mission, it must be defined, whether text, image, or speech messages will be treated and how they shall be analysed.

One criterion depending on the mission is to state the frequency of occurrence of the message types as well as the availability of algorithms for the respective task. For statistical approaches, the classifiers must be trained and thus, sufficient training material must be available.

If no satisfying algorithms are available, a way must be found to feed the process of analysis by operator decisions.

Data storage

The raw data from the sensors needs a lot of disk space. Therefore, the data are often removed after a while (some days, some weeks). The data resulting from the analysis, e.g. reports must be kept for a longer time, these data may be even stored in an archive which does not have any deletion procedure. Alternatively, some data may have a certain expiration date, after which they will be deleted or be checked manually.

Some data, e.g. speech files may be stored for a longer period in order to be able to perform a retraining of a speech classification algorithm. Other data may also be kept to a certain amount for quality measurements etc. These decisions are made by the customer; the way of storing resp. deleting data is manifold and needs a high sophisticated data management model. Of course, redundancy of data is also needed for many data types in case of system failure.

Human experts

Despite all computer algorithms and solutions, human experts are still the central point in any information fusion and analysis systems. The role of human experts can be the following, among others:

- Sensor data: look/listen, analyse, transcribe
- Condensed data: draw conclusions, generate reports, rating of importance and threat
- Expert for automated procedures - evaluate algorithms, retraining, evaluate manual vs. automated analysis

These roles – although there are many more roles – already show that human experts may have quite different roles in this systems. The system must be flexible enough to take maximum advantage of the expertise of humans and to use their abilities at the right place in the system.

Updates

With new missions and new knowledge, it becomes important to update the configuration of the running system. Different types of system changes are possible:

- New mission with new parameters and new setting of importance, e.g. new languages, new regions are higher rated now.
- Regular updates by evaluating the performance of the system, especially when using automated algorithms.

For a continuously good performance, it is important to regularly check the performance, eventually retrain automated algorithms, and to check the mission parameters.

Typical proceeding

1. A proceeding using the principle of the intelligence cycle as shown below can be the following:
2. Request from leadership: a mission is defined and started.
3. Planning: the type and amount of sensors is defined, a finer definition of omitting unimportant issues in order to reduce data is done. Algorithms for automated processing are defined, and trained, if possible.
4. Collection: data assumed as possibly important are chosen. Data which are definitely considered as irrelevant, are discarded. At this step, already automatic classification becomes important in order to add meta-information to each bit of information wherever possible.

5. Processing: According to a defined sequence of available information, data coming from the sensors are processed. This process contains both automated and manual processing, i.e. humans are also involved in this process, supported by computer algorithms.
6. Analysis: this step is mainly performed by humans. The information gathered in the previous steps is evaluated, additional meta-information may help; software tools may be used for generating statistics etc.
7. Reporting and Dissemination: After the analysis, a report is generated and a conclusion is drawn, the results of the mission are disseminated.
8. As a consequence, a new mission may be started, and the information cycle begins again.

CONCLUSION

Traditionally, data are evaluated directly at the sensor. Having more sensors, it becomes difficult to analyse all the data in a similar manner and at the same time.

We present an available system that can be configured by the customer according to different sensors and types of gathered data. Central point of our system is that a pre-processing does a preselection of data at the sensor. All documents can be stored centrally and have a set of meta-information with them, which can be elaborated depending on the available automatic algorithms and on manual analysis.

BIBLIOGRAPHY

- [1] Carr, O. and Estival, D. (2003). "Document Classification in Structured Military Messages", Australasian Language Technology Workshop. Melbourne, Australia.
- [2] Deloule, F. (2007) "Data Fusion for the Management of Multimedia Documents". Proceedings of the 10th International Conference on Information Fusion, Quebec, Canada, 2007.
- [3] FM 34-8-2 Intelligence Officer's Handbook, Field Manual, Headquarters of the Army, Washington DC [pf09ko.0c6]
- [4] Uebler, U. (2006) "A Speech Classification System". In Information Fusion for Command Support (pp. 12-1 – 12-10). Proceedings RTO-MP-IST-055, Paper 12. Neuilly-sur-Seine, France: RTO.
- [5] Uebler, U. (2001), "Multilingual speech recognition in seven languages," in Speech Communication, Aug. 2001, pp. 53–69.