GOOSE: Semantic search on Internet connected sensors

Klamer Schutte[1], Freek Bomhof, Gertjan Burghouts, Jurriaan van Diggelen, Peter Hiemstra, Jaap van 't Hof, Wessel Kraaij, Huib Pasman, Arthur Smith, Corne Versloot, Joost de Wit

TNO, Oude Waalsdorperweg 63, The Hague, Netherlands

ABSTRACT

More and more sensors are getting Internet connected. Examples are cameras on cell phones, CCTV cameras for traffic control as well as dedicated security and defense sensor systems. Due to the steadily increasing data volume, human exploitation of all this sensor data is impossible for effective mission execution. Smart access to all sensor data acts as enabler for questions such as "Is there a person behind this building" or "Alert me when a vehicle approaches".

The GOOSE concept has the ambition to provide the capability to search semantically for any relevant information within "all" (including imaging) sensor streams in the entire Internet of sensors. This is similar to the capability provided by presently available Internet search engines which enable the retrieval of information on "all" web pages on the Internet. In line with current Internet search engines any indexing services shall be utilized cross-domain. The two main challenge for GOOSE is the Semantic Gap and Scalability.

The GOOSE architecture consists of five elements: (1) an online extraction of primitives on each sensor stream; (2) an indexing and search mechanism for these primitives; (3) a ontology based semantic matching module; (4) a top-down hypothesis verification mechanism and (5) a controlling man-machine interface.

This paper reports on the initial GOOSE demonstrator, which consists of the MES multimedia analysis platform and the CORTEX action recognition module. It also provides an outlook into future GOOSE development.

Keywords: Sensor Interpretation, Big Data, Semantic Web, Sensor Discovery, Image Primitives, Ontology Based Matching, Semantic Gap, Scalability

1. INTRODUCTION

An obvious trend is that all sensors are connected by the Internet. With a huge amount of sensors, such as cameras, connected through the Internet an ocean of data will be generated automatically. This growing number of sensors pumps large amounts of location based, time tagged streams of unstructured data into the web, all potentially containing valuable information for a broad variety of (future) users. It is now estimated [1] that about 85% of all produced data consists of unstructured data (email, video, blogs, voice and social media) most prominently produced by devices such as sensors, tablets and smartphones.

Forecasts (Mc Kinsey Global Institute, 2011) [2] indicate that in the coming years billions of devices (smart phones, tablets, smart appliances), the majority of which contains one or more sensors, will be connected to the Internet generating an ocean of data which could be transformed into valuable information. If one can find the sensors, tap into their data stream, and interpret this sensor data to answer one's questions.

This immense pool of data resources will change the way we work and live. To benefit from this vast expanding data space, the grand challenge is to construct new capabilities that enable organizations, society and individuals to extract new and meaningful information from the data repository to feed their decision making and work processes. For military applications especially information serving the instantaneous need to enhance situation awareness is regarded valuable when time critical action is required. In such cases information extracted from all available and relevant IP enabled sensors in the Internet may be vital to successful mission execution. This requires tools to gain access and to exploit the huge collection of camera data streams on the web similar as to finding and using the right webpage with the current

[1] klamer.schutte@tno.nl; phone +31 888664109; www.tno.nl

Next-Generation Analyst, edited by Barbara D. Broome, David L. Hall, James Llinas,
Proc. of SPIE Vol. 8758, 875806 · © 2013 SPIE · CCC code: 0277-786X/13/$18
doi: 10.1117/12.2018112

Internet search engines. From the observed trends in data and the Internet of Things we anticipated a new demand for Google™ type of search engines specifically conceived to answer a user query precisely through searching the entire web of sensors and selection of all relevant sensors and associated data streams to extract- and deliver quality information on demand.

A GOOSE system links user queries to information from all web based sensors to deliver the right answers at the right moment to any user. It provides a GOOgle™-like search engine for SEnsors.

This paper will introduce the GOOSE system ambition and the research objectives for the project. Starting from a reference use case the main characteristics and technology overview are given. Two Big Technology Issues as major challenges for building the GOOSE system: ***"unlimited" scalability*** and the ***semantic gap***. A summary of the research pathways is given.

2. USE CASE

In normal day-to-day life everybody is supported with a smartphone, and continuously is provided information by tweets and photo's uploaded by friends to Facebook™. GOOSE targets similar functions to disclose and share public and owned information available within the coalition force. The military use case serves as a reference for the design of the system.

GOOSE-type capabilities in the Military domain [3] are needed to access, process and use data from all relevant (imaging) sensors discoverable in the coalition network and public internet covering the area of interest. These data streams are additional sources of valuable information to own deployed military sensory systems. A GOOSE system exploits data streams from other sensors in the area of interest as an complementary source of data for intelligence and situation assessment purposes. This supports information operations and decision making at various levels of command, anytime and anywhere. The capability to search any available (image) sensor within the coalition and the public internet makes the over-all situation assessment much more flexible, faster and efficient than the current existing ISR stovepipes. Serving many users simultaneously, GOOSE allows various levels of command, from strategic level downward to platoon peloton level or even individual soldiers, in order to obtain answers to their specific questions, similar to an Internet search engine serves many users simultaneously with the answers to their specific queries.

Information can be shared amongst different units and even between many different forces and organizations that need to access and search in a common repository of sensor data. This makes the use case even more significant if the system supports data exchange with other information centers. With the global internet sensor revolution new data sets become within reach of military observation operations by sourcing public sensor data on the net generated by all sort of new social media and sensor communities.

Public or 'open' [4][5][6] sensor information on the Internet drives the exponential growth of the data space. Below we provide examples how this data space can be exploited in the military domain:

- Enhance support response operations during a bombing incident follow-up, finding all relevant events in the area of interest during the relevant time frame to capture the adversary units or stop further threats.
- Surveillance queries to find abnormal situations and changes in Ad Hoc situations, e.g., to support road clearance.
- Maritime Situation management, query to find/detect threats e.g. pirates within very large area with very large number of sensors and complex situations.

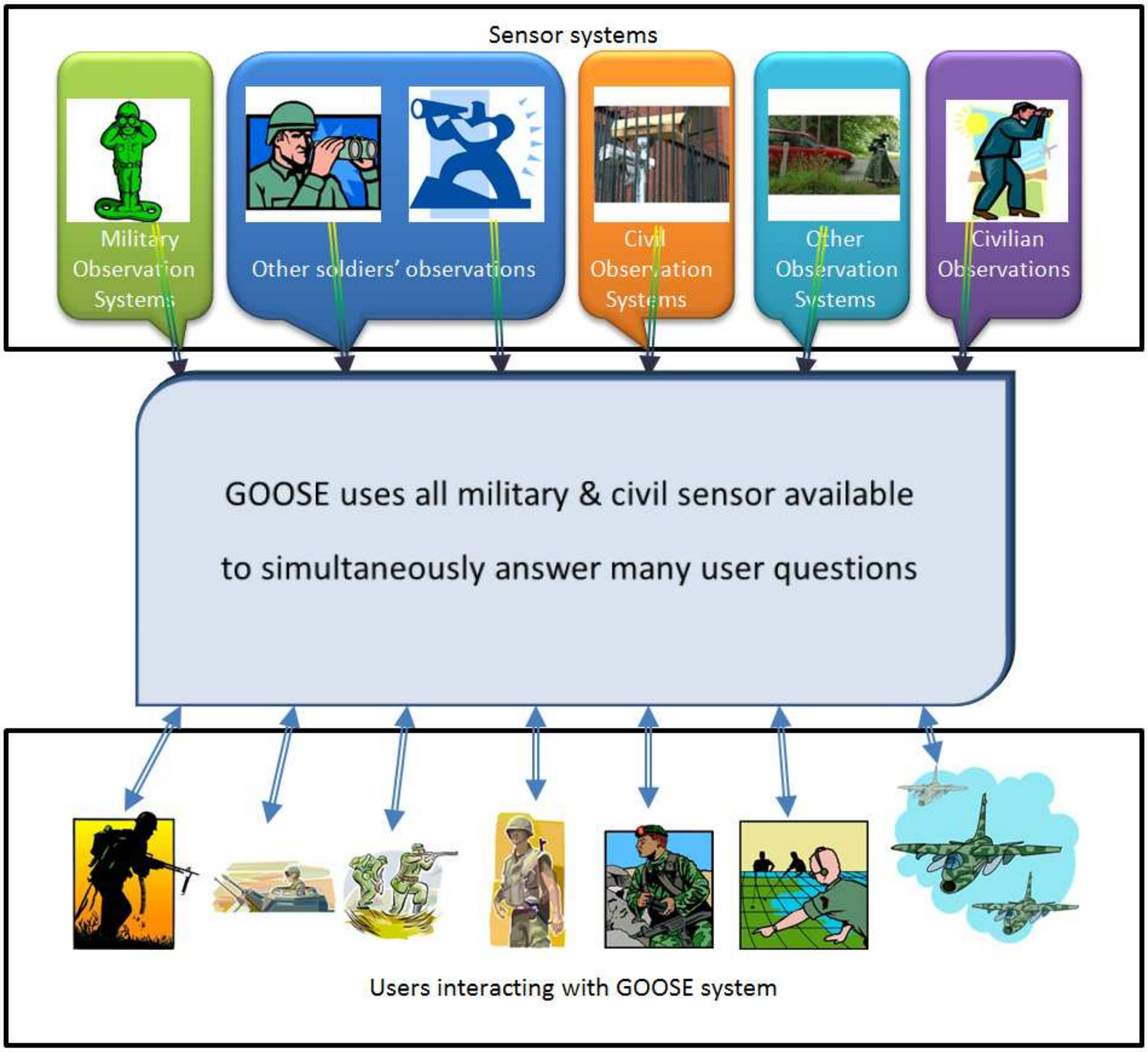

Figure 1: GOOSE top level view, the system linking all military and civil sensors in the Internet with the search engine to answer all user queries simultaneously.

3. GOOSE SYSTEM CHARACTERISTICS

The system will provide the capability to handle **any** user query for **semantic search** to retrieve **any** relevant information from **any** relevant sensor streams in the **entire Internet** of sensors.

This means a processing pipe which includes autonomous extraction of information from real-time captured image sensor data and storing the outcome in a web of data for semantic based search. Automated interpretation and information extraction is indispensable to deliver the right answers to the user query. The main functional characteristics, as depicted in Figure 1, are:

- to select the relevant sensors and their data streams, having applied the right processing, anytime and anywhere,
- to maintain the sensor database by automated enrollment of new sensors,
- to scale the computational power to accommodate any number of user queries at any time,
- to provide a semantic user interface -front end- for queries to task the search engine to select the relevant sensors and to extract the right information in an easy universal way,

- to extract information and content from sensors data streams by adaptive intelligent processing based upon semantic queries.

In contrast to legacy and state-of-the-art system such as described in [7][8], the GOOSE system we present works both with closed and proprietary sensor networks used in private, public and/or military domains as well as with the public open Internet of sensors. A sensor look-up service equivalent to the current Domain Name Service to register and manage sensor URL mapping to IP addresses is needed. This mechanism is essential for discovering "every" accessible sensor to maintain - 24/7 - the registry of all searchable sensors.
To support the semantic queries especially sensors with very high information content are of interest. Given current deployment of sensors this typically are imaging sensors with their majority camera systems. Automated video analytics technology is key to condense the enormous quantity of data from multi camera streams in real time. Once analyzed the derived information - meta data - has to be indexed in preparation to be "searchable" by anybody authorized to access the sensor data network.

4. TWO BIG TECHNOLOGY ISSUES

The analysis of the main characteristics has identified two Big Technology Issues as major challenges for building the GOOSE system: ***"unlimited" scalability*** and the ***semantic gap***.

4.1 Scalability

To build a system capable of extracting information on demand for millions of users from the Web of Sensors delivering the right answer to the query it must be assumed that all relevant sensors in the web will be (become) "***searchable***". Therefore, each sensor must be connected to the web and enabled to respond to internet data exchange standards. The system also needs a scalable architecture for the required computational resources to cater for all user queries at any time. Scalability means dealing with an ultra large – and growing – number of:

- Sensors, that internet enabled, fixed or mobile, private or public,
- Users, professionals like warfighters,
- Queries, in human friendly format,
- Mission scenarios.

A key aspect of a GOOSE system is to *simultaneously* support many users, queries and mission, allowing re-use of sensors to handle requests by external users unprecedented by legacy solutions. To ensure the value of GOOSE support in future operations and missions while both the search query as well as the type and number of sensors are not pre-defined at the moment of design, the system must be able to answer any kind of query in the future without restrictions due to limited insights at design time.

4.2 Semantic Gap

To retrieve the best available information for the user from the Sensor Web requires the system to transform a human friendly query into a "complete" set of operators to initiate all processes to find, select and access any relevant sensor data set. Unlike current search engines which generally work on matching the users query "key" words to vocabularies derived from indexed web pages in the search engine's repositories, the system must draw on (automated) interpretation of the user's query into "context awareness" parameters to search for the right information. To process true semantic queries GOOSE must include mechanisms like:

- Decomposition of user queries into semantics to facilitate computational tasks,
- Abstraction of information from sensor data sets towards users' answers.

For any given query the system generates answers on the basis of multimedia information extraction processes and relevant ontology/vocabularies which link all relevant data sets to deliver results in (near) real time. Each human friendly query must invoke an answer from the system presenting immediately usable ("consumable") information.In fact, this is the ultimate value of the GOOSE system for the user in general. This capability can be considered as is a kind of (virtual) "buddy" available at any time to collect and present valuable information derived from the sensor web, even augmented with results from additional information retrieval services.

5. TECHNICAL OVERVIEW

The design of an architecture which provides the necessary components to construct the complete chain of information extraction and processing pipelines to perform semantic search, filtering, analysis and ranking of every relevant sensor in the web presents quite a challenge. An outline of the proposed architecture is presented in Figure 2. The main areas of research are sensor discovery, information extraction and semantic search & retrieval. Notionally sensor discovery takes place in the top part of Figure2; semantic search & retrieval in the bottom part of Figure 2; and information retrieval is split over both the top and bottom parts. These areas are explored in the GOOSE project and results will be integrated in an experimental model to test and demonstrate the functional characteristics while building experience on the impact of real world constraints for performance of the system. Initial work on GOOSE has started at TNO in 2012, leading to a baseline architecture and a preliminary demonstration (see Figure 6) where the CORTEX [8][9][10][11][12][13] visual analytics have been combined with the MES (Metadata Extraction Services) multimedia infrastructure platform [14]. CORTEX here has the role of an information retrieval module where the MES framework is transformed to fit the Figure 2 architecture. This architecture currently is extended with technology for efficient search with visual examples [15] evaluated in the context of TRECVID [16].

5.1 Sensor discovery

Ideally a sensor on the web should publish itself with a standardized list containing many parameters which allows the GOOSE system to index and describe the sensor properties to assess its relevancy. Processing the user queries must lead to the detection of sensors of interest which match the specific user query search primitives. Structured sensor discovery mechanisms are already under research to achieve common standards for sensors in the web. In the GOOSE project sensor discovery is addressed by Sensor Web Enablement and the European funded project iCore [17].

Sensor Web Enablement

For sensor discovery the effort and results of the Sensor Web Enablement [18] (SWE) are explored. This effort is directed to uniform protocols for integration of all-kinds of sensors into the web infrastructure to enable applications to discover and access the sensor data. The SWE initiatives include, amongst others, solutions to enable enhanced interface between applications and the sensor and its data on the basis of a model language SensorML [19]. SensorML provides basic standardized ontologies for describing generic sensors and measurement processes, but application specific ontologies (e.g. describing concepts like soldier, or patrol car, etc.) are not standardly available. The open structure of SWE allows the definition of local or domain-specific ontologies which will not be enforced centrally. Sensor descriptive ontologies are expected to be constantly evolving due to the different needs and constraints. The SWE principles do allow adaptations to future needs.

iCore

For the implementation of sensor discovery mechanism the results of the iCore project [17] can be used. The iCore objectives are to establish an architecture for the Internet of Things to effectively support:

- flexible (ad-hoc) multi-use of sensors and actuators such that the integrity of data is not impacted,
- multi-party use of sensors and actuators with access control, authentication and billing such that multi-party use is safe and stimulated by economic drivers,
- semantic technology to discover and combine sensors and actuators.

5.2 Information extraction

Information extraction must adequately address the question: *How to find the "**most** relevant" sensor(s) in the GOOSE index?* This is on the basis of the generated data sets containing meta-data derived from the raw content stream. The relevance of a sensor is based on its geo- and spatial information, performance characteristics and the primitives it can recognize. The system is set to build processing chains automatically on the basis of the proprietary Remote Data Access (RDA) implementation of TNO to build flexible high performance real-time image processing chains for intelligent detection, fusion, ranking and selection of the data sets accumulated in its repositories. One of the most challenging aspects of GOOSE is the construction of primitives and concepts with sufficient detail that allow effectively screening scenes of interest in the sensor data set. Automated concept constructing to represent unique characteristic features for detecting the right sensor data set is inevitable to answer the queries simultaneously from many, many users.

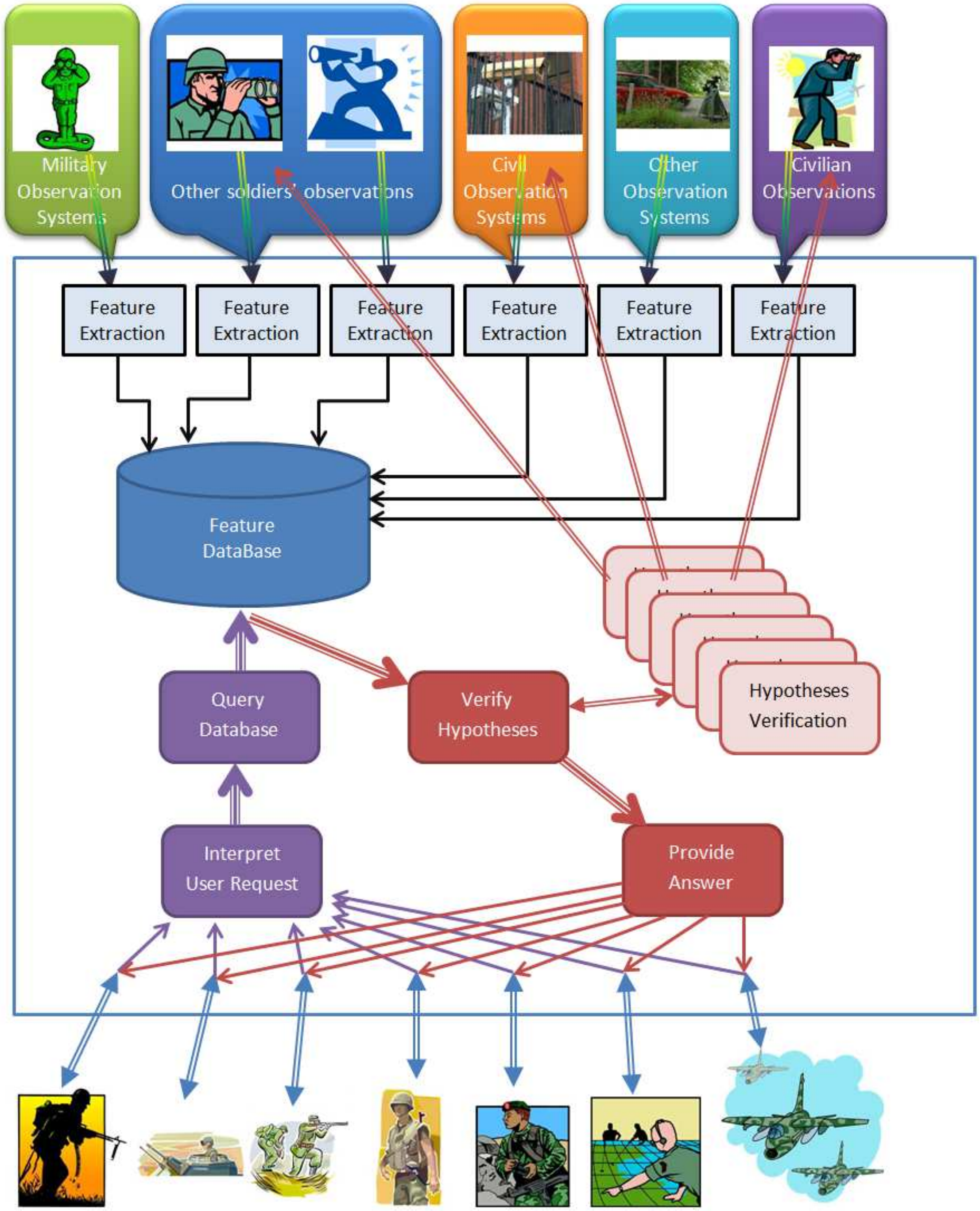

Figure 2: Proposed GOOSE architecture. All sensors provide generic low-level primitives which are stored in the Feature Database. User requests are translated to queries in this database; all hits are possible answers. These are verified by sensor-specific Hypotheses Verification processing. Verified hypotheses are provided as answers to the users, who optionally can refine their request to obtain more specific results.

5.3 Adaptive Querying of Sensors

The following queries are typical in modern warfare – yet no current military system is capable of providing adequate answers to them:

- *"I am in the city and I don't have sensors on or with me – is somebody covertly approaching me?",*
- *"Is there somebody with binoculars close to road A?",*
- *"There is a group of adversaries in the city – where are they?".*

Our research goal is to propose a GOOSE system that is able to handle such queries in the near future. Below we sketch the research pathways that we will take in our future work.

The GOOSE system should answer a broad range of military relevant queries about the mission theatre in near real-time. Many sources of information are available to the military: maps, intelligence reports, local websites, newspapers, photos, briefings, and we project that in the near-future military, civil and personal sensors are ubiquitous.
Query-adaptive search in big data is an active field of research with many open problems. Typical military queries involve locations, events, time, (groups) of people, vehicles, and specific categories of items such as weapons and equipment. The range of queries is too broad to pre-learn good categories of answers to each of them beforehand. Yet, the typical military queries can be broken down into meaningful 'query classes' (events, actors, vehicles), which promises to be very effective to load a specific model to answer class-specific queries (e.g., events). The search for locations on maps, buildings in land registers, crowded areas in camera imagery, tracking of individuals, is solved or will be solved soon. For events, groups of people, vehicles and items, advanced detectors from sensory data are readily available (e.g., video concept detectors). Both geo-spatial referencing and personal context (e.g., location of self) have been exploited effectively in narrowing down the search for concepts. These capabilities answer parts of the queries above, but provide no good answers to them. The critical query elements are 'somebody covertly approaching A', 'somebody with binoculars close to position X', 'group of adversaries somewhere'. To the best of our knowledge no systems deal with such advanced queries yet.
Our research targets the principle question: *"What is so hard about these military queries that it makes them go beyond the state-of-the-art?"* And: *"Which research directions lead to solutions for them?"* The query elements which should be solved are described in the following scenarios in more detail:

'Somebody covertly approaching A'

Person re-identification is becoming a standard tool in imagery-based situation awareness. This enables tracking across multiple cameras. Yet tracking alone does not solve the query of searching somebody who is approaching A.

- **Challenge #1**: The search engine has to infer from maps and registers the various ways in which A can be approached and whether the currently observed track is likely to be one of them and not any of the thousands of other tracks in the city that do not approach A.
- **Research Goal #1**: Combine path-planning algorithms with local statistics of tracks and the currently observed (partial) track to assess whether a person has the deliberate intention to go to a particular place.

'Somebody with binoculars close to position X'

Binoculars may be detected from imagery but will be error-prone due to low resolutions and bad visibility. Even when constraining to detections close to X, the detection results will be poor.

- **Challenge #2**: The search engine should take advantage of contextual information that narrows down the list of results. For instance, people with binoculars are observing a place and therefore they will loiter around. Loitering can be detected reliably and will improve search results.
- **Research Goal #2**: Expand the query from items and activities to related topics, such that additional clues can be gathered to improve the search result.

‘Group of adversaries somewhere’

Groups of people can be detected reliably. But, many groups will be around in the city, many neutral, and few adversaries. A problem is that adversaries cannot be identified by their appearance. Recent military lessons-learned hint that adversaries have to be identified by their behavior.

- **Challenge #3**: Separate the neutral entities from potential adversaries, using intelligence reports on normal and adversary behaviors.
- **Research Goal #3**: Incorporate prior knowledge from intelligence to provide contextual awareness on normal (e.g., a market every Wednesday, or, groups hanging out along the crossroad X) and adversary activities. Adversary activities may be location-bound, e.g., in neighborhood Y it is expected that some suspects have collaborators, and may involve discriminative behaviors, e.g., aggression, special signals, etc. These additional sources may be included to improve the search for adversaries.

The research goals above will be addressed in by GOOSE . They include the automated assessment of potential adversaries, which is believed to be feasible by gauging additional sources of information related to the query at hand. The main additional sources are intelligence reports, maps, statistics of normal behavior. Combinations between these sources are explored in a research project. The ultimate goal is to invent a military search engine that can answer the broad range of queries that can be asked about locations, vehicles, people and events.

5.4 Semantic search & retrieval

As a logical follow up on the current Internet, Tim Berners Lee coined the term ***“semantic web”*** to denote the web of data, also referred to as ***“linked data”***. Because a sensor’s output is in fact data, this vision is closer to what we require for GOOSE. Because sensors generate data both at a low semantic level (raw sensor data) and at a high semantic level (the interpreted, annotated sensor streams), it is important to distinguish the different ways the data is described. In the semantic web ontologies are used for data description to provide constrained and well defined vocabularies.

For the GOOSE information architecture many aspects of a standard linked data structure are applicable where data is stored distributed over the web, and can be queried via dedicated applications. However, the sensor data that is processed by the GOOSE system is much more dynamic than the relatively static data in today’s linked data applications. This requires a major extension for the current linked data search engines to be able to operate on ultra large collections of heterogenic and dynamic data sets generated by a diverse pool of sensors (ranging from simple temp. sensors to complex imaging sensors). The different sensors with their specific data characteristics require different classification algorithms with different semantic concepts to induce data sets for publication in the ***“web of linked data”***[20]. With the current Internet all available sensors can be accessible for search and retrieval if the data they generate are published according to the linked data architecture layers given in Figure 3.

Current developments

Some dedicated linked data search engines already are developed, such as Information workbench [22], SWSE [23] and Sig.ma [24]. Figure 4 illustrates all current data sets available which are mapped on the basis of the linked data priciples. In the linked data web published data sets can exist permanently. New data sets become “connected” and the mappings are updated. New mappings are added when new relations between exisiting vocabularies of data sets are created. For the GOOSE system this web will include dynamic data sets. For the linked data web scalability and the semantic gap are also big issues. There are useful technology solutions available from the Linked Data Community which are applicable to the GOOSE system [21].

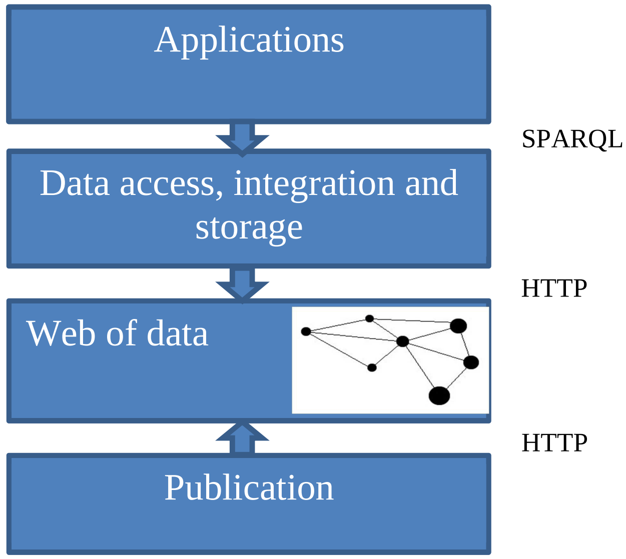

Figure 3 Linked-data information stack consisting of the following layers: the Publication layer refers to the “upload” of sensor data in RDF format, where wrappers can be used to translate native sensor data into RDF; the Web of data layer contains all RDF data sources distributed over webpages i.e. URL’s these data sources have different data schemas but statements represent relations between different schemas in mapping vocabularies; the Data access, integration and storage layer, processes the data sets in the web of data to enable search & retrieval by query using e.g. RDF triple store like Sesame and maintaining vocabulary mappings; the top layer Applications concerns the interfaces with the like a search app, and alert system or any other query based user app.

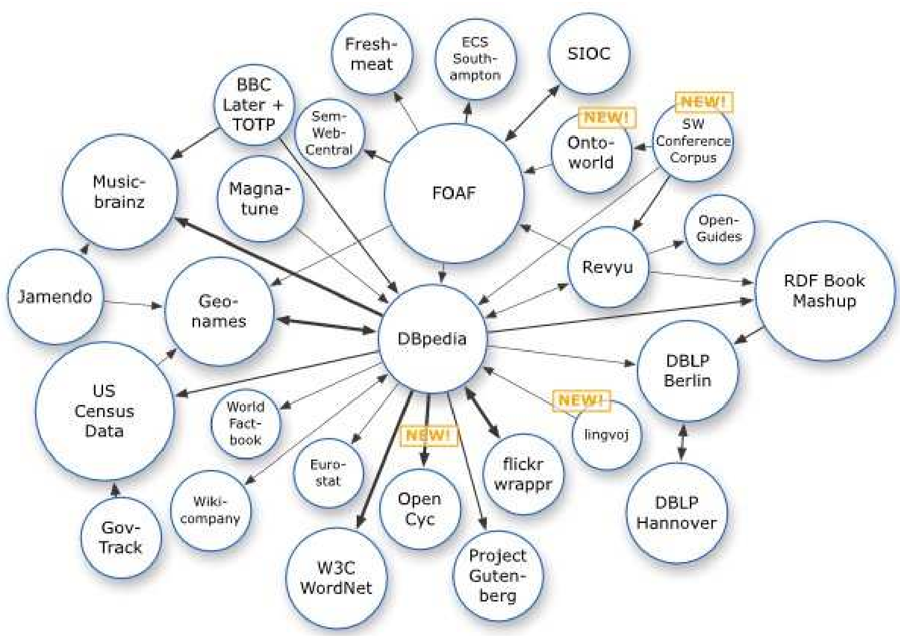

Figure 4 Current linked datasets on the web.

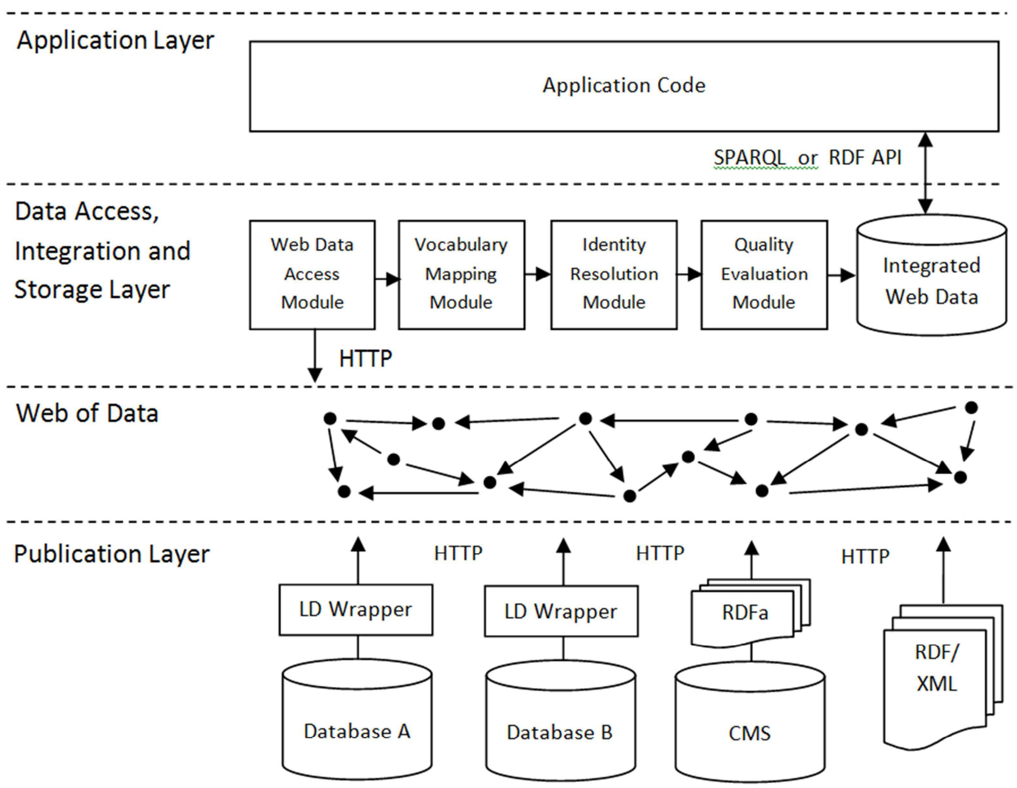

Figure 5 Standard architecture for Crawling pattern for linked data applications, taken from [20]

Using linked data technology for GOOSE

For each of the layers the research targets are described on the basis of a standard architecture for the crawling pattern for Linked Data applications. This standard architecture is illustrated in Figure 5.

1. Publication layer, in the standard architecture it is shown how the databases and CMS repositories are connected through linked data wrappers (e.g. transferring a relational DB table into RDF format output); the GOOSE system will add sensor data as well as classifiers using:
    a. Sensor description based on SensorML and iCore,
    b. Automated sensor discovery based on SensorML. And iCore,
    c. Quantities, Units, Dimensions and Types ontology [25],
    d. Presentation of sensors as a web-service (e.g. using WSDL [26]) resulting in a SOA (service-oriented architecture) construct to execute on-demand classifications like PTZ command to steer a camera FOV.
2. Web of data layer, concerns the standardization of semantics to create and build different ontologies (or vocabularies) which are mapped using mapping vocabularies. the objective is to construct . standardized top-level ontologies using commonly available concepts like the metric system, location representation, or generic concepts like objects and persons. This is as many top-level concepts domain specific standardization increases the interoperability between queries en sensor-output. For GOOSE it is essential that the Linked Data introduces different levels of semantics because the sensors intrinsically deliver low level of information.
3. Data access, integration and storage layer concerns the query federation pattern, the repository of components to build query pipes, directly working on the data sets without replicating. These should perform:
    a. Ontology reasoning, e.g. synonym, hypernym, spatial reasoning etc. A drawback is the high computational resource requirements.

   b. Integration, to provide the capability to query multiple different sensors, even including "social sensors" like twitter streams. The system can combine results from different types of sensors delivering a more accurate answer.
   c. Access control/ Billing, in case of proprietary or closed user group sensors. The linked data access control policies are important instruments to manage proper usage of sensor data sets.
   d. Trust, Authenticity, Quality, Manipulation, as seen already with the Open data initiatives to manage the reliability attributes of data sets correctness, originality, bad quality, or authenticity. Many ready available technologies from the media industry can be used.
4. Applications layer where the user interacts with the GOOSE system addresses query writing support to aid the user acquiring the specific information from the web of sensors, like auto-completion with frequently used search key words:
   a. Natural language search, possible with the LD architecture with natural language to SPARQL plugin [27],
   b. Data-concept search rather than text-based search like graphical geo area selection or by providing popular pre-defined classes based search. The implementation of user semantic search can evolve from text-based UI through concept-based to full semantic search,
   c. Workflow editing, in case the system is built on SOA principles with sensors and classification algorithms are incorporated as services. Such an implementation strategy requires more specific attributes in the user query to control which sensor classification algorithms must be used to optimize the results. There are solutions providing a GUI like Taverna [28], which aid the user to indicate which sensor data set must be used and which classification algorithms to operate on the data, and specific, where fusion of the results is expected.

It must be noted that it is not trivial to designate a particular place in the GOOSE system for the sensor information extraction and interpretation. It is possible and preferable to put it in the publication layer because it scales more easily by adding new URLs connecting new sensors and it also reduces the amount of data sets to be absorbed by the web. However, bringing the data sets to the data, access, integration, and storage layer makes sense from the application builder point of view as this layer gives more control on optimizing the sensor information extraction and interpretation pipes over the available resources. For true innovative GOOSE capabilities the second option is very promising.

5.5 Resource scaling

The scalable and distributed computing infrastructure needed to process the sensor information puts additional challenges on the design of the information architecture of the GOOSE system. In order to be able to search sensors, there must be some (real-time) pre-computed information available. To exploit powerful algorithms which can deliver the right information on user queries like ***"what cameras are detecting movement?"*** or ***"what cameras are 'seeing' human beings?"***, computational resources for (pre-)processing should be scalable, both with the number of sensors and with the number of required pre-processed features. Many of these requirements are already under development within the Big Data arena and are eligible for application in the GOOSE system.

As it is the strategy in the GOOSE project to automate meta data extraction at the sensor end the information architecture implements a highly scalable computational infrastructure, focused on "near-sensor" computational resources already available in today's smartphones. This is recognized as one of the most frontier game changing technology areas and vital to the success of information extraction capability. The machine intelligence –reasoning processes- is designed to support computational mechanisms to create and maintain the linked sensor data web and to handle the semantic transformations in the user domain to construct primitives for the search engine to answer the user queries.

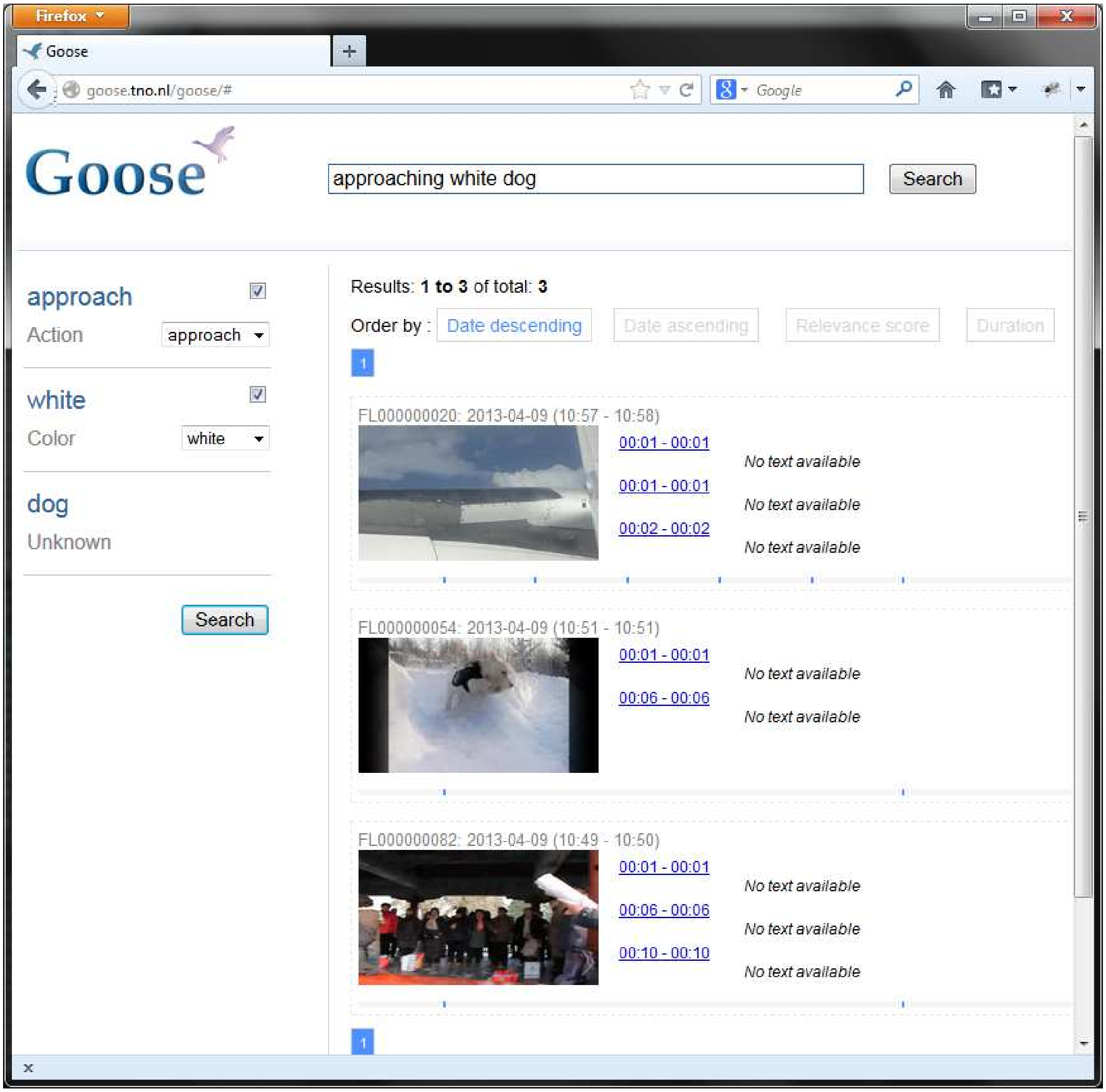

Figure 6 Preliminary GOOSE system in action, where the text query is parts into search primitives. These are mapped to initial feature extraction and second stage hypothesis verification modules.

6. SUMMARY

The research project to conceive and build a GOOSE system is driven by the urgent need to facilitate decision making on all levels of combat on the basis of new ways to exploit huge volumes of available information generated by sensors. Information extraction technologies exploit sensor data to aid users in the process of finding key information in text, audio, images, and video data sets are under development. An integral part of the GOOSE system architecture is the capability to automatically extract relevant pieces of intelligence and answer the query such that the user understands the information easily to make decision faster and more solid.

Two Big Issues have been identified and programmed into the research plan, which includes work in the field of sensor discovery, automated information extraction and linked data information architectures. The solution space is large and

the reference architecture should envisage all links between the required technology components to deliver the end-to-end capability.
It is acknowledged that the construction of GOOSE can start any time now following strategies similar to those in the Big Data arena featuring today's existing data processing and storage solutions while maintaining an eye on the horizon to ensure that the GOOSE framework can evolve on the waves of novel information technology continuously keeping up with the exponential growth of the number of sensors in the web.

REFERENCES

[1] Demistifying Big Data, TechAmerica foundation (2012).
[2] Big Data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute, 2011
[3] Air Force Research Laboratory, Information Directorate, AFRL-RI-RS-TM-2007-15, "Multimedia Information Extraction" 2007
[4] http://www.cosm.com, "Where the Internet of Things is being built"
[5] http://www.dropcam.com
[6] http://www.webcams.travel/community/, "Webcams worldwide"
[7] Dijk, J., Rieter-Barrell, Y., Rest, J. van, Bouma, H., "Intelligent Sensor Networks for surveillance", Journal of Police Studies: Technology-Led Policing 3(20), 109-125 (2011)
[8] Bouma, H., Baan, J., Landsmeer, S., Kruszynski, C., van Antwerpen, G., Dijk, J., " Real-time tracking and fast retrieval of persons in multiple surveillance cameras of a shopping mall,", Proc. SPIE 8756, (2013)
[9] Bouma, H., Hanckmann, P., Marck, J.W., Penning, L., Hollander, R., Hove, J.M., Broek, S.P., Schutte, K., Burghouts, G., "Automatic human action recognition in a scene from visual inputs," Proc. SPIE 8388, (2012)
[10] Hanckmann, P., Schutte, K., Burghouts, G.J., "Automated textual descriptions for a wide range of video events with 48 human actions," ECCV workshop on Video Event Categorization, Tagging and Retrieval, (2012).
[11] Burghouts, G.J., Schutte, K., Bouma, H., Hollander, R.J.M., "Selection of negative samples and two-stage combination of multiple features for action detection in thousands of videos," Machine Vision and Applications, (2013).
[12] Burghouts, G.J., Schutte, K., "Spatio-temporal layout of human actions for improved bag-of-words action detection," Pattern Recognition Letters, (2013).
[13] Bouma, H., Burghouts, G.J., de Penning, L., Hanckmann, L., ten Hove, J.M., Korzec, S., Kruithof, M., Landsmeer, S., van Leeuwen, C., van den Broek, S.P., Halma, A, den Hollander, R, Schutte, K., "Recognition and localization of relevant human behavior in videos," Proc. SPIE, vol. 8711, (2013).
[14] http://mes-project.tno.nl/, "Metadata Extraction Services: Understanding multimedia"
[15] John Schavemaker, Corne Versloot, Joost de Wit, and Wessel Kraaij. TNO instance search submission 2012. In Proceedings of TRECVID 2012, 2013. NIST, USA.
[16] Paul Over, George Awad, Martial Michel, Jonathan Fiscus, Greg Sanders, Barbara Shaw, Wessel Kraaij, Alan F. Smeaton, and Georges Quéot. TRECVID 2012 -- An Overview of the Goals, Tasks, Data, Evaluation Mechanisms and Metrics. In Proceedings of TRECVID 2012, 2013. NIST, USA.
[17] http://www.iot-icore.eu, "Empowering IoT through cognitive technologies"
[18] The Sensor Web Enablement, *http://www.ogcnetwork.net/SWE*
[19] SensorML: *http://www.ogcnetwork.net/SensorML*
[20] Tom Heath and Christian Bizer (2011) Linked Data: Evolving the Web into a Global Data Space (1st edition)
[21] Diggelen, J. van, Beun, R.J., Eijk, R.M. van, Werkhoven, P.J., Efficient Semantic Information Exchange for Ambient Intelligence. *Computer Journal* 53, 8, 1138-1151. (2010)
[22] Danh Le Phuoc, Manfred Hauswirth , "Linked Open Data in Sensor Data Mashups", Proc. 2nd Int. Workshop on Semantic Sensor Networks (SSN09), in conjunction with ISWC 2009, Vol-522, CEUR, 2009.
[23] SWSE, http://swse.deri.org/
[24] Sig.ma, http://sig.ma/
[25] Quantities, Units, Dimensions and Data Types in OWL and XML, http://www.QUDT.org
[26] Web Services Description Language (WSDL), http://www.w3.org/TR/wsdl
[27] Natural language to sparql, http://www.molto-project.eu/sites/default/files/presentation.pdf
[28] Taverna, http://www.taverna.org.uk/