NORTH ATLANTIC TREATY ORGANIZATION

SCIENCE AND TECHNOLOGY ORGANIZATION

AC/323(IST-144)TP/981

www.sto.nato.int

STO TECHNICAL REPORT **TR-IST-144**

Content Based Multi-Media Analytics (CBMA)

(Analyse multimédia basée sur le contenu (CBMA))

Summary Report of the NATO STO IST-144 Task Group.

Published October 2020

Distribution and Availability on Back Cover

NORTH ATLANTIC TREATY ORGANIZATION

SCIENCE AND TECHNOLOGY ORGANIZATION

AC/323(IST-144)TP/981

www.sto.nato.int

STO TECHNICAL REPORT **TR-IST-144**

Content Based Multi-Media Analytics (CBMA)

(Analyse multimédia basée sur le contenu (CBMA))

Summary Report of the NATO STO IST-144 Task Group.

The NATO Science and Technology Organization

Science & Technology (S&T) in the NATO context is defined as the selective and rigorous generation and application of state-of-the-art, validated knowledge for defence and security purposes. S&T activities embrace scientific research, technology development, transition, application and field-testing, experimentation and a range of related scientific activities that include systems engineering, operational research and analysis, synthesis, integration and validation of knowledge derived through the scientific method.

In NATO, S&T is addressed using different business models, namely a collaborative business model where NATO provides a forum where NATO Nations and partner Nations elect to use their national resources to define, conduct and promote cooperative research and information exchange, and secondly an in-house delivery business model where S&T activities are conducted in a NATO dedicated executive body, having its own personnel, capabilities and infrastructure.

The mission of the NATO Science & Technology Organization (STO) is to help position the Nations' and NATO's S&T investments as a strategic enabler of the knowledge and technology advantage for the defence and security posture of NATO Nations and partner Nations, by conducting and promoting S&T activities that augment and leverage the capabilities and programmes of the Alliance, of the NATO Nations and the partner Nations, in support of NATO's objectives, and contributing to NATO's ability to enable and influence security and defence related capability development and threat mitigation in NATO Nations and partner Nations, in accordance with NATO policies.

The total spectrum of this collaborative effort is addressed by six Technical Panels who manage a wide range of scientific research activities, a Group specialising in modelling and simulation, plus a Committee dedicated to supporting the information management needs of the organization.

- AVT Applied Vehicle Technology Panel
- HFM Human Factors and Medicine Panel
- IST Information Systems Technology Panel
- NMSG NATO Modelling and Simulation Group
- SAS System Analysis and Studies Panel
- SCI Systems Concepts and Integration Panel
- SET Sensors and Electronics Technology Panel

These Panels and Group are the power-house of the collaborative model and are made up of national representatives as well as recognised world-class scientists, engineers and information specialists. In addition to providing critical technical oversight, they also provide a communication link to military users and other NATO bodies.

The scientific and technological work is carried out by Technical Teams, created under one or more of these eight bodies, for specific research activities which have a defined duration. These research activities can take a variety of forms, including Task Groups, Workshops, Symposia, Specialists' Meetings, Lecture Series and Technical Courses.

The content of this publication has been reproduced directly from material supplied by STO or the authors.

Published October 2020

ISBN 978-92-837-2301-1

Table of Contents

List of Figures

List of Tables

IST-144 Membership List

CO-CHAIRS

Dr. Elizabeth K. BOWMAN
Army Research Laboratory (Ret.)
UNITED STATES
Email: elizabeth.k.bowman.civ@mail.mil

Prof. Bhopinder MADAHAR
Dstl
UNITED KINGDOM
Email: bkmadahar@dstl.gov.uk

MEMBERS

Dr. Gertjan BURGHOUTS
TNO Research Institute
The NETHERLANDS
Email: gertjan.burghouts@tno.nl

Dr. Michael HARRIES
Dstl
UNITED KINGDOM
Email: mharries@dstl.gov.uk

Dr. Peter HARVEY
Dstl
UNITED KINGDOM
Email: PRHARVEY@mail.dstl.gov.uk

Dr. Lasse OVERLIER
Norwegian Defence Research Establishment (FFI)
NORWAY
Email: Lasse.Overlier@ffi.no

Dr. Raghuveer RAO
ARL
UNITED STATES
Email: raghuveer.rao@us.army.mil

ADDITIONAL CONTRIBUTORS

Mr. Alan KNOX
Dstl
UNITED KINGDOM
Email: awknox@dstl.gov.uk

Mr. Tod HAGAN
Securboration Inc.
UNITED STATES
Email: thagan@securboration.com

PANEL/GROUP MENTOR

Prof. Bhopinder MADAHAR
Dstl
UNITED KINGDOM
Email: bkmadahar@dstl.gov.uk

Content Based Multi-Media Analytics (CBMA)

(STO-TR-IST-144)

Executive Summary

The NATO Information Systems Technology (IST) Panel Research Task Group (RTG), 144, on Content Based Multi-media Analytics (CBMA) undertook Research and Development (R&D) over three years starting early 2015. It was the outcome of an exploratory team (IST-ET-86), completed in 2014, which reviewed the State-Of-The-Art (SOTA) in the fields of "Complex Event Processing for Content-Based Text, Image and Video Retrieval". The R&D by a team from NLD, NOR, UK and USA, has advanced capabilities in the following areas and has shown the 'art-of-the-possible' through concept demonstrations at high profile NATO events:

- Intelligent capture and indexing of motion imagery;
- Expansion of Machine Learning (ML) / Deep Learning (DL) approach to semantic video analytics;
- Cross-cueing from text analytics to drive/exploit video and imagery indexing and retrieval; and
- Explored potential architectures and frameworks that could optimise the implementation of multi-media analytics in distributed coalition environments.

The specific analytics capabilities developed and available at component level are:

Text Analytics	Image/Video Analytics	Semantic Correlations
Text analytics of unstructured data Social-media Understanding and Reasoning Framework (SURF) Inferring entities of interest and relationships	Live social media video analytics Fast indexing on-line videos Searching and identifying objects of interest	Semantic analysis of images/videos Cross-cueing between social media and images/videos

These advances are needed to address the defence challenges raised by NATO alliance and its member nations being increasingly threatened by the global spread of terrorism, humanitarian crises/disaster response, and public health emergencies. These threats are informed and/or influenced by the unprecedented rise of information sharing technologies and practices, even in the most underprivileged nations. In this new Information Environment (IE), agile data algorithms, ML software, and threat alert mechanisms must be developed to automatically create alerts and drive quick response.

Expected impact of the work, for NATO it is in heterogeneous data driven decision making and intelligence analysis applications. This is to:

- Boost agility of military operations both in the physical and IE through deep understanding of adversary perspectives, intent, and threats. This integration of analytics is especially critical in the current threat environment where adversaries operate with information-based tactics designed to achieve strategic goals with non-kinetic methods; and
- Significantly increase in the Commander's and Staff's ability to understand adversary perspectives, intent and threats.

Yet to be most effective, these advanced technologies must be integrated into an open distributed system of systems for the coalition providing the services needed, be balanced with awareness of the underlying context to accurately interpret machine-processed indicators and warnings and recommendations. Furthermore, human involvement will always remain critical in the coalition decision process so a human-centric approach is needed. This includes socio-technical aspects such as Ethics and Privacy. Hence the key recommendations for further NATO led research study, working between panels (e.g., IST and HFM) are to:

- Investigate designs for a human-centric distributed system of systems, leveraging commodity cloud and ML developments, and emerging knowledge representation approaches, to provide the multi-modal analytics services needed by NATO applications. This includes experimentation to explore the human analysts' ability to rapidly and accurately incorporate multi-modal information with a variety of visual displays;
- Determine the state-of-the-art in robustness and accountability for ML systems. Especially DL systems with complex and large models which are virtually impossible to manage by humans;
- Explore potential approaches for the coalition to address emerging socio-technical issues, such as Ethics and Privacy, arising with human-machine systems.

Analyse multimédia basée sur le contenu (CBMA)

(STO-TR-IST-144)

Synthèse

Le groupe de recherche de la commission OTAN sur la technologie des systèmes d'information (IST) RTG-144, intitulé « Analyse multimédia basée sur le contenu (CBMA) », a entamé début 2015 des travaux de recherche et développement (R&D) s'étalant sur trois ans. Ce RTG résultait d'une équipe exploratoire (IST-ET-86) arrivée à son terme en 2014, qui avait examiné l'état de la technique (SOTA) dans le domaine du « traitement d'événement complexe pour la récupération de texte, d'image et de vidéo à partir du contenu ». L'équipe, dont les membres venaient des États-Unis, du Royaume-Uni, de Norvège et des Pays-Bas, a travaillé dans les domaines suivants et montré ce qui était « techniquement possible » par des démonstrations de concept lors d'événements OTAN de premier plan :

- Capture et indexation intelligentes d'images de mouvement ;
- Élargissement de la démarche d'apprentissage automatique (ML, machine learning) ou d'apprentissage profond (DL, deep learning) à l'analyse sémantique de la vidéo ;
- Repérage croisé à partir de l'analyse de texte pour entraîner/exploiter l'indexation et la récupération de vidéos et d'images ; et
- Étude d'architectures et de cadres potentiels pouvant optimiser la mise en œuvre de l'analyse multimédia dans les environnements de coalition répartis.

Les capacités d'analyse spécifiques mises au point et disponibles au niveau du composant sont :

Analyse de texte	Analyse d'image/de vidéo	Corrélations sémantiques
• Analyse de texte de données non structurées • Cadre de compréhension et de raisonnement des médias sociaux (SURF, Social-media Understanding and Reasoning Framework) • Déduction d'entités d'intérêt et de liens	• Analyse de vidéo en direct dans les médias sociaux • Indexation rapide de vidéos en ligne • Recherche et identification d'objets d'intérêt	• Analyse sémantique d'images/de vidéos • Repérage croisé entre les médias sociaux et des images/vidéos

Ces progrès sont nécessaires pour répondre aux menaces en matière de défense auxquels l'Alliance de l'OTAN et ses pays membres sont de plus en plus confrontés à cause de la diffusion mondiale du terrorisme, des crises/catastrophes humanitaires et des urgences de santé publique. Ces menaces sont informées et/ou influencées par le développement sans précédent des technologies et des pratiques de partage des informations, même dans les pays les moins privilégiés. Dans ce nouvel environnement d'information (IE), il est impératif de mettre au point des algorithmes de données agiles, des logiciels de ML et des mécanismes d'alerte, afin de lancer automatiquement des alertes et d'enclencher une réaction rapide.

Ces travaux devraient trouver, au sein de l'OTAN, des applications en matière d'analyse du renseignement et de prise de décision à partir de données hétérogènes. Cela servira à :

- Stimuler l'agilité des opérations militaires, dans l'environnement physique et l'IE, par une compréhension fine du point de vue, de l'intention et des menaces des adversaires. Cette intégration de l'analyse est particulièrement capitale dans l'environnement actuel des menaces, où les adversaires appliquent des tactiques basées sur les informations, conçues pour atteindre des objectifs stratégiques par des méthodes non cinétiques ; et
- Augmenter sensiblement la capacité du commandant et de l'état-major à comprendre le point de vue, l'intention et les menaces des adversaires.

Cependant, pour être les plus efficaces possible, ces technologies perfectionnées doivent être intégrées dans un système de systèmes réparti et ouvert fournissant les services nécessaires à la coalition et être contrebalancées par la connaissance du contexte sous-jacent qui permet d'interpréter correctement les indicateurs, avertissements et recommandations provenant de machines. Par ailleurs, l'implication humaine restera toujours essentielle au processus décisionnel de la coalition. Une démarche axée sur l'humain est donc nécessaire. Cela inclut des aspects socio-techniques, tels que l'éthique et la confidentialité. Par conséquent, les recommandations clés en vue d'autres recherches menées par l'OTAN, dans le cadre de travaux entre commissions (par exemple, IST et HFM), sont les suivantes :

- Étudier les modèles applicables à un système de systèmes réparti et centré sur l'humain, en utilisant l'évolution du cloud grand public et du ML et les démarches émergentes de représentation des connaissances, pour fournir les services d'analyse multimodaux nécessaires aux applications de l'OTAN. Cela inclut l'expérimentation portant sur la capacité des analystes humains à incorporer avec rapidité et précision des informations multimodales dans divers affichages visuels ;
- Déterminer l'état de la technique en matière de robustesse et de responsabilité des systèmes de ML, en particulier les systèmes de DL avec des modèles complexes et vastes, qui sont virtuellement impossibles à gérer par les humains ;
- Explorer les démarches potentielles de la coalition visant à traiter les questions socio-techniques, telles que l'éthique et la confidentialité, qui découlent des systèmes humain-machine.

CONTENT BASED MULTI-MEDIA ANALYTICS (CBMA)

1.0 INTRODUCTION AND AIMS

1.1 Introduction

Gaining and maintaining Information Superiority, i.e., an Information Advantage, over potential adversaries, now and in the future, is a strategic objective for the NATO alliance and its member nations[1]. This is in recognition of the challenges, opportunities and threats posed by the unprecedented and increasing global rise, at exponential rates, of data and information permeating every sector of the global hyper-connected society, in developed or developing nations. Such a dynamic, rich and global Information Environment (IE) presents new opportunities to our adversaries and a persistent challenge to the coalition in addressing the increasing threats posed, for example, by the global spread of terrorism, natural or man-made humanitarian crises/disasters and public health emergencies. These threats are informed and/or influenced by the unprecedented rise of information sharing technologies and practices, even in the most underprivileged nations. In this new IE, agile data algorithms, machine learning software, and threat alert mechanisms must be developed to automatically create alerts and drive quick response. Yet these advanced technologies must be balanced with awareness of the underlying context to accurately interpret machine-processed indicators and warnings and recommendations; human involvement will always remain critical in the coalition decision process.

This report provides a summary of the research developments of the NATO information Systems Technology (IST) panel Research Task Group (RTG), IST-144-RTG on Content Based Multi-media Analytics (CBA) carried out by a team from USA, UK, NOR and NLD. It covers the SOTA developments in analytics of heterogeneous media sources with a focus on text, video and images. This is within the overall context of providing better automated support to military analysts and decision makers in extracting salient and meaningful information from multi-media sources pertinent to their IE. Support includes bringing together information retrieval strategies from heterogeneous media sources (text, video and images) and human assessment. As a result multiple heterogeneous data sources can be exploited by content based information retrieval and multi-media analytics to deliver timely and accurate synopses of data with information that can be combined with human intuition and understanding to develop a comprehensive 'view' of the problem/solution space. Such interoperable tools that cross-cue knowledge obtained from one method to generate taskings in another are needed by NATO Coalition military leaders, commanders, and intelligence analysts to accelerate situational awareness and decision making and deal with the complexity of the defence IE.

Following a summary of the aims, an outline of the research approach is provided in Section 2. This includes encapsulation of a generic CBA process for defence analysts in cartoon image form – similar to the well-known defence OODA (Observe, Orient, Decide and Act) loop. This image not only helped communicating the essence of the research to others, experts and non-experts, in a simple form but also helped the research developments of the team, from the different nations, to maintain a common vision and focus. A detailed description of the core technical components is provided in Section 3. This is followed in Section 4 by description of their combined application in a concept demonstrator addressing a fictitious scenario, but nevertheless realistic, representative of defence challenge facing the alliance. Section 5 presents

[1] "…our ability to respond faster through cleverer decision-making which is enabled by the flow of information, is actually frankly as important if not even more important than whether our tanks out-range an anti-tank missile." General Sir Gordon Messenger, UK MOD, Vice Chief Defence Staff [1].

"An information oriented navy, ready to use information as a weapon and tool." UK MOD , The Future Navy Vision: Royal Navy today, tomorrow and towards 2025 [2].

Information declared the 7th Joint function and recognised as an instrument of national power by the US DOD, Joint Doctrine Note 1-18, 25 April 18 [3].

several operational evaluations conducted in different military venues to establish the value and maturity of products to Defence customers. The report is concluded with the findings and recommendations in Section 5 for further work that is needed to address limitations, including technical and system level gaps.

1.2 Aims

The project was initiated at the end of 2015 but started in earnest in April 2016 and completed in April 2019. It was preceded by, and built upon, the outcomes of an exploratory team (NATO IST-ET-86) that involved two additional nations, CAN and TUR, than the RTG. The exploratory team reviewed the State-Of-The-Art (SOTA) in the fields of "Complex Event Processing for Content-Based Text, Image and Video Retrieval" and issued a comprehensive report [4]. It concluded that this was an emerging field with rapid growth in the civil sector but that technologies remained at low levels of maturity for defence applications with specific research gaps. Key recommendation was to establish a RTG to "…enhance 'real-time' analytics of heterogeneous multi-media streams (image, video, text, speech, etc.) enabled by enhancements in the contextual understanding of complex events through advances in computational/human processing techniques". This through focus on the following technical aspects, and research gaps therein:

- Intelligent Capture and indexing of motion imagery;
- Expansion of Machine Learning / Deep Learning approach to semantic video analytics;
- Cross-cueing from text analytics to drive/exploit video and imagery indexing and retrieval; and
- Explore potential architectures and frameworks that could optimise the implementation of multi-media analytics in distributed coalition environments.

Hence the aims of the task group were to address the key recommendation with a clear focus on the above technical areas and with the following expectations.

Expected technical outcomes:

- Multi-media contextual analytics techniques that can harvest forensic social and digital media; and
- Correlate video and text analytics for cross-cueing between data types.

Expected impact:

- Boost agility of military operations both in the physical and IE through deep understanding of adversary perspectives, intent, and threats. This integration of analytics is especially critical in the current threat environment where adversaries operate with information-based tactics designed to achieve strategic goals with non-kinetic methods; and
- Significant increase in the Commander's and Staff's ability to understand adversary perspectives, intent and threats.

2.0 RESEARCH DEVELOPMENTS

2.1 Overview of Approach

The approach used is similar to other NATO RTG activities in that it links to, and leverages, current developments in the participating nations relevant to the objectives of the RTG. Research teams are brought together through regular meetings (video/phone conferences and on-site meetings at participating laboratories) to progress the agenda. In this RTG, to focus the application of the research, we deliberately included towards the end articulation of concepts through realising aspects of what NATO refers to as Concept Demonstrators of Technologies (CDTs).

Each of these instruments was exploited in full, and in addition, the RTG organised a NATO Research Specialist Meeting (RSM) which was held on 6 – 8 September 2017 at the University of Middlesex in the UK (see Annex A, programme information). This was to review the RTG work to date, and to solicit additional inputs and advice from the experts in attendance, including the key notes, on forward direction and recommendations. There were 25 attendees, including the research team, and two key notes[2] plus a visit to our host's, Prof. William Wong, lab at the University. This was to see their research on the EU VALCRI (Visual Analytics for Sense-Making in Criminal Intelligence Analysis) project[3] focussing on visualisation of complex multi-media data for decision making, which is a human factors research complement to the RTG.

RSM [5] indicated that the research was on the right track and offered two recommendations. One is to continue the shared video indexing experiment being carried out by the project in support of concept demonstrators. The second is to propose a follow-on human subjects experiment to explore the human analysts' ability to rapidly and accurately incorporate multi-modal information with a variety of visual displays. This overlaps with the NATO Human Factors and Medicine Panel regarding human factors and has not been progressed due to on-going commitments on current panel research activities.

With the RSM report and that from the ET, mentioned earlier [4], the reader should be able to glean a comprehensive view of the leading technical developments related to analytics of multi-modal streams. Here, in this report and the following sections, we provide details of the technical components that underpin specific technical capabilities considered to be essential to the goal of the RTG. These capabilities and components are summarised in Table 1.

Table 1: Technical Capabilities and Components.

Text Analytics	Image/Video Analytics	Semantic Correlations
• Text analytics of unstructured data. • Social media Understanding and Reasoning Framework (SURF). • Inferring entities of interest and relationships.	• Live social media video analytics. • Fast indexing on-line videos. • Searching and identifying objects of interest.	• Semantic analysis of images/videos. • Cross-cueing between social media and images/videos.

2.1.1 Generalist Specialist Analyst Work

In order to bring these technical capabilities to life, a consideration of the work of an analyst is needed. This could range from an analyst supporting decision making in general or specialists supporting particular areas such as intelligence. There are in general well established processes that a search on Wikipedia or Google Scholar will reveal. One at the very high level is Boyd's Observe – Orientate – Decide – Act (OODA) loop that is much cited as high level processes for decision making in operations. Another, at a lower but much more detailed level, and based on research findings of Cognitive Task Analysis (CTA), is the Intelligence process cycle by Pirolli and Card for intelligence analysts [6]. This is shown in Annex B and its bottom-up stages, such as the foraging loop, align with the capabilities and components being developed by the RTG.

[2] On forgery detection in images by Dr. Katrin Franke Professor of Computer Science, Centre for Cyber and Information Security, Norwegian University of Science and Technology.
On use of Motif detection by Mr. Tod Hagan, Securboration Inc, Melbourne, FL, USA.

[3] Details available at http://valcri.org/.

In their paper, Pirolli and Card define the bottom-up stages as:

- Search and filter (e.g., collecting relevant data/information into some store (the “shoebox” in the diagram) for further processing);
- Read and extract (e.g., extract nuggets of evidence that may be used to draw inferences, or support or disconfirm theory triggering new hypothesis/searches);
- Schematize (e.g., information representation and visualization);
- Build case (e.g., marshalling of existing/new evidence to support or disconfirm hypotheses); and
- Tell story (e.g., Publication/presentation of evidential case).

The RTG considered how best to exploit these processes but represent them in a much simpler and meaningful way for the research as described next.

2.1.2 Abstracted Intelligence Analyst Cycle

One of the students supporting a research team member used their artistic and creative skills to draw a simple picture of the intelligence cycle in cartoon form. With further suggestions and refinements from other team members, the project converged on the following schematic as an excellent abstract representation of the intelligence cycle [7] (Figure 1).

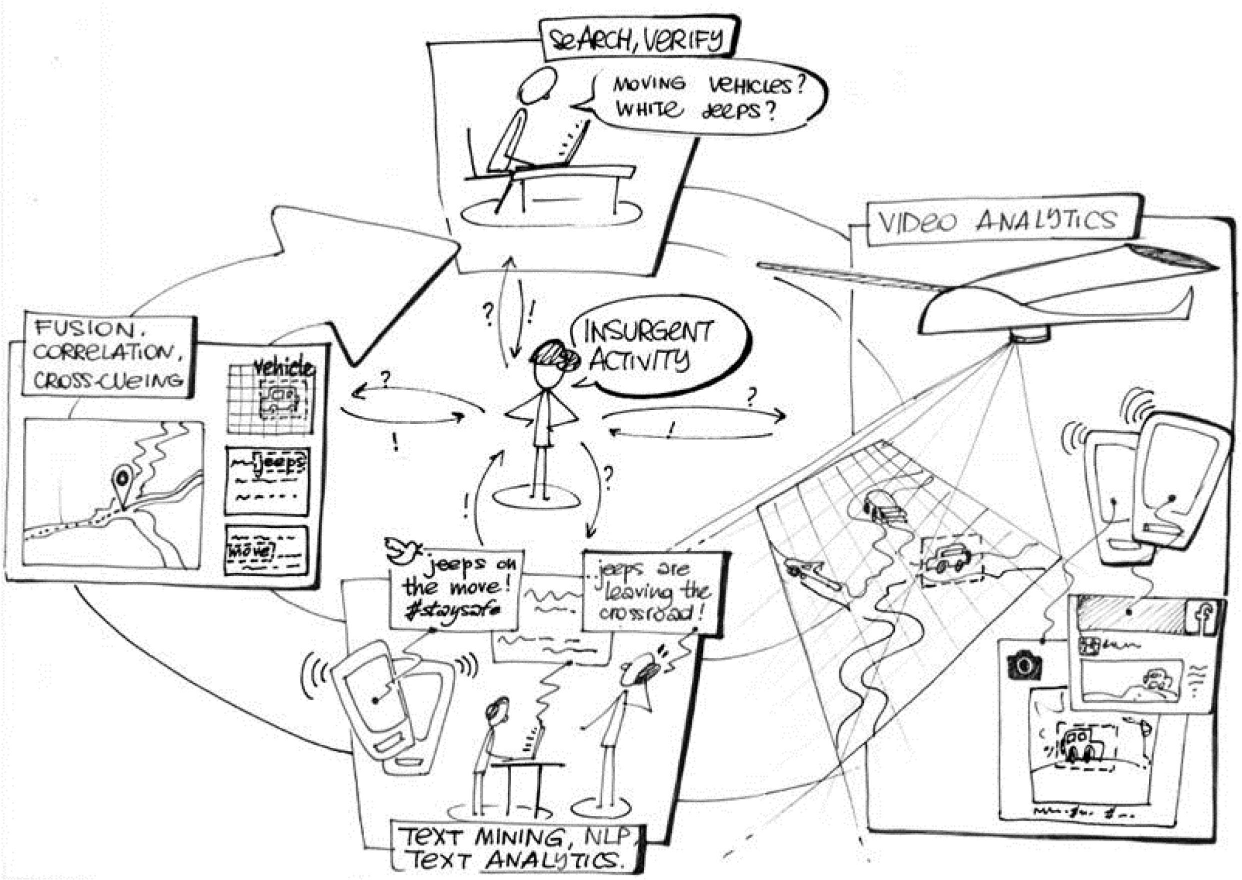


Figure 1: Intelligence Cycle for Multi-Modal Data.

This captures the essence of the type of services the core technical components developed by the RTG need to be engineered and integrated to provide. Central point of this iterative loop is the task that an analyst is trying to fulfil – marked as ‘insurgent activity’ in the picture. That is the identification of such an activity in the operational space and extracting the meaningful information from the analytics to build the supporting evidential case. Early foraging stages, text analytics, have indicated potential adversaries and modes of operation. Further stages and text/video analytics are extracting facts in answer to queries (or hypothesis) and iterative refinement whereby relevant elements are linked together, relationships inferred, and added to the evidential case. End product is a final set of correlated and fused information that pieces together and presents a compelling picture/story to trigger a decision and action.

It is important to note that the processes are not necessarily cyclical but that the central query or hypothesis (the goal) triggers actions from, and between, one or more, of the asynchronous parallel analytics processes. This is important as the multi-modal analytics system has to be able to provide services in parallel to address multiple queries. Overall utility will be much more evident when a representative scenario for the applications of the developments is reported in later sections.

The researchers found this picture to be a simple but powerful image to communicate the research to other and to focus the developments from the different teams towards a common goal. Hence it is referred to extensively in this report and has been used extensively by the project in presentations, demonstrations and publications in peer reviewed journals. These also provide additional technical details on related developments by the research team and are listed in the reference section ([8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24]) for readers to obtain a more comprehensive overview of the research.

3.0 TECHNICAL COMPONENTS

Details of the technical components supporting the technical capabilities listed in Table 1 are provided in the following sub-sections. They are listed under the primary heading from the table.

3.1 Text Analytics of Unstructured Data

Data from documents or social media and its analysis, such as through Natural Language Processing (NLP) and text analytics, can provide a better understanding of the content, behaviour, attitudes, intentions and relationships of the entities and actors involved. A number of generic and bespoke analytics tools are available to an analyst and the RTG selected the following as best fit for purpose.

3.1.1 Text Analytics

Dstl has developed a flexible and robust framework for text analytics research called BALEEN which has been provided as open source software on GitHub [25]. This framework enables users to extract entities and information from unstructured and semi-structured text from different sources (e.g., documents, social media). It is constructed with open source libraries, and bespoke Dstl developed code, and is an open source (Apache 2.0 license) Java framework built on top of Apache Unstructured Information Management Applications UIMA [26]. The latter enables applications to be decomposed into components (e.g., “language identification”; “language specific segmentation”; “sentence boundary detection”; “entity detection” (person/place names, etc.)) with interfaces for each defined by the framework that also provides self-describing metadata via XML descriptor files. The framework manages these components and the data flow between them.

There is common data structure called the CAS (Common Analysis System), and the framework initiates this, then passes it in turn to a sequence of “annotators”. The annotators can read from, and write to, the

CAS. Each annotator can see (and modify/delete) any annotations from upstream annotators. Annotations are defined by an extensible type system, and individual annotators depend only on the annotation types they consume and produce (and not each other). The first and last annotators in the pipeline have special status, and are called the "collection reader" and "consumer" respectively. The collection reader is responsible for getting text into the CAS, and the consumer is responsible for doing whatever is required with the final set of annotations. There are Java interfaces defined for collection readers, annotators and consumers – so it possible to customise and extend in whatever way is required. What any particular running instance of BALEEN does is determined by the annotators in its pipeline – driven by a configuration file. BALEEN comes with a rich set of collection readers [25] and its overall architecture is shown in Figure 2.

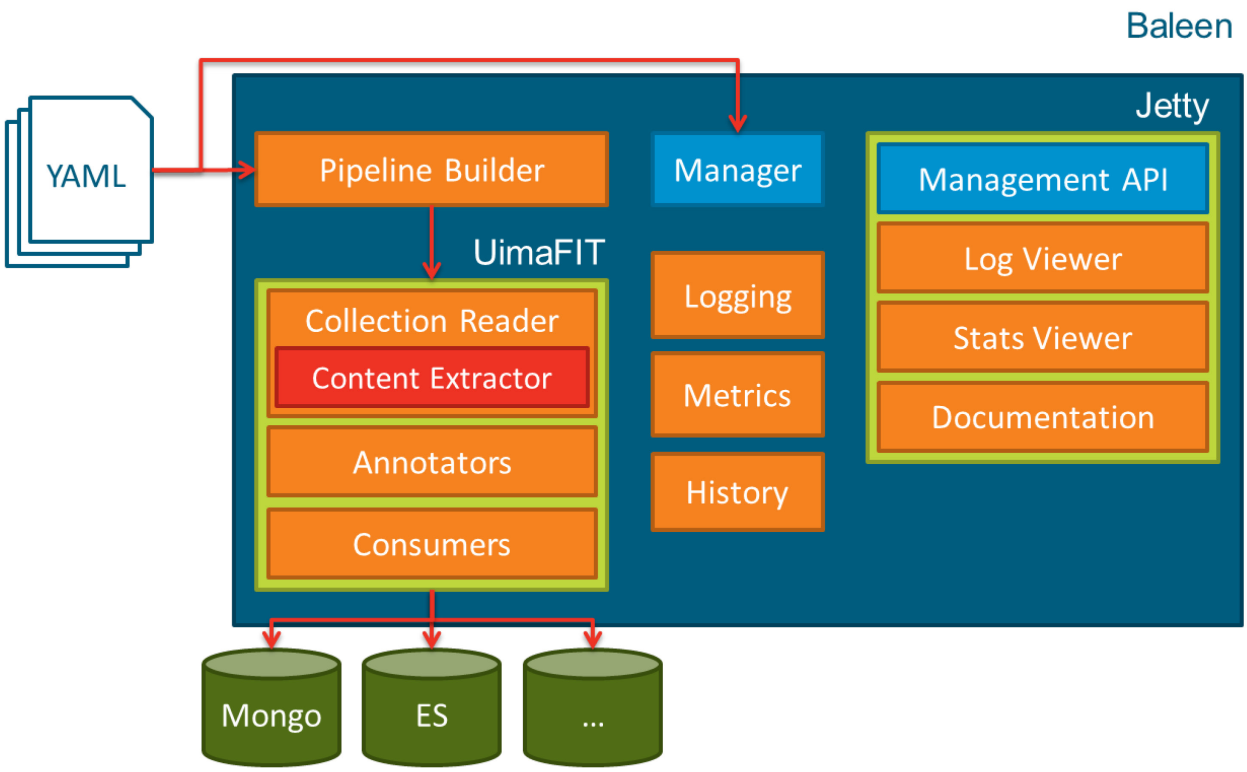


Figure 2: Architecture of BALEEN.

3.1.2 Social Media Understanding and Reasoning Framework (SURF)

The Social Understanding and Reasoning Framework (SURF) [27] is a text analytics platform that determines threat levels of individuals based on their social media interactions using machine learning and bioinformatics-inspired algorithms to find and accurately classify group membership. The tool currently operates with Twitter messages (Tweets) and is language-agnostic because classification is based on network patterns. SURF has been successfully tested on Arabic, English, French, and Spanish languages. The classification algorithm is based on a proven formula used to classify biological entities from their protein interaction network. The SURF adaptation to social media is >85% accuracy using cross-folded validation with ten random K-folds. Group models are built from examples of true positive and true negative messages; currently four models exist. These are businessmen, hackers, Islamic State of Iraq and Syria (ISIS) 1 and ISIS 2 (e.g., earlier and later ISIS group adaptations). SURF provides two categories of group membership: nodes and ego networks. The first identifies individuals active in the Twitter network under study and the second considers friends and followers of individuals. The second category is important to consider because of the combined influence friends and followers can have on individuals.

3.1.3 Inferring Entities of Interest and Relationships

The SURF capability to infer entities of interest and relationships within a social media network was demonstrated in May 2019 at the Army Enterprise Challenge 2019 Exercise at Ft. Huachuca, AZ. Over 15 Open Source Intelligence (OSINT) analysts exploited a social media dataset in the realm of Multi-Domain Operations (MDO). They used SURF in a MDO scenario for preparatory intelligence activities to characterise adversary actions and formulate a credible US information narrative. Key results from this interaction with analysts included:

- The ability to classify data elements (social media users, signals, events) into membership groups using pattern association machine learning algorithms provides a language-agnostic analytic that is flexible and robust.
- The ability of analysts to modify current membership groups and create new classes makes SURF adaptive to dynamic battlefield operations.
- SURF visualizations would be enhanced with geospatial overlays correlating to areas of operation for relevant echelons of command.

The impacts of the results from this exercise verify that the SURF tool is a valuable component in a larger toolchain designed to classify levels of threat from adversaries and validate that the key components of the SURF tool match Intelligence requirements for Multi-Domain Operations. The successful demonstration and characterization of SURF during EC19 allows for smoother transition opportunities within the Army system, especially with regard to future capability drops for decision making tools. SURF is currently used by elements of the Intelligence Community (IC) and the EC19 demonstration and characterization results will strengthen those relationships.

3.2 Image/Video Analytics

The sheer number of social media postings for almost any area, in a given period of time, makes it impossible for an analyst to view and mark the content relevant or not. This is especially true for postings containing images and videos. Are the posted images real? Are the acclaimed times of recording correct? Have the images or videos been posted anywhere before? Questions such as these must be considered. We have built a tool [28] to shorten the list with flexible triggers for deeper analysis. This tool can be used by an analyst to specify the relevant tags, users and keywords, and the monitoring of the active social media platform(s) will then result in a filtered list of the relevant postings and links for options and possible actions for further analysis as shown in Figure 3.

The tool works with several aspects of Tweets, to include text, image/video clips, geolocation markers, user names, relevant links, and locations. The capability of an early version, shown in Figure 3, is to identify objects of interest in videos, keywords from text in images, make viewing of new postings of interest simpler, and identify whether a video clip has been observed on-line before, and if so, by whom, when, and where. The user is presented with the option to manually specify the relevant tags, users and keywords. The monitoring of the active social media platform(s) will then result in a filtered list of the relevant postings and links for options and possible actions for further analysis as shown in the left part of Figure 3. This list will enable the analyst to rapidly address the detailed content of the posting and look for more relevant material through the user behind the posting. Using one of the items in Figure 3 as an example, the analyst can on the left choose to more deeply examine the marked video for accuracy as shown to the right in the figure. This in-depth analysis will happen from the locally stored copy of the video identified in the posting, and will not go on-line again to look for postings that may have been changed or deleted.

The system is configured to follow a set of tags, users and/or keywords. The tool enables automatic pull down of the complete content of related messages, together with any content linked to in the messages, like cloud-based images and videos. Due to limitations in access to old data information may be fetched from last two weeks. But if earlier historical data is needed this must likely be purchased from commercial services. Any image or video linked/referenced in a social media posting on-line which matches our set of criteria will now be downloaded and automatically analysed in-depth with regards to:

- Does the posting have identifying marks, like geolocation markers, specific content, names, groups, etc.?
- Does the included or referenced image/video contain marked objects like tanks, weapons, drone strikes, etc.?
- Has the image/video clip been observed on-line before? When, where, by whom?
- Does the image/video contain text like posters, airplane identifiers, names, and/or textual propaganda?

For future inclusion in tool we have also researched:

- What is the textual content and its context for this posting?
- Does the image/video contain faces, and do the faces exist in the database?
- Does the video contain any events marked for more interest, like object exchange between humans, digging a hole in the ground, throwing an object, etc.?

For each of the in-depth image and video analyses there are special tools and systems working in the background of our architecture. Identifying marks are pulled from the text and metadata with each social media entry. These are fetched with the original stream of postings and used to pull down the necessary extra data related to this posting, like images, videos, geolocation tags, etc.

The rest of the technologies used in the tool and listed above will be explained in the following sections.

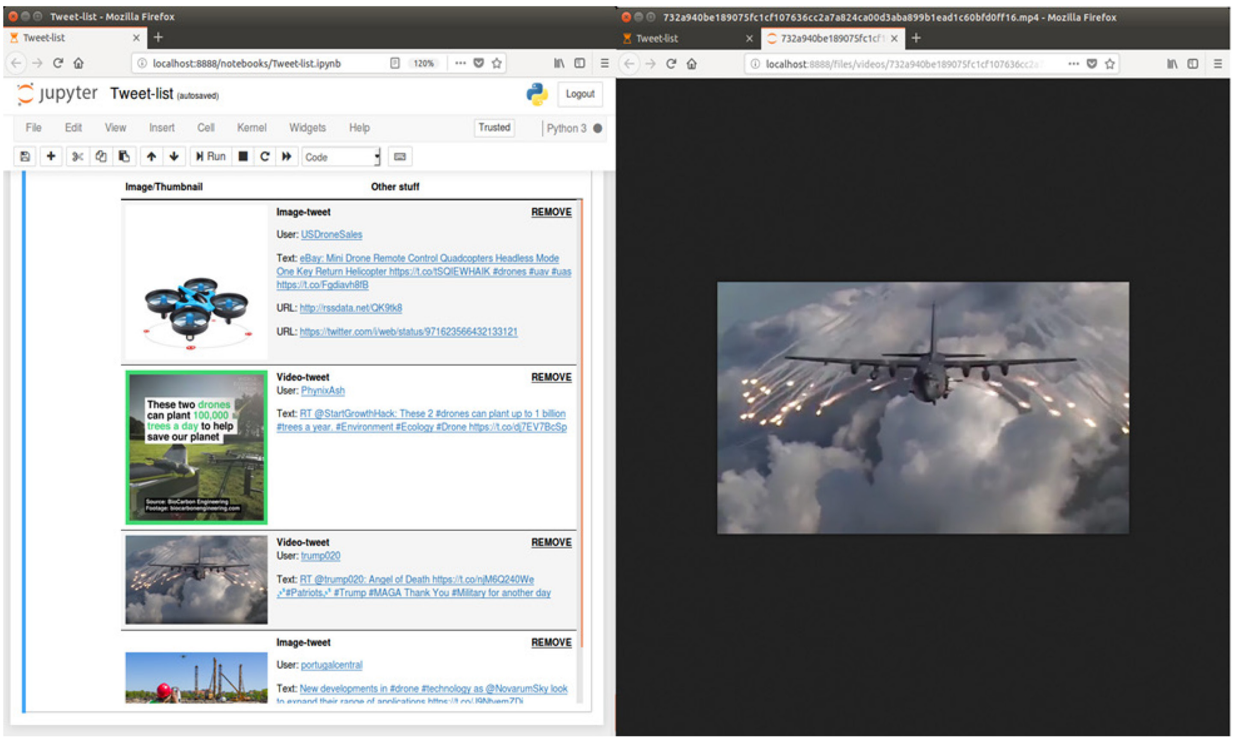


Figure 3: Stream of Relevant and Filtered Images and Videos Shown in List to the Left; on the Right Manual Inspection of Item in List from Local Repository.

3.2.1 Objects Marked for Special Interest

Using machine learning for detecting objects in images is a growing area of research and many machine learning models have been created for this including deep learning techniques [29], [30]. These work out-of-the-box for the predefined classes of objects, but if you need to detect new classes you have to either train your own model or retrain the existing model with new and/or additional data. We want to recognise objects not in the major existing detection models and have used TensorFlow to retrain an existing model with new data from an idea and description made on-line (see Refs. [31], [32]).

In order to demonstrate the object recognition of new objects we downloaded a set of 100+ images of tanks and 100+ set of drones and used TensorFlow to retrain the final layer of an existing deep learning model for object detection. The process of retraining for object detection involves manually tagging each image with items of interest with a corresponding bounding box for the object as shown in Figure 4.

This enables the tool to return matching objects and include a bounding box for the location within the image the object has been detected.

The retraining of the system was performed on the SSD_mobilenet_v1 [32] model which has more focus on speed than accuracy. The main challenge with detection within each image in a video is also in this case the amount of processing power for medium sized videos. In our case a good laptop with a dedicated graphics processor was able to run the detection algorithm at 10 – 20 frames per second. Simple optimizations can be done by looking for significant changes in scenes and/or just using every 10 frames.

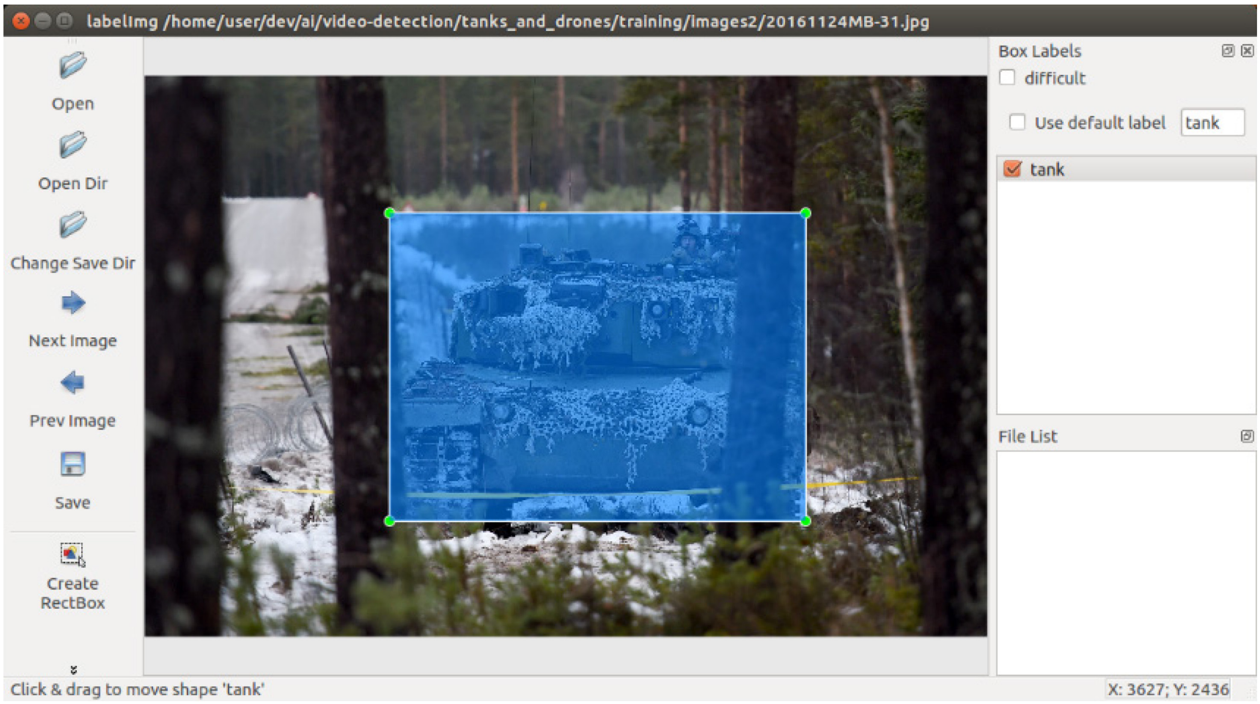


Figure 4: Labelling Images for Retraining Deep Learning Models to Look for New Objects.

3.2.2 Pre-Existing Video Clip or Image

One problem with social media postings is that they may go viral before any kind of authenticity of the content has been established. If a video posting of a bombing of civilians in a remote area of the world is made, and goes viral, a lot of opinion can be formed regardless of whether this is true or not. It may be that the posting is from a prior event with scrambled, edited, tiny fractions of clips put together for this purpose only.

The tool will be able to determine whether a video posted on-line does exist in our prior database or not within a reasonable delay. Currently we have a system for searching through large amounts of prior videos, but the real challenge is in the startup and collection of the database. The indexing technology being used is called perceptual hashing and is described primarily for use with images, but has also been used for videos [33] which is what we have optimised to help us search for video clips.

There are many ways perceptual hash may be calculated, but the main principle is shown in Figure 5 and we have used the same principle as described in Ref. [34]. To convert an image to a perceptual hash you first convert the image into a small representation of your image in "greyscale" values as shown in the middle of Figure 5 as an 8 x 8 image. This representation can focus on removing minor details and small shifts by just taking an average representation of the underlying area, or it can be a Fourier transformed representation of the image to enable more detailed filtering of the level of details [34]. Usually this representation is made into the same size as the number of bits in the resulting perceptual hash. There are also many methods for converting this representation into the resulting perceptual hash, like average hash, pHash and distance hash. Average hash is made by calculating the average value of the representation pixels and marking each bit with regards to their value being above with "1", and below with "0" and thereby end up with a 64-bit representation of the image, which may be several megabytes originally.

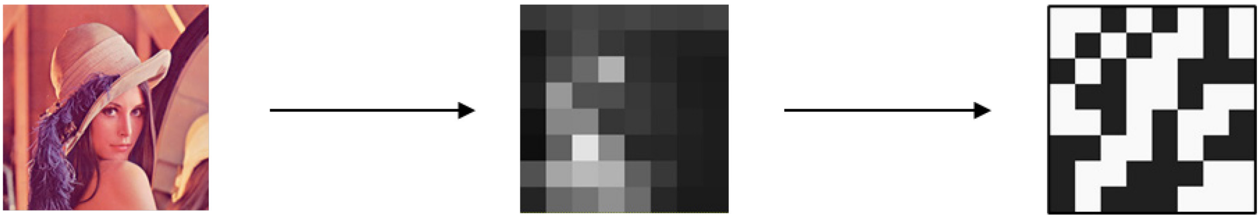


Figure 5: Process of Transforming Image into a Tiny Representation (Middle) and into a 64-bit Perceptual Hash (Right).

By calculating this on the stream of images in a video you get identifiers for each period of the video. The tool builds a database of these identifiers and compares incoming video clips to this database to see if the video has been published earlier.

3.2.3 Textual Content of Image or Video

Optical Character Recognition (OCR) technology is a completely separate area of research outside the scope of our research area, but an open source system [28] has been tested and integrated and works partially with plain English text inside the images on monotone background. As shown in Figure 6 this software does not work very well on angled text nor with text on image backgrounds, but we have still integrated it to be able to make some automatic triggers have information from images and videos.

Another important challenge with the open source system of text recognition is the speed when being used for video clips. The best system we have tested uses approximately one tenth of a second for each image, and while this is quite acceptable for images pulled from the feed, a video of a few minutes will now become a challenge. One solution may be to have the video divided up into scenes and only scan each major change in video content for new textual content. The technology for scene splitting does exist [35], [36] and will enable much higher speed for the OCR module, but is not integrated into the current system.

More complex commercial systems can also be integrated into the architecture if they exist. At the time of writing we do not know of any available system with satisfying performance.

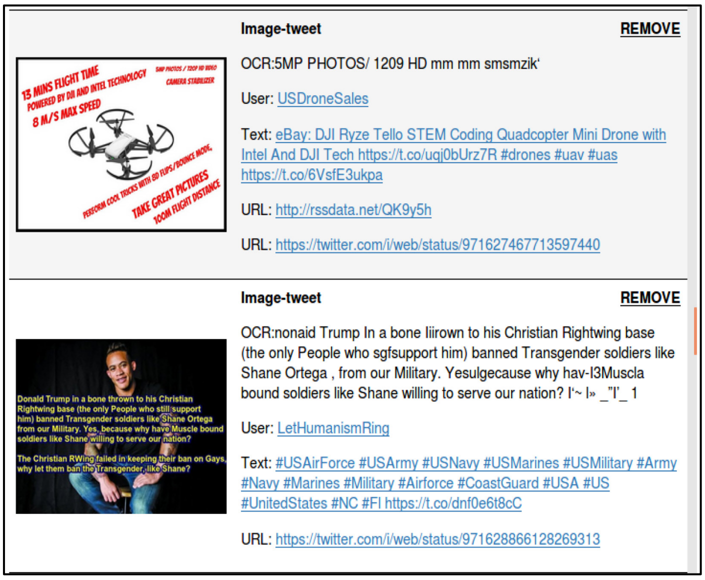


Figure 6: Current Integrated OCR Tool Does Not Work Well with Angled Text or Changing Background.

3.2.4 Deep Learning for Detection of Military Objects

Deep learning is becoming a popular machine learning technique to detect, localise and classify objects. For detection of objects, several deep neural network types are available that show differences in accuracy and speed. The network used in this study is the Single Shot MultiBox Detector (SSD) [28] which is a deep convolutional neural network that produces a set of bounding boxes at hypothesized object locations. That is, it indicates regions in the image that resemble the objects from the training images. In order to train the network, a substantial number of training images is required. For aerial images, such training images are scarce because the viewpoint is oblique (most training images are frontal or straight from above in case of satellite imagery), i.e., there are no large datasets of people or vehicles in aerial imagery. To overcome this, a novel methodology is explored for creating artificial training samples: by simulating images in a virtual environment.

Virtual reality can provide highly realistic images from aerial viewpoints. These virtual images can be captured and annotated. For our purpose, we used an adjusted version of the game Grand Theft Auto 5 to quickly generate highly realistic images. The images were created for a specific type of van and two types of tanks. The viewpoints can now be taken at bird's eye view, and from all sides of the vehicle, see Figure 7 for a few examples. In addition to viewpoints from around the vehicle, also different scenes and weather conditions were simulated.

In total, nearly 1700 images have been simulated among the variations of vehicle type, scene, position in scene, weather conditions, viewpoint and camera height. All images have been annotated by indicating the bounding box of the vehicle.

The tank images have been used to train a SSD network. Again, these images are scarce due to the oblique viewpoint and because tanks are not very common in images. Some real aerial images showing tanks have been used to test the detector, see Figure 8. Again, these images are scarce due to the oblique viewpoint and because tanks are not very common in images. With help from the Netherlands Ministry of Defence, we obtained a small number of useful images. The detections generated by this detector are most accurate when tested on data that look similar to the virtual data. This supports the view that a training data set is required that covers the many varieties of aerial data. Although the number of different tank models in the training data was limited to only two, this is not a limitation for learning the overall shape of a tank.

Figure 7: Creation of Training Samples by Taking Images from Virtual Reality. The top four images show a tank in the field from different viewpoints. The bottom four images show a tank in desert and forest scenes, and a van in a sunny and cloudy city scene.

Figure 8: Some Examples of Detections of Our Tank Detector on Real Imagery. The viewpoints in these test images are close to the ones in the training data, see Figure 7.

3.3 Video Analytics

3.3.1 Object Localization in Aerial Imagery

To limit the search, persons and vehicles are detected in the video. Persons are detected by the ACF detector [37], which provides bounding boxes in video frames. The boxes are checked against the expected size of persons, which is derived from the metadata about the video. The ACF detector is limited to persons that appear relatively large within the image, typically larger than a height of 50 pixels. Persons that appear smaller in the image, are detected when they move, by a Moving Object Detection (MOD) method. For MOD, the method in Ref. [38] was adopted and modified to search specifically for moving persons by considering the expected size of persons as derived from the metadata. The boxes provided by ACF and MOD are tracked to obtain temporal consistency and to remove erroneous boxes. The tracker is based on kinematics of the boxes, where speed and orientation in world coordinates are derived from the metadata. The association of boxes is performed by considering the consecutive overlap between the projected previous box and the current box. Two examples of ACF and MOD are provided in Figure 9.

Figure 9: Three Complementary Detectors: ACF (Green), and Two MOD Variants (Yellow and Red).

3.4 Social Media Analytics

The ability to extract, characterise, and visualise elements of an adversary's information campaign is central to the Coalition forces' situational understanding. This includes the ability to accurately identify individuals and groups associated with a specific threat, recognition of strategies employed to reach and influence others, and estimation of strategies that might be employed to counter or disrupt adversary activities within applicable rules of engagement. Within the emerging threat environments that Western democracies can be expected to face in the future this can be an extremely complex undertaking, given the complex nature of the social terrain. The socio-cultural landscape of non-combatants is increasingly dynamic, as hybrid conflicts are occurring in urban environments and the social media platforms by which information is exchanged is adapted to technology advances and social preferences. In addition, there are increasing constraints, in particular in Western democracies, due to ethical and privacy considerations regarding the collection and use of data and information and data. These "socio-technical-political" aspects have not been addressed in the study and technical developments have been viewed in isolation.

To deliver an effective set of capabilities to assist Warfighters in gaining a deep understanding of adversarial information campaigns, several technology applications must be pursued. These typically involve leveraging a variety of Natural Language Processing (NLP) tools for extracting meaning in multi-media. These might focus on extracting topics, sentiment, entities and relationships, and patterns of interest according to some established models of adversarial behaviour and belonging. Machine Learning (ML) approaches are often used to sort and categorise data to classify various actions or activities within the context of a problem of interest. Software tool development is most effective when specifically focused on small problems that may reflect shifting or temporary challenges. Take, for example, a software tool designed to identify messages in a social media stream that are machine produced, known as bots. Bots are used as force multipliers in the IE that are used to spread narrative and attack the counter narrative [39]. In the short period of time that these elements have been active, their use and behaviours are rapidly shifting with more coordination among the machines becoming apparent. Accordingly, the software used to identify and track bot behaviour must be flexible to accommodate new threat behaviours and strategies. Another example could be a software tool designed to classify the top users in a social media dataset according to adversary models. Given the nature of threats posed by rogue nation states, non-state actors, and new adversary group actors, monolithic threat models would be unhelpful and need to be updated as new threats are discovered. In this case, ML algorithms must be monitored and shaped as true positive and true negative cases (that are used to train the classification algorithms) are identified.

A basic NLP application to understand adversary messaging is the wide range of speech and language translation tools. Common technologies for translation include Automatic Speech Recognition (ASR), Machine Translation (MT), and Optical Character Recognition (OCR). OCR is potentially useful in extracting information from imagery when an analyst is trying to determine the location represented in the image. For example, an image might contain a symbol that could be used for identification purposes, such as a flag, road sign, or building sign. Translation technologies have seen significant progress in resources with languages with high resources.

Social media analytics provides an additional challenge of managing the data exploitation problem. The information stream that must be monitored is a large mix of structured and unstructured data that includes text, imagery, and video coming from a mix of social media and open source platforms, each with their own data types and structures and metadata components. To compound this situation, the popular social media platforms can be transitory and may make tools that concentrate on one data type to be outdated rapidly. Due to the quickness with which information is posted, shared, and used to influence groups, software applications also need to be modular and flexible in terms of design time and ability to integrate into a larger workflow. From a Defence acquisition standpoint, traditional timelines of technology development taking years can no longer be feasible. This is very true within the perspective of exploiting the information space (especially in the social media domain) when considering the contrasts between Coalition forces and adversary State and non-State actors. While the former is constrained by legal and ethical considerations for information use, processing, and exploitation, the latter choose to be unfettered from the same restrictions. This ultimately causes delays in countering adversary messaging and may impact influence campaigns.

3.5 Biometric Analytics

The Augmented Reality for Tactical Edge Analysis system (ARTEA) is a technology developed for the Combat Capability Development Center-Army Research Lab (CCDC-ARL) as a Small Business Innovative Research (SBIR) project that progressed to Phase II. ARTEA was designed to meet intelligence requirements for Soldiers at the tactical edge of today's military operations to address the gap that current Intelligence, Surveillance, and Reconnaissance (ISR) systems and processes do not leverage wearable computing technologies and evaluate Social Media (SM) in an accurate and timely manner. This is a major gap since social media contains near real-time information related to persons of interest, Area of Interest (AOI) activities and underlying attitudes of the population. ARTEA is based on three complementary pillars:

Big Data / social media analytics; Augmented Reality (AR); and Collaborative Sense Making (CSM). Each is unique and necessary to provide edge Soldiers with the information and awareness they need to be successful. The Phase I effort successfully demonstrated facial and document recognition with Google Glass and Big Data / SM analytics, particularly in the areas of user account classification and relevant motif or pattern detection. These advancements not only increase the Soldiers' situational awareness, but also the ability to forecast the populace's likely next courses of action.

The functional view for the ARTEA system is displayed in Figure 10. The system is capable of ingesting multi-source data inputs from a variety of social media platforms. These data are processed in the social media pipeline to filter noise, identify patterns, calculate features, and create tactical meaning such as anomalies and threats in the surrounding area. In the upper right section alerts are shown being sent to the Soldier via Google Glass technology. These alerts can range from threats to recent events to sentiment expressed by the local populace. The technology links tactical edge Soldiers to higher level tactical operation centres. Figure 11 provides a Civil-Military scenario for the use of ARTEA.

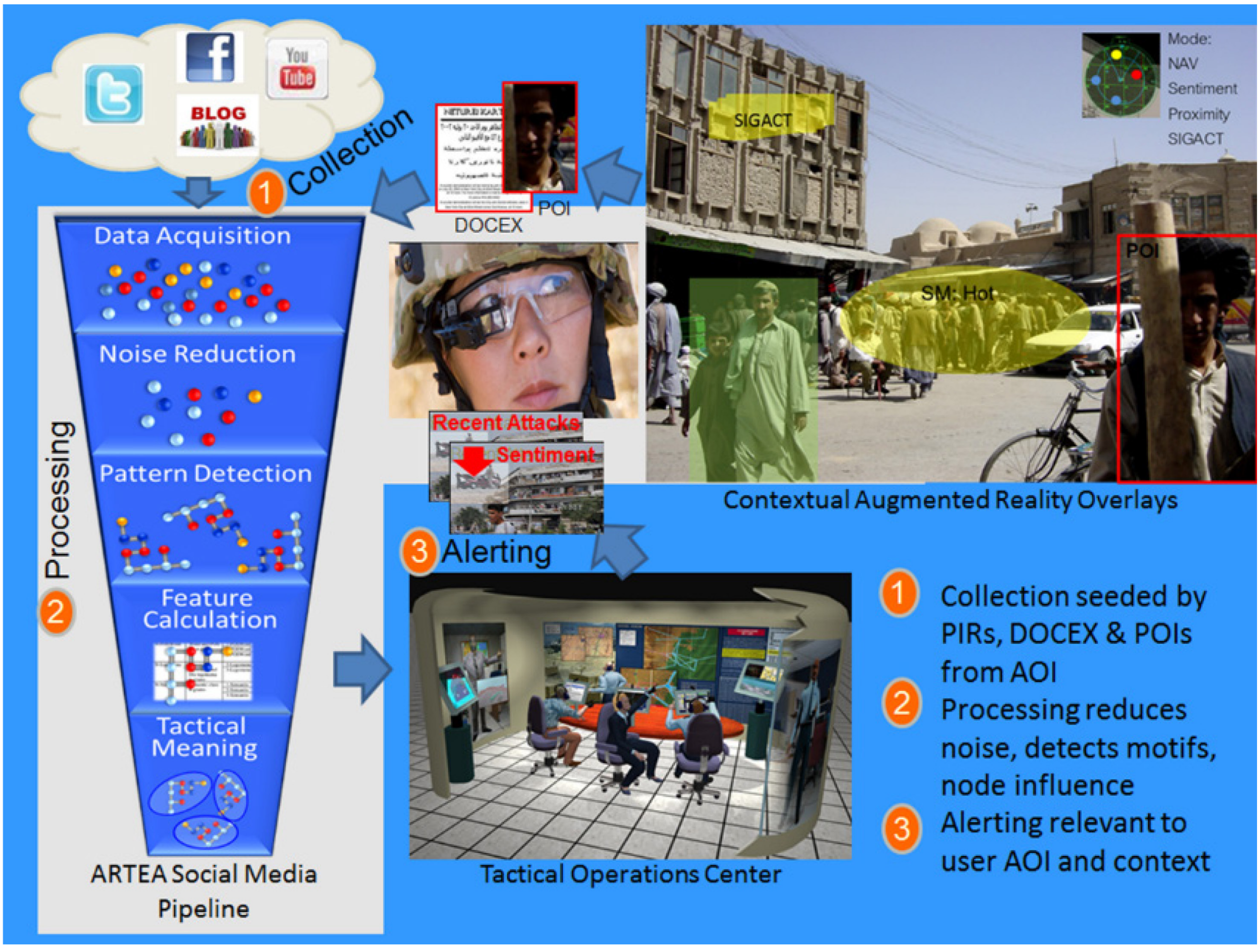


Figure 10: The ARTEA Concept of Use Demonstrated in a Tactical Scenario.

Figure 11: ARTEA Scenario Described Within the Context of Civil-Military Operations.

4.0 CONCEPT DEMONSTRATOR

4.1 Story Board

A realistic and representative work detail for intelligent analysts supporting coalition operations had to be defined by the research time in order to bring the essence of the multi-modal analytics ‘to life’. A common story, with applications in many national and coalition operations, is the identification of potential insurgent/terrorist activities and pre-emptively thwarting their actions by gathering evidential indicators/intelligence that is actionable. A story board outlined in Table 2 was considered to be most pertinent in line with the technical focus of the research which is analytics of text, image and video modalities. That is improving decision making by integration of text and video information. It aligns with, and brings into context, the abstract processes described earlier and depicted in Figure 1. The table has two columns, the first outlining the query/hypothesis and the second the potential outcome of the analytics leading to forming a decision or action.

Table 2: Story Board Integrating Text and Image/Video Analytics.

Flag	Query (Operations Area)	Outcome of Analytics
1	Rising insurgency indicators on social media, inter-agencies and intra-agencies intelligence reports?	Text analytics of reports, social media and linked relationships with other data indicate: • Insurgents in a particular area; • May be using vehicles (jeep?).

Flag	Query (Operations Area)	Outcome of Analytics
2, 3	Availability of imagery, including Full-Motion Video (FMV), and any indicators therein? Moving jeeps correlated with reports?	UAV available and video analytics indicate: • Moving vehicles in area; • Identification of Jeeps, indexed and being tracked.
4	Where is the specific jeep?	Updated deep learning models to search for jeep of interest in imagery: • Located and activity imaged; • Cross-cue links.
5, 6	Information available from other modalities?	Text mining reveals social media posts available and analytics reveal; • Suspicious behaviour; • Possible insurgents and roles.
7	Same insurgents at different points?	Database queries and biometric analytics reveal: • Identities of insurgents.
8	What will they do next?	Cross-cueing between different data sources enables the analyst to verify and interpret the information leading to action to interdict.

4.2 Scenario

Though the story board can be applied to other scenarios by changing the queries, the following fictitious, but realistic and representative, scenario was in mind of the researchers. It is based on a topical region for coalition operations, Afghanistan, and involves a Coalition Brigade Combat Team (BCT) operating in a region south of Kandahar. This is shown in Figure 12 with overlays depicting key locations and some examples of analytics for the scenario and the operational context for the analysts. Fuller details and additional schematics are available in Refs. [23] and [24].

In summary, the analysis cell has received unconfirmed reports that some of the leadership from one of the primary ISIS groups currently located in Karz is splintering and moving to an unknown sector. The basis of the split, the specific leaders of the splinter group, and the intended location of any move are currently unknown but serve as priority information requirements. The BCT has established four sectors from Kandahar to Marjah, a town 170 km to the west. Five Checkpoints (CPs), see Figure 12, exist along the Ring Road connecting Kandahar to the western edge of the Helmand Province.

The two potential targeted cities are Lashkar Gah and Gereshk, which lie to the South and North of the Ring Road, respectively and marked in Figure 12 with red circles. The analysis challenge is to monitor all available communications to alert NATO troops in the western Helmand regions of suspicious Taliban movement activity towards CPs indicating a surge of fighters. Of particular concern is CP3, where the route to either target city will become clear and pre-emptive action can be taken.

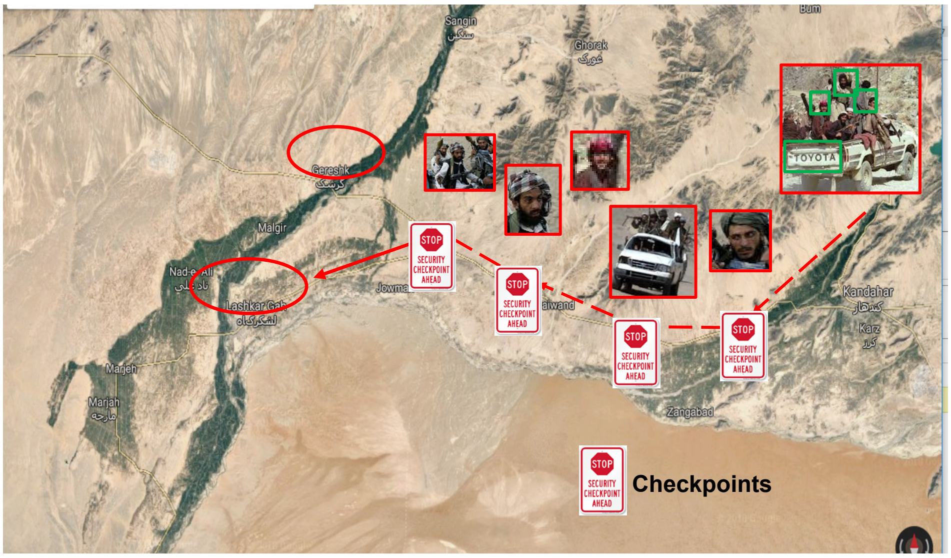


Figure 12: Military Scenario with Applied Analytics.

4.2.1 Applied Analytics

The analytics capabilities researched by the RTG can be used by the analysis cell to address the challenge. An overall process was depicted by the story board that was described earlier. In Figure 13, we have summarised the steps in a montage linked to the abstract process diagram that was introduced earlier. Though the steps are not enumerated one-to-one with the story board, the sequence and the applied analytics are equivalent to those described in the story. Implementation of these within an integrated system, and with relevant and available modal data and databases, would lead to improved decision making and informed pre-emptive action using text and image/video analytics.

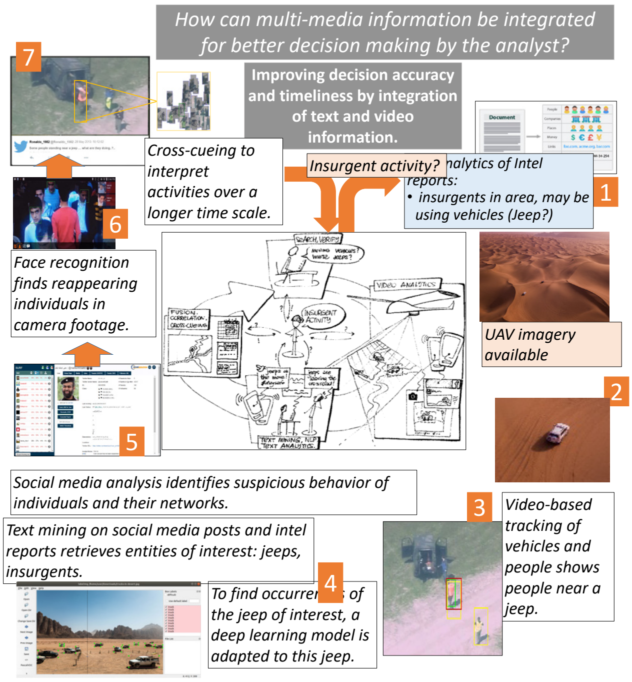


Figure 13: Montage of Applied Analytics to Military Scenario.

4.3 Components and Capabilities

4.3.1 Cross-Cueing Between Imagery and Social Media Sources

In this subsection, we give an example of how text and video analytics can be integrated to support the analyst to rapidly form situational awareness. The basis is the video analytics from Section 3.2, which localises objects in imagery. In this subsection, we build on top of this, and show how it can enable the analyst to perform cross-cueing between various imagery sources and social media sources.

The starting point for the analyst's search is a simulated Twitter message containing text and an image, displayed in Figure 14. The image displays the jeep with the woman and the two men. Analysts report that the incoming videos from social media and surveillance assets show increasing volume of vehicles and people. They need to determine types of activity that might be occurring to indicate intentionality. Twitter reports suggest that a woman in a jeep with two men might be of particular interest. The analyst wants to cross-cue from the Twitter reports to surveillance data from other sources including UAVs.

Figure 14: The Twitter Message that Suggests a Woman in a Jeep with Two Men Might be of Interest.

The analyst will search for the woman in surveillance data, consisting of Full-Motion Video, recorded by a UAV. In total the video spans 32 minutes, including infrared and daylight imagery, in various field of views. The recordings are accompanied by metadata about the position of the UAV and its field of view, as shown in Figure 15.

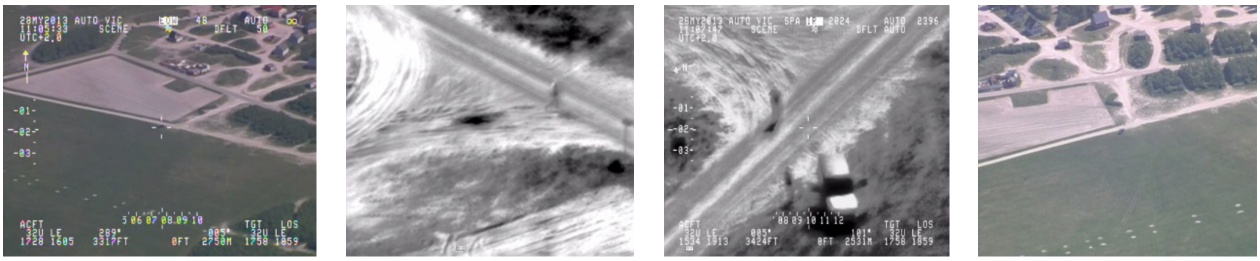


Figure 15: Impressions of the Full-Motion Videos Recorded by the UAV.

The analyst searches the woman in the Twitter image in the UAV imagery. The analyst indicates the woman in the image, by selecting a bounding box, from which a 'snippet' is extracted to show the picture of the woman, as depicted in Figure 16.

Figure 16: The Analyst Selects the Woman as a Starting Point for the Search.

The analyst wants to use this snippet to cross-cue to the UAV videos. The snippet of the person of interest, i.e., the woman, is now searched within all tracks in the UAV videos. From the tracks, snippets are extracted. The snippet of the woman (i.e., the query snippet), is compared with all other snippets (i.e., the candidate snippets). For comparison, the re-identification algorithm from [40] is deployed. It provides a score between two snippets of 1 (similar) and 0 (dissimilar). This score demonstrated a large degree of robustness for a change of camera, viewpoint and illumination [40]. This is beneficial for searching in a large volume of surveillance data, potentially recorded by various cameras. Pair-wise scores are computed between the query snippet and the candidate snippets.

The similarity scores are presented to the analyst on an intuitive two-dimensional canvas. For the projection onto the canvas, the t-SNE embedding [41] is applied. This embedding groups similar snippets together, with the rationale that local structure is preserved, and dissimilarities are less important. The grouping of similar snippets aids the analyst to search for the woman in the surveillance video data.

A Graphical User Interface (GUI) has been developed to sift through the canvas of snippets. In practice, there may be a large amount of snippets. These will be projected on top of each other. The GUI enables the analyst to show more or less snippets on the canvas, and to navigate through the canvas. A single snippet often does not provide a good impression of the person's activity. When a snippet is clicked by the analyst, the accompanying part of the video is showed. While the analyst is searching, snippets can be confirmed and added to the search results. The confirmed search results can be showed on a timeline viewer, such that the analyst can playback the related video parts and analyse intentionality.

The graphical user interface with the image snippets of persons is shown in Figure 17. The interface is a canvas of snippets, where their position is based on the similarity embedding (see previous section). On the left of Figure 17, all snippets are shown. On the right, the zoom-in mode is shown, which are the snippets inside the red rectangle depicted on the left. When a snippet is clicked by the analyst, the video part plays, which is shown in Figure 6 inside the pop-up window. This enables the analyst to explore the persons in the imagery data and their activities. Figure 17 illustrates this by the search for the person of interest (the woman). She was found in a UAV video, while carrying a shovel.

The person of interest (the woman) was found multiple times in the imagery data. This is shown in Figure 17 on the right side. The analyst has confirmed these appearances by double-clicking the snippets. In Figure 18 this is depicted by the green contours. The analyst has given these appearances the label 'woman at jeep'. This tag enables the analyst to trace-back to the original imagery data based on a text query.

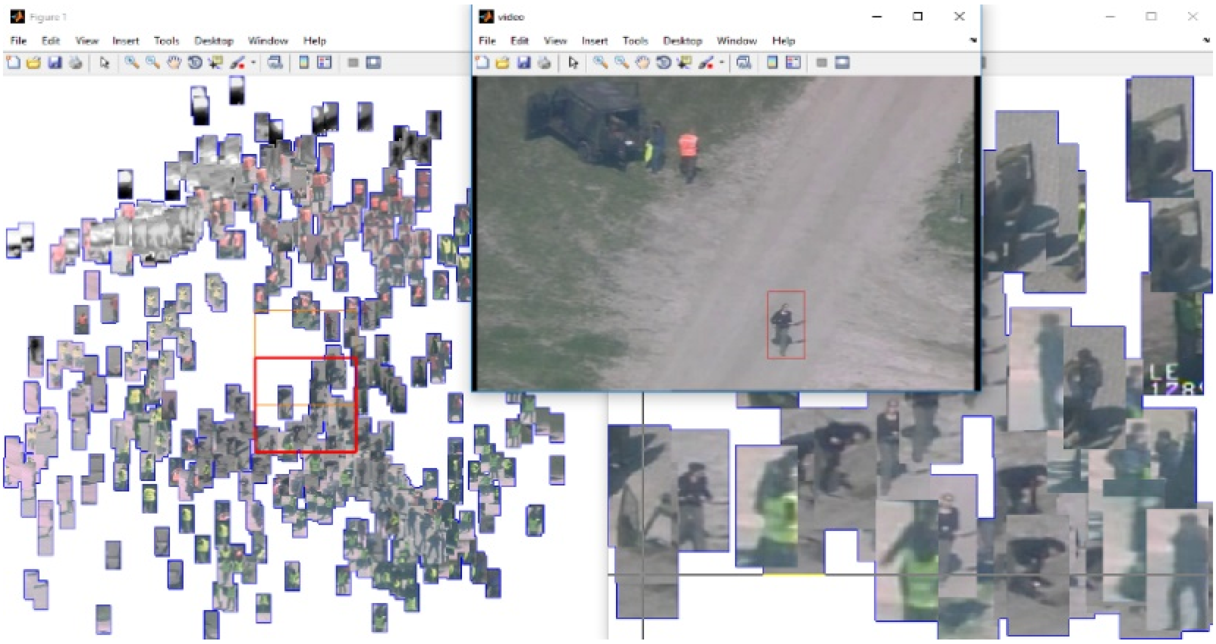


Figure 17: The Graphical User Interface with the Image Snippets of Persons and their Local Grouping by the Embedding Based on the Similarity Measure (See Text).

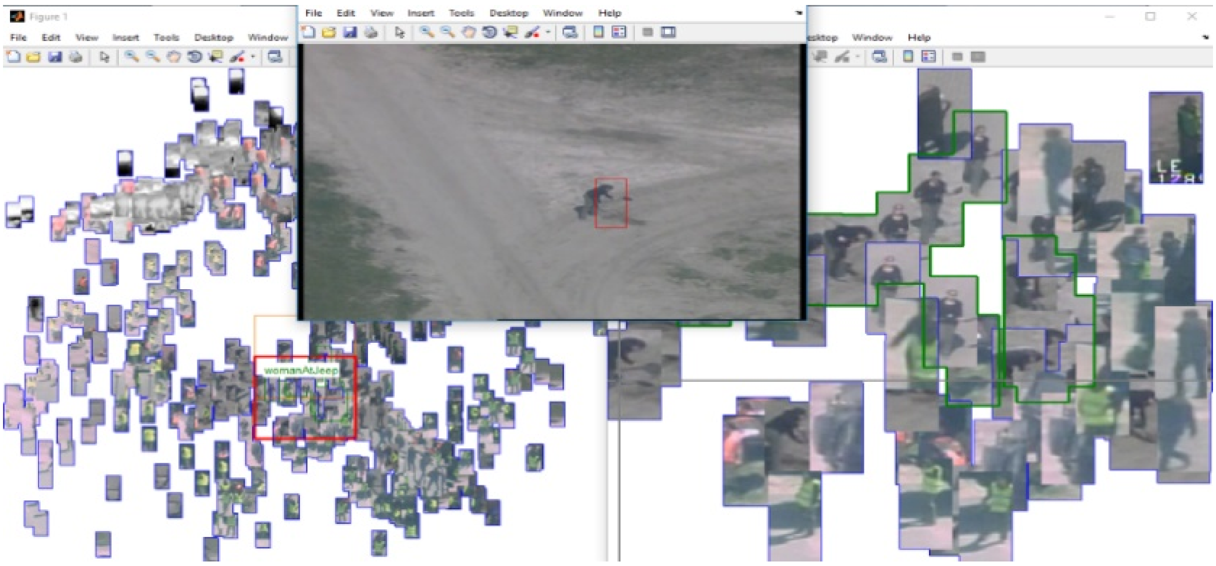


Figure 18: Multiple Appearances of the Woman Have Been Found by the Analyst (in Green).

To determine intentionality, the analysts needs to analyse what happened where and when. To aid the analyst with this task, the tagged appearances of the person of interest (the woman) are shown on a timeline (when), map (where), accompanied by the imagery (what). Figure 19 shows the timeline viewer, with its timeline (bottom), map (left) and imagery (right). On the timeline, the analyst can mark a time point (one image) or interval (part of the video). The interesting behaviours can be indicated by a bookmark with a tag. Figure 19 shows the moments that the woman steps out of the jeep, then digs, and finally leaves the area.

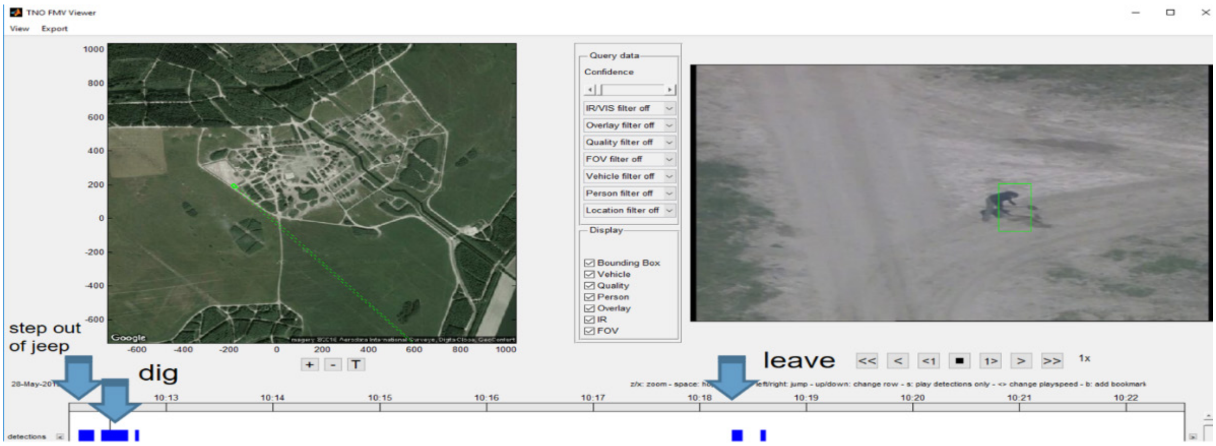


Figure 19: Timeline View of the Found Appearances of the Person of Interest and Her Activities.

5.0 OPERATIONAL EVALUATIONS

5.1 User Evaluations for Concept Demonstrator

Though attempts were made to gather feedback from concept demonstrator events, while positive on the whole, they were qualitative and anecdotal. More confidence on the quality of the work, the direction and approach was gleaned from the specialist meeting. This as reported earlier provided positive feedback and also indicated a strong recommendation to include human factors aspects which is included in the recommendation section.

One area where more quantitative feedback was provided by users was for video analytics as reported in the next section.

5.1.1 Feedback from Military Users About Video Analytics

User evaluations of video analytics were performed by military Image Intelligence (IMINT) analysts of the Netherlands MoD. Three different experiments were performed:

1) Experience sharing: Each analyst shared his/her experience about one task that was difficult to perform. Qualitative findings:

   a) Tasks get hard when an event has spread across a long time, large space, or across different recordings.

   b) Sometimes a significant amount of details should be visible in the image to determine who it is or which vehicle it is. Finding the right image with sufficient detail.

   c) Analysts have an incredible quality of interpreting situations based on many different pieces of information.

2) A serious game: The analysts performed the task on a fictitious but realistic scenario, i.e., to find the insurgent who planted an IED. A map and an intel report provided certain facts to the analysts as a starting point. Vehicles and people were displayed on game cards that were added to the map whenever the analysts asked 'the system' a question. The system was actually a researcher who knew every detail about the scenario and replied to questions of the analysts if it was an automated system.

Qualitative findings:

a) A serious game is a good way to abstract from any specific implementation details.

b) Any decision support should provide a sense of the time and spatial scale, which are needed by the analyst to make assessments about whereabouts and travelling.

c) It is important for the analyst to know where the automated system may have failed.

d) There are many ways in which the analyst is doing the task, yet a common finding was that analysts want to have an understanding of a location and time to start from there and extend the analysis step by step towards bigger time scale and/or bigger spatial scale.

3) Working with a real system: The GUI in Figure 20 displayed 5 hours of data in which the analysts had to perform another task, i.e., to find an insurgent who took a civilian hostage. Some details about sightings of the civilian at certain places and timing were provided to the analysts as a starting point. Note that this GUI is the one from Section 1.3.2.1, where the image analytics (Section 1.2.2.1), video analytics (Section 1.2.3.1) and cross-cueing (Section 1.3.2.1) were brought together.

Qualitative findings:

a) At first, it is really hard for analysts to work with a highly specific GUI with many options and sometimes complicated outputs.

b) Filters for selecting parts of the data are very important, including: the type of sensor that generated the data, the location of the data, the time of the data, the type of detected objects, the field-of-view of the camera. As an example, to identify a person, the analyst would turn on the filters for: RGB (EO sensor), zoom-in, object of interest = persons.

c) A compass in the GUI (in this case an arrow pointing North) is necessary to have a sense of where somebody is travelling.

Figure 20: GUI Used for User Evaluations.

5.1.2 Feedback from Military Users About Text Analytics with Video Inputs

The SURF capability, described in Section 3.1.2, was used as an exemplar social media exploitation capability that provides text, imagery, and relationship linkages of individuals linked in social networks by systems of classification models. This was demonstrated in May 2019 at the Army Enterprise Challenge 2019 Exercise at Ft. Huachuca, AZ. Over 15 Open Source Intelligence (OSINT) analysts exploited a social media dataset in the realm of Multi-Domain Operations (MDO). They used SURF in a MDO scenario for preparatory intelligence activities to characterise adversary actions and formulate a credible US information narrative. Key results from this interaction with analysts included:

- The ability to classify data elements (social media users, signals, events) into membership groups using pattern association machine learning algorithms provides a language-agnostic analytic that is flexible and robust.
- The ability of analysts to modify current membership groups and create new classes makes SURF adaptive to dynamic battlefield operations.
- SURF visualizations would be enhanced with geospatial overlays correlating to areas of operation for relevant echelons of command.

The impacts of the results from this exercise verify that the SURF tool is a valuable component in a larger toolchain designed to classify levels of threat from adversaries and validate that the key components of the SURF tool match Intelligence requirements for Multi-Domain Operations. The successful demonstration and characterization of SURF during EC19 allows for smoother transition opportunities within the Army system, especially with regard to future capability drops for decision making tools. SURF is currently used by elements of the Intelligence Community (IC) and the EC19 demonstration and characterization results will strengthen those relationships.

5.1.3 Feedback from NATO Chiefs of Transformation Demonstration

The results of the Concept Demonstrator were displayed to the NATO Chiefs of Transformation Conference (COTC) in Norfolk, VA December 13, 2018. The COTC is an annual event conducted by Headquarters Supreme Allied Commander Transformation, directed primarily at national Flag/General Officer level or civilian equivalent decision makers, from NATO and Partner Nations, involved in national Transformation and longer-term military capability development. It remains the only venue that gathers the national Chiefs of Transformation, senior leaders from NATO, and Allied Command Transformation’s leadership together for open interaction and discussion to inform and support Alliance and National thought/leadership. The COTC Conference 2018 Theme was “DISRUPTIVE ADVANCES SHAPING WARFARE.” Within this context, members of RTG IST-144 briefed and demonstrated exemplar text and imagery exploitation capabilities to NATO leaders.

6.0 CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

The R&D by a team from USA, UK, NOR and NLD has advanced capabilities in the following areas and shown the ‘art-of-the-possible’ through concept demonstrations:

- Intelligent Capture and indexing of motion imagery;
- Expansion of Machine Learning / Deep Learning approach to semantic video analytics; and
- Cross-cueing from text analytics to drive/exploit video and imagery indexing and retrieval.

Their utility has been shown by application to a realistic military scenario using a story board outlining applied analytics processes. These were mapped to, and represented by, processes for intelligence analysts that the project has defined in an abstract cartoon form. This though simple proved to be invaluable in galvanizing the efforts and focus of the research team. It is offered to other researchers as a complement to Boyd's OODA loop and the Pirolli and Card Intelligence Process Cycle. Being in cartoon form may be more attractive to early career scientists as alternative approaches to capturing and visualising processes. These capabilities were also demonstrated at several events where military analysts and decision makers were able to interact with the technology to determine value to current operational methods and tactics, techniques, and procedures.

This RTG explored potential architectures and frameworks that could optimise the implementation of multi-media analytics in distributed coalition environments. In the area of DL, there is a significant amount of research in both the academic and commercial sectors into large-scale machine learning, which has produced state-of-the art results as outlined in previous sections. The capabilities are now commonly used for commercial use for image processing (Google, Baidu), audio processing (Google, Apple, and Microsoft) and the financial sector. There are many defence and security challenges that could benefit from these techniques, particularly when triaging large data sets (e.g., image exploitation, biometrics and security, speech analysis). As we consider applications, further research and performance measurements are needed, especially false alarm rates with large data sets particular to defence and security. System co-design is needed, algorithms and hardware, to be able to engineer and integrate system architectures that are optimised for CBA and DL in general. Technical areas include, but are not limited to the following:

- Hardware for DL, including neuromorphic chips and low-power, high-performance field programmable gate arrays;
- New optimization methods across the convolutional layers and back propagation techniques;
- Training with fewer examples; and
- Distributed learning across systems.

An overall challenge is how best to integrate with and best exploit the current and future commodity distributed systems (e.g., servers, cloud) and distributed services (e.g., computation, storage) expected to be available to an enterprise. Included in this goal is how those services would be extended to support collaborative endeavors and leverage the distributed systems capability for research, development, and experimentation. A huge improvement for video analytics is expected when contextual information can be successfully incorporated such that the algorithm knows when and where to focus on a subset of the scene and actors.

ML systems are a key element in virtually all decision support systems, autonomous systems, and other systems that are important in NATO operations. These systems will over time have the potential to influence both control and over vehicles, sensors and weapons, as well as the decisions made from sensor inputs. They need to be human-centric as human involvement will always remain critical. Hence consideration of human factors is important as are socio-technical aspects such as Ethics and Privacy.

There are many activities, projects and programs that look at manipulation of Machine Learning Systems (MLS) and how specific systems can be influenced by creative input. But there is too little activity in machine learning research to look at how we can create more robust systems and how such systems might require a fundamental change in training, testing, validation and/or product phases.

One problem might be that commercial MLS may be trained in ways that cannot be verified through the product. Can the products contain back doors in the system, much like software in general, only made by creatively crafting the input/training data? e.g., is it possible to train a missile detection system, that is trained to report no detection on one specific type of missile, and that this manipulation cannot be detected because the machine learning model is too large and complex? A proposed recommendation for an RTG will look into methods for how such training can take place, how training can take place which will avoid these types of challenges, and how systems must be documented in order to avoid being the victim for such solutions as a customer.

Data from military sensors are being fed directly into systems for fast analysis and decisions. Robustness in training phase is only one step towards a more robust overall system. Military systems also need sensor input to be unpredictable enough that the analysis will not be compromised with fake data. Robustness in operations will also be an important area of research.

Another problem might be the accountability of using MLS when decisions have been made. How can the decision be documented at the time of the event in a way that later can be verified was correct with information currently available. This accountability will require machine learning systems, especially dynamic MLSs, to have major changes from today's "take it or leave it" output.

6.2 Recommendations

With the extreme growth of Machine Learning Systems into military equipment, it is important to cover the potential problems listed above. In order to achieve trust in military systems using complex machine learning models and algorithms, the military needs to be able to prove both robustness and accountability. Robustness is important for the availability and integrity of any military system, with or without both sensors and effectors. Accountability is likely a future requirement for such systems, and the more complex a system becomes, the documentation of accountability will grow towards "non-human" complexity. Military decision makers must be able to document how the military decision making systems operates in order to show why their system recommended those specific actions based on the input from these specific sensors.

Accordingly, it is recommended that a follow-on activity be sponsored to achieve the following objectives:

- Determine the SOTA in robustness and accountability for MLS. Especially DL systems with complex and large models which are virtually impossible to manage by humans.
- Examine whether a methodology can be made to verify that commercial MLS, e.g., cloud MLS, comply with a set of criteria. Including what kind of criteria that may be possible, and what should be mandatory in a military setting.

The following topics would be appropriate for study in a follow-on activity:

- SOTA in robustness of MLS – during training and operations.
- SOTA in accountability in MLS.
- Methods of interaction between different nations' MLSs in a federated military environment with regards to robustness and accountability.
- Including possible vulnerabilities and countermeasures.

Such a follow-on activity undertaken by NATO was proposed towards the end of this work as an RTG. It was approved, to start this year, to address the subject of 'Robustness and Accountability in Machine Learning Systems' (IST-RTG-169). Work will further determine the foundational elements of how computational algorithms can be understood and can accurately exploit context in human scenarios. Within that purview it is also recommended that an integrated demonstration on multi-media analytics be undertaken with a validation component with relevant military reviewers. The efforts expended by IST-144 and demonstrated to various military audiences demonstrates the importance and value of integrated multi-media analytics for a variety of defence challenges and should be expanded for further exploitation of emerging technology advances.

Other recommendations relate to wider system of systems aspects, human-machine interactions and socio-technical issues. Research is needed towards an open distributed system of systems for the coalition providing services for analytics and information fusion (e.g., through knowledge representation and reasoning using knowledge graphs.) Human involvement will always remain critical in the coalition process and includes sociotechnical aspects such as Ethics and Privacy.

Hence the two more key recommendations for further NATO led research study, working between panels (e.g., IST and HFM) are to:

- Investigate designs for human-centric distributed system of systems, leveraging commodity cloud and ML developments, and emerging knowledge representation approaches, to provide the multi-modal analytics services needed by NATO applications. This includes experimentation to explore the human analysts' ability to rapidly and accurately incorporate multi-modal information with a variety of visual displays; and
- Explore potential approaches for the coalition to address emerging socio-technical issues, such as Ethics and privacy, arising with human-machine systems.

7.0 REFERENCES

[1] General Sir Gordon Messenger, UK MOD, Vice Chief Defence Staff, "Military Must Wake Up to Information War", https://www.thetimes.co.uk/article/general-sir-gordon-messenger-military-must-wake-up-to-information-war-cqflf90fj#:~:text=%E2%80%9CWe%20have%20to%20wake%20up,an%20anti%2Dtank%20missile%2C%E2%80%9D. Accessed 1 March, 2018.

[2] UK Royal Navy, "The Future Navy Vision: Royal Navy Today, Tomorrow and Towards 2025". https://www.royalnavy.mod.uk/About-the-Royal-Navy/~/media/Files/Navy-PDFs/About-the-Royal-Navy/Future%20Navy%20Vision.pdf. Accessed 8 March, 2018.

[3] US DOD, Joint Doctrine Note 1-18, 25 April 2018. https://www.jcs.mil/Portals/36/Documents/Doctrine/jdn_jg/jdn1_18.pdf?ver=2018-04-25-150439-540.

[4] Boury-Brisset A., et al. "Complex Event Processing for Content-Based Text, Image and Video Retrieval", NATO Exploratory Team Report, RTO IST-ET-86 CBA, November 2015.

[5] Madahar, B. et al. "Technical Summary Report: Specialist meeting (NATO-IST-158-RSM-10) Content Based Real-time Analytics of Multi-Media Streams (CBRAM)", NATO, IST-158-RSM-10, September, 2017.

[6] Pirolli, P., S. Card. "The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis", In Proceedings of International Conference on Intelligence Analysis, volume 2005, pp. 2-4, 2005.

[7] Burghouts, G. "NATO RTG-144 Experiment TNO: Acquiring Intelligence from Text and Video", TNO Report, TNO, The Hague, 2017.

[8] Broome, B. et al. "Next Generation Analyst". SPIE Defense, Security, and Sensing, Baltimore, Maryland, United States, Proceedings Volume 8758, N; 875801, 2013.

[9] Bowman, E., Jobidon, M.E., Guyard, A.B. "Human Factors Considerations in Text and Video Analytics", International Command and Control Research Technology Symposium (ICCRTS), CA, USA, Nov 6 – 8 2017.

[10] Guyard, A.B. “A Contextual Query and Analysis Framework for Heterogeneous Unstructured Text”, International Command and Control Research Technology Symposium (ICCRTS), CA, USA, Nov 6 – 8 2017.

[11] E. Bowman et al. “Complex Event Processing for Content-Based Text, Image, and Video Retrieval”, US Army Research lab, Technical Report, ARL-TR-7705, June 2016.

[12] Bowman, E. et al. “Complex Event Processing for Content-Based Text, Image, and Video Retrieval”, International Conference on Military Communications and Information Systems (ICMCIS), Brussels, Belgium, May 2016, Proceedings, IEEE Explorer, https://doi.org/10.1109/ICMCIS.2016.7496546, June 2016.

[13] Bowman, E. et al. “Advanced Text and Video Analytics for Proactive Decision Making”, SPIE Defense + Security, Proceedings Volume 10207, 2017.

[14] Bowman, E. et al. “Content-Based Multimedia Analytics: US and NATO Research”, SPIE Defense + Security, 2018, Orlando, Florida, United States, Proceedings Volume 10653, 2018.

[15] Text/Video Motifs: 10th Operations Research and Analysis Conference, NATO SACT in Norfolk, Virginia, 17 –18 Oct 2016.

[16] Demonstrating Text/Video Analytics: SPIE Next Generation Analyst Conference, CA, USA, 10 – 12 Apr 2017.

[17] Schutte, K. et al. “Long-Term Behavior Understanding Based on the Expert-Based Combination of Short-Term Observations in High-Resolution CCTV”, Proc. SPIE, vol. 9995, 2016.

[18] Wu, T., Gurram, P., Rao, R.M., Bajwa, W. “Human Action Attribute Learning Using Low-Rank Representations”, Signal Processing with Adaptive Sparse Structured Representations (SPARS) workshop, Lisbon, Portugal, 5 – 8 June 2017.

[19] Bowman, E. et al. “Content-Based Multimedia Analytics: Rethinking the Speed and Accuracy of Information Retrieval for Threat Detection”, NATO Operation Research Analysis Conference, NATO STO publication, MP-SAS-OCS-ORA, 2016.

[20] Bowman, E. et al. “Content-Based Multimedia Analytics for Big Data Challenges”, NATO Conference, Big Data and AI for Military Decision Making, Bordeaux, France, NATO Publication, MP-IST-160-S1-2, May 2018.

[21] Bowman, E. et al. “IST-RTG-144: Content-Based Multi-Media Analytics”, Poster and Concept Demonstrator, NATO ACT.

[22] Bowman, E. et al. “Complex Event Processing for Content-Based Text, Image, and Video Retrieval”. 2016 International Conference on Military Communications and Information Systems (ICMCIS), Brussels, pp. 1-6, 2016. doi: 10.1109/ICMCIS.2016.7496546.

[23] Demonstrating Text/Video Analytics: SPIE Next Generation Analyst Conference, CA, USA, 10 – 12 Apr 2017.

[24] Wu, T., Gurram, P., Rao, R.M., and Bajwa, W. “Human Action Attribute Learning Using Low-Rank Representations”, Signal Processing with Adaptive Sparse Structured Representations (SPARS) workshop, Lisbon, Portugal, 5 – 8 June 2017.

[25] BALEEN Documentation. https://github.com/dstl/baleen. Accessed March 8 2018.

[26] UIMA. https://uima.apache.org. Accessed March 8 2018.

[27] US DOD, "Social Understanding and Reasoning Framework (SURF)," Available from https://www.sbir.gov/sbirsearch/detail/1257717. Accessed March 8 2018.

[28] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C. "SSD: Single Shot Multibox Detector", In European Conference on Computer Vision, pp. 21-37, Springer, Cham, 2016.

[29] Tesseract Open Source OCR Engine, https://github.com/tesseract-ocr/tesseract/. Accessed March 8 2018.

[30] Viola, P., Jones, M. "Rapid Object Detection Using a Boosted Cascade of Simple Features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, pp. I-511-I-518 vol.1, 2001.

[31] Krizhevsky, A., Sutskever, I., Hinton, G.E. "ImageNet Classification with Deep Convolutional Neural Networks", In Proceedings of the 25th International Conference on Neural Information Processing Systems – Volume 1 (NIPS' 12), F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger (Eds.), Vol. 1. Curran Associates Inc., USA, pp. 1097-1105, 2012.

[32] "Tensorflow Object Detection API", On-line resource, Retrieved from https://github.com/tensorflow/models/tree/master/research/object_detection, Accessed March 8 2018.

[33] Testing Custom Object Detector – Tensorflow Object Detection API Tutorial, [Blog post], Retrieved from https://pythonprogramming.net/testing-custom-object-detector-tensorflow-object-detection-api-tutorial/, Accessed March 8 2018.

[34] Wolfgang, R.B., Podilchuk, C.I., Delp, E.J. "Perceptual Watermarks for Digital Images and Video," In Proceedings of the IEEE, vol. 87, no. 7, pp. 1108-1126, Jul 1999.

[35] Kjelsrud, O. "Using Perceptual Hash Algorithms to Identify Fragmented and Transformed Video Files", Master Thesis, Gjøvik University College, Gjøvik, Norway, 2014.

[36] Lin, T., Zhang, H. "Automatic Video Scene Extraction by Shot Grouping", In Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, vol.4, pp. 39-42, Barcelona, 2000.

[37] Dollar, P., Appel, R., Belongie, S., Perona, P. "Fast Feature Pyramids for Object Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1532-1545, 2014.

[38] Michael J. Lanham, M.J., Morgan, G.P. and Carley, K.M. "Social Network Modeling and Agent-Based Simulation in Support of Crisis De-escalation",IEEE Transactions on Systems, Man, and Cybernetics: Systems, 44(1): pp. 103-110, 2014.

[39] Xiao, J., Cheng, H., Sawhney, H., Hang, F. "Vehicle Detection and Tracking in Wide Field-of-View Aerial Video", CVPR, 2010.

[40] Bouma, H., Borsboom, S., Hollander, R. den, Landsmeer, S., Worring, M. "Re-Identification of Persons in Multi-Camera Surveillance Under Varying Viewpoints and Illumination", Proc. SPIE, vol. 8359, 2012.

[41] Maaten, L. van der, Hinton, G. "Visualizing Data Using t-SNE", Journal of Machine Learning Research, pp. 2579-2605, 2008.

Annex A – NATO SPECIALIST MEETING, IST-158-RSM

NATO OTAN SCIENCE & TECHNOLOGY ORGANIZATION

COLLABORATION SUPPORT OFFICE

SPECIALISTS' MEETING

CONTENT BASED REAL-TIME ANALYTICS OF MULTI-MEDIA STREAMS

IST-158 /RSM-010

organised by the

Information Systems and Technology Panel

to be held in

Middlesex University, The Burroughs, Hendon, London, United Kingdom

Wednesday 6 September - Friday 8 September 2017

This Specialists' Meeting is NATO UNCLASSIFIED open NATO nations and Sweden, Finland, Australia

ENROLMENTS:

Participants are requested to enrol on-line

https://events.sto.nato.int

Deadline for enrolments is 1 September

There are no conference fees.

If you are unable to enrol via the internet, please contact the IST Panel Assistant at: aysegul.apaydin@cso.nato.int

S&T organization CSO

Background

Information Systems Technology Panel (IST) is one of the seven Panels whose role it is to implement, on behalf of the Science & Technology Board, the STO Mission with respect to Information Systems Technology. The advancement and exchange of techniques and technologies to provide timely, affordable, dependable, secure and relevant information to war fighters, planners and strategists, as well as enabling technologies for modelling, simulation, and training are the focus of this Panel. The Information Systems Technology Panel covers the fields of Information Warfare and Assurance, Information and Knowledge Management, Communications and Networks and Architecture and Enabling Technologies.

Theme - Objectives –Topics

Content-based information multi-media retrieval and analytics, (Content-based Real-Time Analytics of Multi-media Streams (CBMA)), is a means to allow military experts to exploit data from multiple sources in a rapid fashion for sense making, decision support and knowledge generation. Elements of CBMA include contextual understanding of complex events through computational/human processing techniques, event prediction through the automated extraction of spatio-temporal features, hidden clusters, network structures and resource flows, and the use of machine learning and processing for automated translation, parsing, information extraction, & summarization of unstructured and semi-structured data from multiple streams. These types of complex analyses cannot be done in isolation. NATO and coalition military leaders, commanders, and intelligence analysts need interoperable tools that cross-cue knowledge obtained from one method to generate taskings in another. This requires a focus on building the cross-cued solution from advances in multi-media data analytics (e.g. text/image/audio). Results will significantly improve NATO abilities to generate knowledge from extremely large stores of text, imagery, and video caches to speed situational awareness and decision making.

TOPICS:

Example topics include, but are not limited to:

- Capture and indexing of motion imagery
- Architectural design concepts
- Human evaluations of exploitation results
- Imagery exploitation
- Deep learning for semantic video analysis
- Index generation for content retrieval
- Text Analytics to exploit video indexing
- Integrated text and video analytics
- Others

The Specialists' Meeting will commence with two keynote presentations and followed by approximately ten invited presentations to provide initial directions for the Specialists' Meeting discussions. The plan is to have an interactive Specialists' Meeting consisting of inter-disciplinary working groups.

Collaboration Support Office - Point of Contact

Mrs. Aysegül APAYDIN
IST Panel Assistant
NATO STO – CSO

Tel : +33 (1) 5561 2282
Fax: +33 (1) 5561 9626
Email: aysegul.apaydin@cso.nato.int

Programme Committee

Programme Co-Chairs

Prof. Bob MADAHAR
Senior Fellow
DSTL
United Kingdom
Email: bkmadahar@dstl.gov.uk

Dr. Elizabeth K. BOWMAN
Airforce Research Laboratory
United States
Email: Elizabeth.k.bowman@mail.mil

Members

Dr. Marie-Eve JOBIDON
DRDC Canada
Canada
Email: marie-eve.jobidon@drdc-rddc.g.ca

Dr. Alexandre BERGERON-GUYARD
DRDC Canada
Canada
Email: alexandre.bergeronguyard@drdc-rddc.gc.ca

Dr. Gertjan BURGHOUTS
TNO
The Netherlands
Email: gertjan.burghouts@tno.nl

Dr. Lasse OVERLIER
FFI'
Norway
Email: lasse.overlier@ffi.no

Dr. Diana MILITARU
ACCTM
Romania
Email: dmilitaru@acttm.ro

Dr. Peter HARVEY
DSTL
United Kingdom
Email: prharvey@mail.dstl.gov.uk

Mr. Ranjeev MITTU
NRL
United States
Email: ranjeev.mittu@nrl.navy.mil

Dr. Raghuveeer RAO
ARL
United States
Email: raghuveer.m.rao.civ@mail.mil

Wednesday 6 September 2017

BUILDING: HATCHCROFT BUILDING; ROOM: H116

09:00 **REGISTRATION and TEA/COFFEE**

10:00 **OPENING CEREMONY**

Welcome:
Prof. William WONG, Middlesex University, GBR

10:15 Introduction to the Specialists' Meeting by Co-Chairs:
Dr. Liz BOWMAN, ARL, USA and Prof. Bob MADAHAR, DSTL, GBR

10:45 **BREAK**

11:15 **KEYNOTE SPEECH 1:**
Using Motif Detection to Predict Group Membership from Multi-Source Data
by Mr. Tod HAGAN, Securboration, USA

12:15 **BUFFET LUNCH**

SESSION 1 – SEMANTIC MULTIMEDIA ANALYSIS
Chair – Dr. Lasse OVERLIER, NOR

13:15 **1** Ad-Hoc Search in Video Data
by Ms. Maaike de BOER, TNO, NLD

13:45 **2** Compressed-Domain Deep Learning for Semantic Video Analytics
by Dr. Ioannis ANDREOPOULOS et al, UCL, GBR

14:15 **3** Combination of Video and Text Analytics for Multi-Source Intelligence
by Dr. Gertjan BURGHOUTS, TNO, NLD

14:45 **BREAK and TOUR** of Visual Analytics Lab. and VALCRI Project
Escorted by Drs. William WONG and Neesha KODAGODA

16:00 **BREAKOUT SESSION 1**
Establish Groups and ways of working

16:30 Feedback from Groups

17:30 – 20:00 **DRINKS RECEPTION & FINGER FOOD**

Thursday 7 September 2017

BUILDING: TOWN HALL; ROOM: COUNCIL CHAMBERS

09:00 **ARRIVALS**

SESSION 2 – IMAGERY EXPLOITATION
Chair – Dr. Gertjan BURGHOUTS, NLD

09:30 **KEYNOTE SPEECH 2:**
Forgery Detection in Images
by Dr. Katrin FRANKE, NTNI, NOR

10:30 **BREAK**

11:00 **4** A Deep Convolutional Network for Traffic Congestion Classification
by Dr. Chris WILLIS et al., BAE Systems, GBR

11:30 **5** Collaborative Autonomy Using Deep Learning for Emergency Scene Assessment and Response
by Dr. Stefano CAVAZZI et al., ENVITIA, GBR

12:00 **LUNCH**

SESSION 3 – EMERGING ISSUES
Chair – Dr. Liz BOWMAN, USA

13:00 **6** Exploring the Value of Target Motion Dependency
by Drs. Simon JULIER, UCL, GBR and Raghuveer RAO, ARL, USA

13:30 **7** A Collaborative Experiment for Image Indexing
by Dr. Lasse OVERLIER, FFI, NOR

14:00 **8** Intelligent Outreach to AI Capabilities
by Mr. Phil GIBSON et al., DSTL, GBR

14:30 **BREAK**

15:00 **BREAKOUT SESSION 2**
Detailed group work to address questions regarding key areas of research for NATO IST

16:30 Feedback from Groups

17:00 Conclusions and Summary (non-committee members can leave if not interested in contributing to Day 3)

Friday 8 September 2017

BUILDING: TOWN HALL; ROOM: COUNCIL CHAMBERS

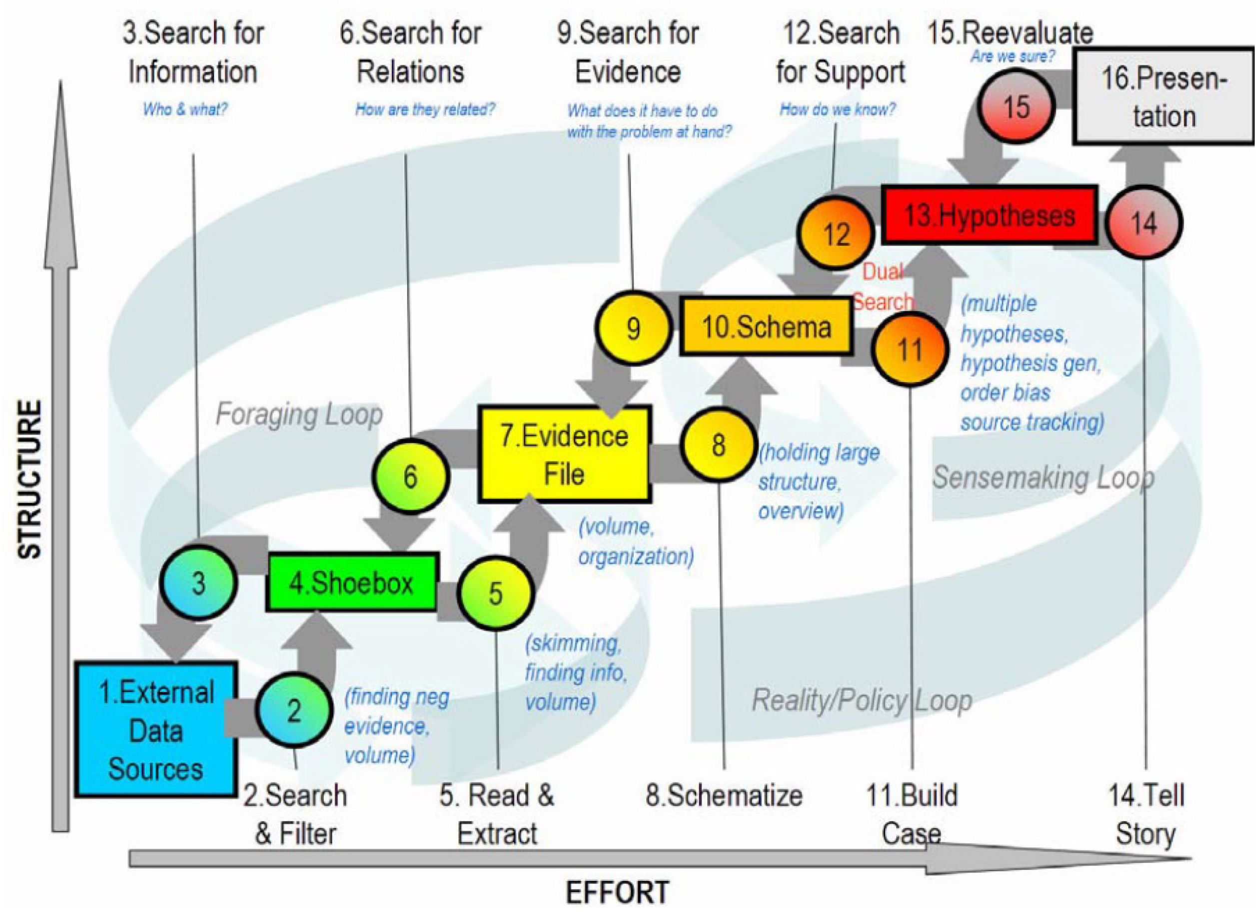
09:00 **ARRIVALS**

SESSION 4 – OUTCOMES AND OUTPUTS
Chair – Prof. Bob MADAHAR, GBR

09:30 Discussion on findings from the RSM and working groups
Report structure and writing initial drafts
by Committee Members and those interested such as breakout group leads and others with research interests

12:30 **CLOSE and LUNCH in Canteen Not Provided**

Annex B – PIROLLI AND CARD: NOTIONAL MODEL SENSE-MAKING LOOP FOR INTELLIGENCE ANALYSIS [6]



REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	STO-TR-IST-144 AC/323(IST-144)TP/981	ISBN 978-92-837-2301-1	PUBLIC RELEASE

5. Originator

Science and Technology Organization
North Atlantic Treaty Organization
BP 25, F-92201 Neuilly-sur-Seine Cedex, France

6. Title

Content Based Multi-Media Analytics (CBMA)

7. Presented at/Sponsored by

Summary Report of the NATO STO IST-144 Task Group.

8. Author(s)/Editor(s)	9. Date
Multiple	October 2020

10. Author's/Editor's Address	11. Pages
Multiple	52

12. Distribution Statement

There are no restrictions on the distribution of this document. Information about the availability of this and other STO unclassified publications is given on the back cover.

13. Keywords/Descriptors

Analytics architectures; Artificial intelligence; Decision making; Event prediction; Image processing; Information environment; Knowledge generation; Machine learning; Machine vision; Multimedia analytics; Semantic analysis; Sensemaking; Social computing; Text analytics; Video analytics

14. Abstract

This report provides a summary of the research developments of IST-144-RTG on Content Based Multi-Media Analytics (CBA) carried out by a team from NLD, NOR, USA, and UK. These scientists have brought together information retrieval strategies from heterogeneous media sources (text, video and images) and human assessment. As a result, multiple heterogeneous data sources can be exploited by content based information retrieval and multimedia analytics to deliver timely and accurate synopses of data with information that can be combined with human intuition and understanding to develop a comprehensive 'view' of the problem/solution space. Such interoperable tools that cross-cue knowledge obtained from one method to generate taskings in another are needed by NATO Coalition military leaders, commanders, and intelligence analysts to accelerate situational awareness and decision making and deal with the complexity of the defence information space. A description of the core technical components is provided along with the description of their combined application in a concept demonstrator addressing a fictitious, but nevertheless realistic, scenario representative of defence challenges facing the alliance. The report is concluded with the findings and recommendations for further work that is needed to address limitations, including technical and system level gaps, such as an open distributed architecture for integration of analytics services, now and expected in the future.

NORTH ATLANTIC TREATY ORGANIZATION

NATO OTAN

BP 25

F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE

Télécopie 0(1)55.61.22.99 • E-mail mailbox@cso.nato.int

SCIENCE AND TECHNOLOGY ORGANIZATION

S&T organization

DIFFUSION DES PUBLICATIONS

STO NON CLASSIFIEES

Les publications de l'AGARD, de la RTO et de la STO peuvent parfois être obtenues auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la STO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus soit à titre personnel, soit au nom de votre organisation, sur la liste d'envoi.

Les publications de la STO, de la RTO et de l'AGARD sont également en vente auprès des agences de vente indiquées ci-dessous.

Les demandes de documents STO, RTO ou AGARD doivent comporter la dénomination « STO », « RTO » ou « AGARD » selon le cas, suivi du numéro de série. Des informations analogues, telles que le titre est la date de publication sont souhaitables.

Si vous souhaitez recevoir une notification électronique de la disponibilité des rapports de la STO au fur et à mesure de leur publication, vous pouvez consulter notre site Web (http://www.sto.nato.int/) et vous abonner à ce service.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Gorch-Fock-Straße 7, D-53229 Bonn

BELGIQUE
Royal High Institute for Defence – KHID/IRSD/RHID
Management of Scientific & Technological Research for Defence, National STO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30, 1000 Bruxelles

BULGARIE
Ministry of Defence
Defence Institute "Prof. Tsvetan Lazarov"
"Tsvetan Lazarov" bul no.2
1592 Sofia

CANADA
DGSIST 2
Recherche et développement pour la défense Canada
60 Moodie Drive (7N-1-F20)
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Acquisition and Logistics Organization (DALO)
Lautrupbjerg 1-5
2750 Ballerup

ESPAGNE
Área de Cooperación Internacional en I+D
SDGPLATIN (DGAM)
C/ Arturo Soria 289
28033 Madrid

ESTONIE
Estonian National Defence College
Centre for Applied Research
Riia str 12
Tartu 51013

ETATS-UNIS
Defense Technical Information Center
8725 John J. Kingman Road
Fort Belvoir, VA 22060-6218

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72
92322 Châtillon Cedex

GRECE (Correspondant)
Defence Industry & Research General Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HONGRIE
Hungarian Ministry of Defence
Development and Logistics Agency
P.O.B. 25
H-1885 Budapest

ITALIE
Ten Col Renato NARO
Capo servizio Gestione della Conoscenza
F. Baracca Military Airport "Comparto A"
Via di Centocelle, 301
00175, Rome

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

PAYS-BAS
Royal Netherlands Military Academy Library
P.O. Box 90.002
4800 PA Breda

POLOGNE
Centralna Biblioteka Wojskowa
ul. Ostrobramska 109
04-041 Warszawa

PORTUGAL
Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide
P-2720 Amadora

REPUBLIQUE TCHEQUE
Vojenský technický ústav s.p.
CZ Distribution Information Centre
Mladoboleslavská 944
PO Box 18
197 06 Praha 9

ROUMANIE
Romanian National Distribution Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6
061353 Bucharest

ROYAUME-UNI
Dstl Records Centre
Rm G02, ISAT F, Building 5
Dstl Porton Down
Salisbury SP4 0JQ

SLOVAQUIE
Akadémia ozbrojených síl gen. M.R. Štefánika, Distribučné a informačné stredisko STO
Demänová 393
031 01 Liptovský Mikuláš 1

SLOVENIE
Ministry of Defence
Central Registry for EU & NATO
Vojkova 55
1000 Ljubljana

TURQUIE
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanliklar – Ankara

AGENCES DE VENTE

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
ROYAUME-UNI

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa, Ontario K1A 0S2
CANADA

Les demandes de documents STO, RTO ou AGARD doivent comporter la dénomination « STO », « RTO » ou « AGARD » selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications STO, RTO et AGARD figurent dans le « NTIS Publications Database » (http://www.ntis.gov).

NORTH ATLANTIC TREATY ORGANIZATION

NATO OTAN

BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@cso.nato.int

SCIENCE AND TECHNOLOGY ORGANIZATION

DISTRIBUTION OF UNCLASSIFIED STO PUBLICATIONS

AGARD, RTO & STO publications are sometimes available from the National Distribution Centres listed below. If you wish to receive all STO reports, or just those relating to one or more specific STO Panels, they may be willing to include you (or your Organisation) in their distribution.

STO, RTO and AGARD reports may also be purchased from the Sales Agencies listed below.

Requests for STO, RTO or AGARD documents should include the word 'STO', 'RTO' or 'AGARD', as appropriate, followed by the serial number. Collateral information such as title and publication date is desirable.

If you wish to receive electronic notification of STO reports as they are published, please visit our website (http://www.sto.nato.int/) from where you can register for this service.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Royal High Institute for Defence – KHID/IRSD/RHID
Management of Scientific & Technological Research for Defence, National STO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30
1000 Brussels

BULGARIA
Ministry of Defence
Defence Institute "Prof. Tsvetan Lazarov"
"Tsvetan Lazarov" bul no.2
1592 Sofia

CANADA
DSTKIM 2
Defence Research and Development Canada
60 Moodie Drive (7N-1-F20)
Ottawa, Ontario K1A 0K2

CZECH REPUBLIC
Vojenský technický ústav s.p.
CZ Distribution Information Centre
Mladoboleslavská 944
PO Box 18
197 06 Praha 9

DENMARK
Danish Acquisition and Logistics Organization (DALO)
Lautrupbjerg 1-5
2750 Ballerup

ESTONIA
Estonian National Defence College
Centre for Applied Research
Riia str 12
Tartu 51013

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc – BP 72
92322 Châtillon Cedex

GERMANY
Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Gorch-Fock-Straße 7
D-53229 Bonn

GREECE (Point of Contact)
Defence Industry & Research General Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HUNGARY
Hungarian Ministry of Defence
Development and Logistics Agency
P.O.B. 25
H-1885 Budapest

ITALY
Ten Col Renato NARO
Capo servizio Gestione della Conoscenza
F. Baracca Military Airport "Comparto A"
Via di Centocelle, 301
00175, Rome

LUXEMBOURG
See Belgium

NETHERLANDS
Royal Netherlands Military Academy Library
P.O. Box 90.002
4800 PA Breda

NORWAY
Norwegian Defence Research Establishment, Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

POLAND
Centralna Biblioteka Wojskowa
ul. Ostrobramska 109
04-041 Warszawa

PORTUGAL
Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide
P-2720 Amadora

ROMANIA
Romanian National Distribution Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6
061353 Bucharest

SLOVAKIA
Akadémia ozbrojených síl gen M.R. Štefánika, Distribučné a informačné stredisko STO
Demänová 393
031 01 Liptovský Mikuláš 1

SLOVENIA
Ministry of Defence
Central Registry for EU & NATO
Vojkova 55
1000 Ljubljana

SPAIN
Área de Cooperación Internacional en I+D
SDGPLATIN (DGAM)
C/ Arturo Soria 289
28033 Madrid

TURKEY
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanliklar – Ankara

UNITED KINGDOM
Dstl Records Centre
Rm G02, ISAT F, Building 5
Dstl Porton Down, Salisbury SP4 0JQ

UNITED STATES
Defense Technical Information Center
8725 John J. Kingman Road
Fort Belvoir, VA 22060-6218

SALES AGENCIES

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
UNITED KINGDOM

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa, Ontario K1A 0S2
CANADA

Requests for STO, RTO or AGARD documents should include the word 'STO', 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of STO, RTO and AGARD publications are given in "NTIS Publications Database" (http://www.ntis.gov).

ISBN 978-92-837-2301-1