Eszter Katalin Bognár [1]

Novel IT Technologies on the Digital Battlefield: The Application of Big Data and Data Mining Technologies[2]

Korszerű eszközök a digitális harcmezőn: Big Data és adatbányászati technológiák alkalmazása

In modern warfare, the most important innovation to date has been the utilisation of information as a weapon. The basis of successful military operations is the ability to correctly assess a situation based on credible collected information. In today's military, the primary challenge is not the actual collection of data. It has become more important to extract relevant information from that data. This requirement cannot be successfully completed without necessary improvements in tools and techniques to support the acquisition and analysis of data. This study defines Big Data and its concept as applied to military reconnaissance, focusing on the processing of imagery and textual data, bringing to light modern data processing and analytics methods that enable effective processing.

Keywords: Big Data, data analytics, digital image processing, text mining

A modern hadviselés kapcsán leginkább szembetűnő változás az információ mint fegyver megjelenése. A katonai műveletek alapját a megszerzett információ révén elérhető helyzetértékelési képesség adja. Manapság a katonai műveletek során nem elsősorban az adatok megszerzése jelenti a fő kihívást, sokkal fontosabbá vált az adatokból a parancsnoki döntéshozatal számára hasznos és releváns információ kinyerése. Ez az igény csak az adatok megszerzésére és elemzésére szolgáló technológiák fejlesztése révén érhető el. A tanulmány ismerteti a Big Data definícióját

1   University of Public Service, Doctoral School of Military Engineering, PhD student, e-mail: bognarek@uni-nke.hu, ORCID: https://orcid.org/0000-0002-3697-7871

és jellemzőit a katonai felderítési adatok vonatkozásában, valamint feltárja a képi és szöveges adatok elemzésére alkalmazható technológiákat.

Kulcsszavak: **Big Data, adatelemzés, digitális képfeldolgozás, szövegbányászat**

1. Introduction

In modern warfare, an obvious change is the emergence of information as a weapon. The basis of military operations is the ability to assess the situation, made possible by relevant collected information. Nowadays, there is an increasing number of military data that can be easily collected, but the real challenge is to extract the useful information from the collected data, which can be converted to helpful reconnaissance data, a necessary knowledge for command decision making.[3] While half a decade ago the time lapse for processing data and then determining the target was as much as two days, command decision-making today requires real-time processing, while the volume of reconnaissance data has grown exponentially (see Figure 1).

Since each activity is organised around collected information, the defensive sphere is paying particular attention to the newest technologies which help to develop information processing, especially regarding the acquisition of information and its procession more efficiently and quickly. The computerised command and control function supported by reconnaissance makes it possible to collect differently structured high volumes of data, but at the same time makes it very difficult to process it in an efficient and expeditious manner. A modern system based on data analytics can integrate the data provided by various reconnaissance data sources. Current data-mining technology helps manage information overload. The recognition of the different patterns of data, and diverse data reduction and filtering methods would produce timely, wide-ranging, and accurate reconnaissance information. The solution to the problem can be through modern Big Data technologies, data-mining algorithms, and machine learning.

From an operational point of view, it has become more important for the intelligence-gathering element to process massive datasets, identify specific signatures, and take appropriate action. It must be done in seconds rather than hours or days, since it requires continuous rapid data analysis. This requirement cannot be accomplished without necessary improvements in tools and techniques to support the acquisition and analysis of data. To find necessary information in a flood of real-time data is crucial. At the same time, collection, storage, and processing must conform to applicable laws and regulations protecting the rights of citizens.

The study defines 'Big Data' as applied to military reconnaissance data and focuses on the processing of imagery and textual data, revealing the modern data processing and analytics methods which help process them effectively.

---

[3] Zs. Haig, *Információs műveletek a kibertérben* (Budapest: Dialóg Campus, 2018). 346 p.
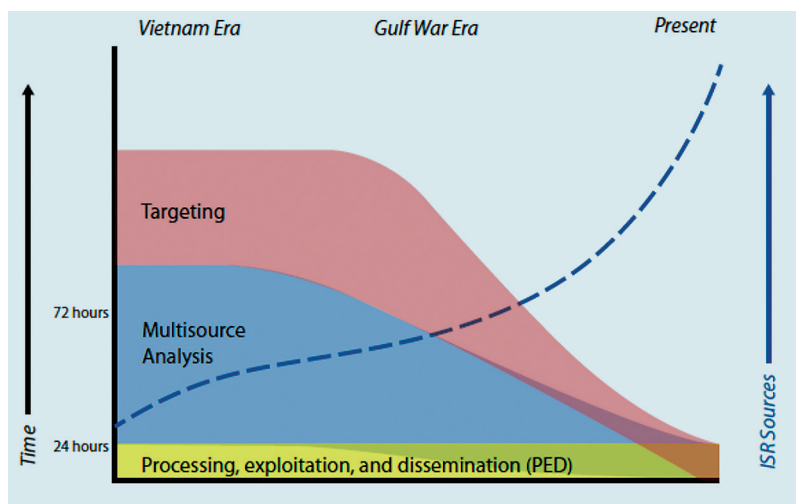
Figure 1

The impact of information revolution on collection, analysis and targeting

Source: P. S. Hamilton and P. M. Kreuzer, 'The Big Data Imperative. Air Force Intelligence for the Information Age,' *Air & Space Power Journal* 32, no 1 (2018), 4–20.

2. Big Data and data analytics

Today the term 'Big Data' receives much attention. The growth of Big Data is the result of the ever-increasing channels and variety of data in today's world. There are many definitions of Big Data, but the easiest way to imagine it is an extremely huge amount of data from various sources in an unstructured format which cannot be easily managed (store, search, share, visualise, and analyse) with traditional technologies.

According to most definitions: Big Data is a large pool of data that can be captured, communicated, aggregated, stored, and analysed, but for these we need novel technologies instead of the traditional ones.[4]

The characteristics of Big Data are usually described by the so-called four Vs: volume, velocity, variety, and value.[5] Volume and velocity can be defined as data growing significantly in volume with extremely high speed. Variety means that the generated data comes from multiple sources, thus it is in various and diverse formats and structures. It causes considerable problems, because the unstructured data is difficult to process with traditional relational databases. This issue has brought about new database technologies like NoSQL databases. Value means that there is a vast amount of data, but only a small part of it has any value useable to make informed business decisions.

---

[4] *National and Transnational Security Implications of Big Data in the Life Sciences*, American Association for the Advancement of Science, 2014.

[5] 'Big Data for defence and security,' Royal United Services Institute, ocassional paper, 2013.

In the area of military surveillance, Big Data refers to the following types of data:
- machine-generated /sensor data – data collected by different types of military sensors (imagery, seismic, magnetic, infrared and so on). Today, machine data is generated by the movement of ships, aircraft and vehicles, satellites in space, drones, unmanned aerial vehicles (UAVs), reconnaissance aircraft, sensors, and battlefield surveillance radars (BFSR);
- open source intelligence (OSINT) data – data from any publicly-available data source: online news, posts on micro-blogging sites like Twitter and social media platforms like Facebook.

The application of Big Data tools helps to gain big advantages in national defence. For example, a medium-altitude long-endurance system such as the MQ-9 Reaper in the UK can collect ca. 20 laptops' worth of data per sortie.[6] The US Argus autonomous real-time ground ubiquitous surveillance imaging system collects more than 40 GB of information per second. It has a 1.8 gigabyte pixel video camera taking 12 frames per second (fps) and touts 368 sensors. It collects 6,000 terabytes of data/imagery per day and feeds it to Homeland Security. In the armed forces it became a burning issue how to interpret and derive value from the large volumes and rapid flow of real-time data within their operations.[7] Telemetry, surveillance, and military sensors bring the internet of things to the battlefield as well as operational areas.[8]

Big Data analytic technologies are a fast-developing field and are crucial components in dealing with the information overload they face. In these usage areas, security and real-time processing of data are extremely important. That is why the military is keenly aware of the importance of the different Big Data issues and are working to discover and implement these new technologies in processes.

2.1. Infrastructure Requirements

The Big Data processing pipeline usually consists of three main steps: acquisition, organisation, and analysis.[9] At each step, different technologies must be utilised to maximise the speed and effectiveness of the data being processed. I will examine these three steps and explore the latest technologies and frameworks to be used to get the most out of the data, together with the required security and privacy considerations.

Data acquisition refers to the collection and storage of data. Since Big Data usually comes from different sources and in various formats, it is important to handle the acquisition process efficiently. The environment in which Big Data is generated and collected is dynamically changing and has a major impact on the data structure.

---

6   'Big Data for defence and security.'
7   M. Haridas, 'Redefining military intelligence using big data,' *Scholar Warrior*, Autumn 2015, 72–78.
8   F. Loaiza, J. Shah and R. Rolfe, 'Real-Time Information Extraction from Big Data,' Institute for Defence Analysis, Virginia, 2015.
9   Bao En ME4 Toh, 'Swimming in Sensors, Drowning in Data – Big Data Analytics for Military Intelligence,' *Pointer, Journal of The Singapore Armed Forces* 42, no 1 (2016), 51–65.

In addition, in most of the cases, particularly in case of military surveillance data, it must be collected and processed with the least possible delay, preferably in real time.

Instead of traditional relational database management systems for the acquisition of Big Data, the use of distributed NoSQL (non-SQL) databases are the most suitable method. NoSQL databases do not have any predefined structure; thus, they are perfect for the collection and storage of the unstructured data. In comparison with traditional relational database systems, most NoSQL databases store data in simple key-value pairs, column, graph or document-based data structures so there is no need to precisely identify the connection between data attributes; it simply captures all data without categorising and parsing it into fixed schema. This simple and dynamic structure allows changes to take place without costly reorganisation in the storage layer. The simplicity of design, flexibility, scalability and application of distributed storage and file systems such as HDFS[10] (Hadoop Distributed File Systems) make these databases more suitable for storing Big Data than traditional database management systems. There are many different types of NoSQL databases to choose from; one of the most popular implementations is MongoDB.[11]

Moving forward, organisation of Big Data refers to data integration. Integration of data involves data preprocessing: filtering, transforming, and sorting of data that comes from various sources and in various formats and structures to achieve the final integrated, consistent, and structured input dataset for further analysis. It is desirable to organise data at its original storage location to avoid moving large volumes of data in and out of the storage system. Apache Hadoop[12] is a free Java-based programming framework that enables the collection, storage, and organisation of data in a distributed computing environment.

Another useful tool from Apache is MapReduce,[13] a software framework for processing vast amounts of data in parallel across a distributed cluster of processors or stand-alone computers to speed up execution of operations on data.

The last step in Big Data processing is data analysis and visualisation. Analysis is used to obtain insights into the data, discover primary patterns and create value from the data set. After discerning hidden correlations between records, it can be used to make more accurate decisions.

The analysis of Big Data must be done within the distributed environments, preferable at the location of the data (in-database analytics) and the tools must allow deeper analysis of data using statistical methods, data mining, and so on. The biggest challenge is to speed up processing time. Statistical tools like Python[14], R[15] and Tableau[16] for visualisation can be integrated with the above-mentioned Big Data solutions.[17]

---

[10] For HDFS see https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html (30. 05. 2020).
[11] For MongoDB see www.mongodb.com/ (30. 05. 2020).
[12] For Apache Hadoop see https://hadoop.apache.org/ (30. 05. 2020).
[13] For Apache MapReduce see https://hadoop.apache.org/docs/r2.8.0/hadoop-mapreduce-client/ (30. 05. 2020).
[14] For Python see www.python.org/ (30. 05. 2020).
[15] For R see www.r-project.org/ (30. 05. 2020).
[16] For Tableau see www.tableau.com/ (30. 05. 2020).
[17] 'Big Data for the Enterprise,' Oracle White Paper, June 2013.

2.2. Data security and privacy concerns

In the case of Big Data, the diversity of data sources and the special infrastructural requirements imply new security and privacy challenges. Security and privacy must be seen from different perspectives according to the dynamically changing environments in which data is collected, stored, processed, and analysed.

According to the Cloud Security Alliance,[18] the security challenges of Big Data are divided into four categories. The top ten security and privacy issues for Big Data processing according to the work of Duygu Sinanc Terzi, Ramazan Terzi and Seref Sagiroglu are listed below:[19]

- infrastructure security: secure computing in distributed programming frameworks and security practices in nonrelational data stores;
- data privacy: privacy-preserving analytics, cryptographically enforced data-centric security and granular access control;
- data management: secure data storage and transaction logs, auditing, and data provenance;
- data integrity and reactive security: end-point validation and filtering, real time monitoring.

During data acquisition, proper input validation and filtering must be applied. In the case of storage, nonrelational databases were designed for superior performance and scalability prioritised over proper security standards. Thus, nonrelational databases lack many of the security features of traditional database management systems, causing much critical security flaws in NoSQL. To overcome them, the best practices for nonrelational data stores must be utilised, such us data encryption, better control between clusters, security policies for middleware, and so on. By organising and preprocessing the data, secure computing must be applied; while in the analysis stage, privacy-preserving data mining (for example anonymised datasets) should be used to preserve the privacy of individuals.

Beside these, proper data management must be ensured during all stages through granular auditing, access control, real-time security monitoring and data provenance.

In this chapter, the definition and main characteristics were introduced together with the most recent technologies that can be used in the different stages of processing Big Data. Based on the above, we can see that Big Data and associated technologies play an important role in today's connected world. To handle vast amounts of unstructured data, the appropriate IT technologies must be utilised with consideration of the increased security and privacy scenarios. Figure 2 shows the Big Data infrastructure, together with proposed security and privacy solutions.

---

[18] 'Expanded Top Ten Big Data Security and Privacy Challenges,' Cloud Security Alliance, April 2013.
[19] Duygu Sinanc Terzi, Ramazan Terzi and Seref Sagiroglu, 'A survey on security and privacy issues in big data,' *Proceedings of the 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST)*, London, 2015. 202–207.
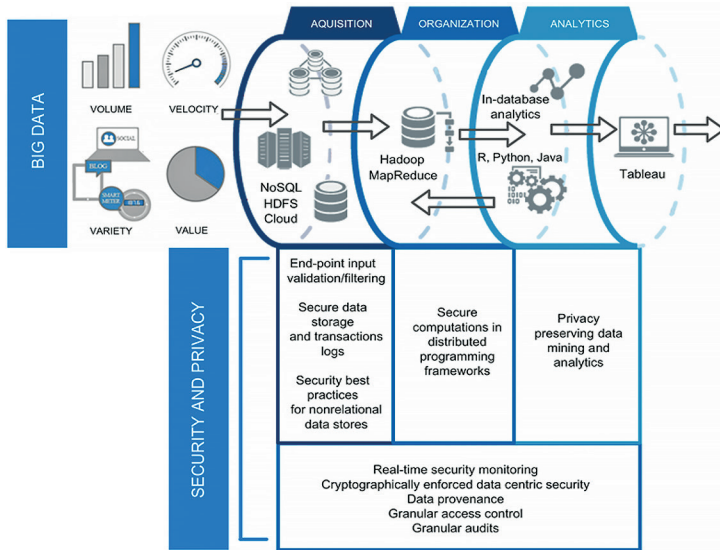
Figure 2
Big Data infrastructure with security and privacy recommendations
Source: edited by the author

3. Data mining for military intelligence analysis

In today's information-saturated battlespace, Big Data represents a process for rapidly compiling, storing and accessing large amounts of data and information from numerous sources using varying structures.

With an ever-growing reliance on network-centric operations, governments have payed significantly more attention to improving their ability to collect and analyse intelligence data. Big Data analytics represent the tools and processes that can transform Big Data into insights; from intelligence preparation of the operational environment to threat warning, predictive battlespace awareness or targeting. These insights in turn shape decision-making across the range of military and diplomatic operations, from strategic deterrence operations to near real-time tactical engagements.[20]

Different intelligence sources (the INTs – signals intelligence [SIGINT], geospatial intelligence [GEOINT], imagery intelligence [IMINT], human intelligence [HUMINT], open-source intelligence [OSINT], and measurement and signals intelligence [MASINT]) enable the delegation of effort into separate data issues that could be analysed individually in their parts by specialists, with all source intelligence answers produced by combining component parts. The conceptual layout of a Big Data application-based intelligence gathering system can be seen in Figure 3.

---

[20] R. D. Thiele, 'Mit Daten siegen – Big Data verändert Wirtschaft und Streitkräfte,' ISPSW Strategy Series: Focus on Defense and International Security, No. 393 (2015), 1–7.

Most of the intelligence data comes from IMINT and OSINT. The next part of this chapter will explore the most important technologies of image and text data mining.
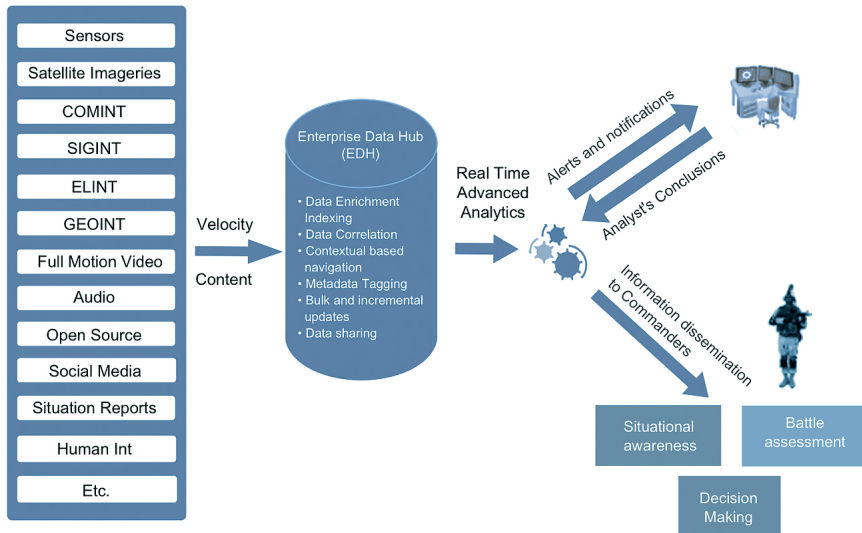

Figure 3
Conceptual layout of Big Data application-based intelligence gathering system
Source: Haridas, '*Redefining*', 76.

3.1. Processing imagery intelligence (IMINT) data

Image data are collected by high resolution satellites, large manned aircraft and by smaller unmanned platforms, tiny handheld devices and unattended ground sensors.

Today, the high-definition digital electro-optic and infrared sensors are creating more and more data on the battlefield and these devices can provide more detailed imagery at longer ranges than was possible before.

How can soldiers get that information quickly enough to act on it in a timely manner? The urgent need for clear, actionable imagery intelligence demanded the application of real-time digital image processing techniques on the battlefield, so fighters could obtain the situational awareness they need to protect themselves and to act decisively against threats within tactical timelines.

3.1.1. Digital image processing

The images collected by different platforms may be distorted as a result of the sensor's motion, vibration from the engine carrying the device, flight maneuvers and

turbulence. To get the most accurate result, the target must be consistently and accurately tracked.

Today, analogue imagery sensors have mostly been replaced by high-resolution digital devices. A big advantage of digital imagery is that it does not have to be converted and can be immediately processed by the computer to enhance image quality. The capabilities of visual surveillance can be extended and enhanced in challenging lighting and weather conditions by either improving the hardware, primarily the camera, or digitally processing the resulting images to enhance their quality. Sophisticated processing is still necessary to collect, stabilise, track and compress this digital video imagery so that it can be transmitted to users without overwhelming data links.[21]

In this chapter, I will discover the most often used image processing techniques, starting with the different preprocessing techniques such as filtering, stabilisation and so on, to enhance image quality, and the most sophisticated techniques to interpret the images and get the most accurate result of the actual battlefield situation.

Without attempting to be comprehensive, this chapter focuses on the most important digital image processing techniques without discussing the algorithms themselves in depth.

Image enhancement techniques

The goal of image enhancement techniques is to process images so that the result is more suitable than the originals for a specific application. This can dramatically improve visual appearance and gain the most useful information from collected images by analysing the preprocessed, enhanced images.

In particular, the following image enhancement techniques can be used:[22]

- Histogram equalisation: the process of adjusting the intensity values of the image to enhance the images' contrast. Occasionally, images contain disproportionate values in the dark or bright ranges. By stretching the range of intensity values, image quality can be significantly improved. Figure 4 shows enhancement options of an infrared image using histogram equalisation methods.
- Filtering and morphological operations: filtering is a technique for modifying or enhancing an image. Different filter masks can be used to emphasise certain features or remove other features in the images. Image processing operations implemented with filtering include smoothing, sharpening, and edge enhancement. Morphological operations include erosion, dilation and opening process images based on shapes. In a morphological operation, each pixel in the image is adjusted based on the value of other pixels in its neighborhood.
- Distortion correction: digitally eliminating image distortion caused by analogue imaging optics.
- Stabilisation: the removal of random camera shaking from video footage.
- Mosaicking: the process of combining images to get a larger field of view.

---

21 V. V. D. Shah, 'Image Processing and its Military Applications,' *Defence Science Journal* 37, no 4 (1987), 457–468.
22 'Use real-time HD image processing for military and civilian surveillance,' EDN, 2013.

- Fusion: combining images from different and complementary wavebands, using image fusion techniques, to yield a single composite image that maximises relevant information.
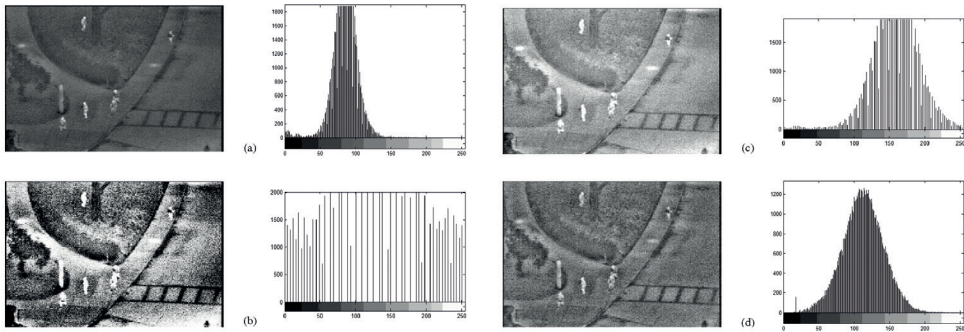

Figure 4

Enhancement of infrared image by applying different histogram equalisation algorithms

Source: S. Erturk, 'Improved Region of Interest for Infrared Images Using Rayleigh Contrast-Limited Adaptive Histogram Equalization,' 2013.

3.1.2. Sophisticated image processing techniques

With advancements in computer storage capacity and parallel processing, Big Data has become omnipresent. Related to Big Data, the application of artificial intelligence (AI), particularly different machine learning (ML) algorithms, has become very popular performing sophisticated analyses on images to better assess activity on the depicted scene.

After preprocessing images applying basic digital image enhancement techniques, these algorithms can be used to recognise and track objects, the primary purpose of military surveillance. To achieve comprehensive situational awareness, actions and interactions, as well as a given series of actions and interactions (pattern of life), can be also identified.

The most important techniques according to the work of Dijk et alii are:[23]

- Object detection: object detection is the process of finding real-world objects such as people, vehicles and buildings in images or videos.
- Tracking: tracking keeps the target within the center of the sensor's field of view by sending steering commands to the gimbal, based on target information derived from the sensor data, using object recognition algorithms. Objects can be tracked throughout the video sequence by making use of their position,

---

[23] Judith Dijk, Adam W. M. van Eekeren, Olga Rajadell Rojas, Gertjan J. Burghouts and Klamer Schutte, 'Image processing in aerial surveillance and reconnaissance: From pixels to understanding,' Proceedings of SPIE (The International Society for Optical Engineering), Electro-Optical Remote Sensing, Photonic Technologies, and Applications VII; and Military Applications in Hyperspectral Imaging and High Spatial Resolution Sensing, 88970A (15 October 2013).

movement, and possible appearance. Tracking will provide an association of multiple detections belonging to a single entity and will designate the path that the entity has followed in the scene.

- Change detection: its goal is to identify changes in multi-temporal data sets. It is commonly used for identifying changes in surveillance data such as the appearance or disappearance of vehicles and individual persons.
- Activity and interaction recognition: recognising human actions and interactions during surveillance. Figure 5 shows an example of activity recognition in images.
- Pattern of life: action and interaction detection data can be used for event detection and for providing situational awareness. Detected objects and actions are used to see a certain event and identify habits to better assess what is occurring. Pattern-of-life analysis works on the assumption that if behavior patterns of, for example, a population, town or street are continuously observed, regular patterns can be identified and deviations (or anomalies) to this pattern can also be detected. The challenge for airborne systems is that the period of time they can observe a certain area is limited, which makes it challenging to determine the normal pattern.
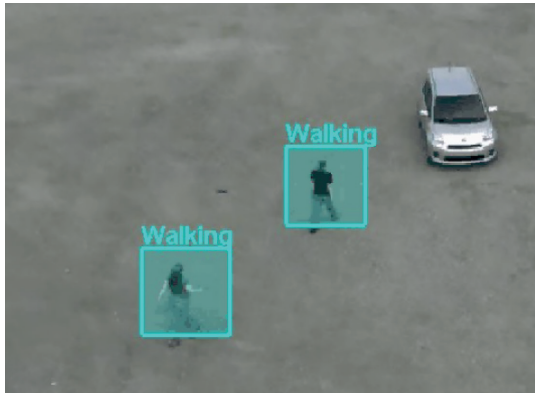

Figure 5
Activity recognition in images
Source: Dijk, Eekeren, Rojas, Burghouts and Schutte, 'Image processing.'

3.2. Processing open source intelligence (OSINT) data

OSINT is a new intelligence discipline that appeared as the result of the proliferation of the internet and social media. Today, the number of publicly available data sources like high resolution Google Earth satellite imagery of nearly the entire planet, Google Street View covering many countries, people's posts on Twitter and other social media platforms allow endless possibilities for data collection.

The information revolution has led to a new online culture of sharing. Through Twitter, Facebook, Snapchat, blogs, and numerous social media sites, intelligence has access to tens of millions of passive collectors all over the world.

In this chapter, the analytics of open source information will be introduced, focusing on the algorithms used for processing textual data.

Information in news/documents related to a specific person or topic – for example, new articles, publications, white papers, updates on social media – comprise inputs that will be of great help in planning intelligence strategy at a higher level.

Prime topics/concepts being discussed in social media can be monitored and studied specifically, by geography, persons, organisations and so on. Analytics of information sources – for example, affinity of information sources to a specific user group, geography and so on – will have great intelligence value. The sentiments of people regarding a policy or concept can be known and proactive actions taken, as required. Social media graphs identifying users' groups active on a website can be analysed.

3.2.1. Text mining

Text mining is the process of deriving high-quality information or actionable knowledge from textual data while minimising human effort. Text mining usually involves the process of structuring the input text (usually parsing, along with the addition of some derived linguistic features and the removal of others, and subsequent insertion into a database), deriving patterns within the structured data, and finally, evaluation and interpretation of the output. Typical text mining tasks include text categorisation, text clustering, concept/entity extraction, production of granular taxonomies, sentiment analysis, document summarisation, and entity relation modeling (that is, learning relations between named entities). There has been much research done in this field; to mention some, Parekh, Amarasingam, Dawson and Ruths[24] identified terrorist groups on Twitter, Harb and Becker[25] performed sentiment analysis regarding terrorism using tweets, and Vasileios[26] identified sentiments towards refugees.

Text mining involves information retrieval, lexical analysis to study word frequency distributions, pattern recognition, tagging/annotation, information extraction, data mining techniques that include link and association analysis, visualisation, and predictive analytics. The overarching goal is, essentially, to turn text into data for analysis, via application of natural language processing (NLP) and analytical methods.

---

[24] Deven Parekh, Amarnath Amarasingam, Lorne Dawson and Derek Ruths, 'Studying Jihadists on Social Media: A Critique of Data Collection Methodologies,' *Perspectives on Terrorism* 12, no 3 (2018), 3–21.

[25] Johathas D. G. Harb and Karin Becker, 'Emotion analysis of reaction to Terrorism on Twitter,' Proceedings of the SBC 33rd Brazilian Symposium on Databases, 2018, Rio de Janeiro, Brazil, 97–108.

[26] L. Vasileios, *Comparative analysis of the hashtags #RefugeesWelcome and #StopRefuges*. Master thesis, Korinth, 2017. 62.

Without attempting to be comprehensive, the most important text preprocessing and analysing techniques will be covered based on the book of D. Tikk.[27]

3.2.2. Text preprocessing techniques

- Text filtering: To normalise text for further analysis, some filtering should be applied. Punctuation and whitespaces should be removed. In most cases the removal of numbers is advisable, too, if they are not relevant to the analyses. Usually, regular expressions are used to remove unnecessary characters and numbers. Mostly, removal of the most common words in a language like 'the', 'a', 'on', 'is', 'all' is also performed, since these words do not carry important meaning and are usually removed from texts. That is called stop-word filtering.
- Tokenisation: Tokenisation is the process of splitting the given text into smaller segments, called tokens. Words, numbers, punctuation marks, and others can be considered as tokens.
- Stemming: Stemming is the process of reducing words to their word stem, base, or root form (for example, books – book, looked – look). The most common algorithm is the Porter Stemming Algorithm, but there are special stemming algorithms for different languages, for instance, for the Hungarian language, the Snowball Stemmer.
- Part of speech (POS) tagging: Part-of-speech tagging aims to assign parts of speech to each word of a given text (such as nouns, verbs, adjectives, and others) based on its definition and its context.
- Named entity recognition (NER): Named-entity recognition aims to find named entities in text and classify them into pre-defined categories (names of persons, locations, organisations, times and so on). The identified entities should be correlated with other sources of data, such as images of a building at a given location.
- Computation of term frequency–inverse document frequency (tf–idf): it is a numerical statistic that is intended to reflect how important a word is to a document in a corpus (collection of documents). The tf–idf value increases proportionally to the number of times a word appears in the document and is offset by the number of documents in the corpus that contains the word, which helps to adjust for the fact that some words appear more frequently in general. Term frequencies can be used for example for word cloud creation using the most frequent terms or for classifying documents.

3.2.3. Sophisticated text processing techniques

- Natural language processing (NLP): it is a technique that allows a computer to make sense of spoken and written words, including even their nuances. This

---

27   D. Tikk, *Az informatika alkalmazásai: Szövegbányászat* (Budapest: Typotex, 2007). 300.

technology could be used to transcribe communication intercepts, audio/video clips, news reports and handwritten or hardcopy documents; data with key words could be flagged for a human analyst's attention.

- Sentiment analysis (SA): Sentiment analysis is a type of data mining that measures the inclination of people's opinions through natural language processing (NLP), computational linguistics and text analysis, which are then used to extract and analyse subjective information from the Web – mostly social media and similar sources. The analysed data quantifies the general public's sentiments or reactions toward certain products, people or ideas and reveals the contextual polarity of the information. Sentiment analysis is also known as opinion mining. Figure 6 shows the tag cloud of the most frequent positive and negative words regarding the fire at Notre Dame Cathedral.

Figure 6
Tag cloud of the most common positive and negative words connected to the fire in Notre Dame Cathedral
Source: edited by the author

- Social network analysis (SNA): Social network analysis is the process of investigating social structures using networks and graph theory. It characterises networked structures in terms of nodes (individual actors, people, or things within the network) and the ties, edges, or links (relationships

or interactions) that connect them. These networks are often visualised through sociograms in which nodes are represented as points and ties are represented as lines. These visualisations provide a means of qualitatively assessing networks by varying the visual representation of their nodes and edges to reflect attributes of interest. Figure 7 shows the pro- and anti-ISIL metacommunities on Twitter.
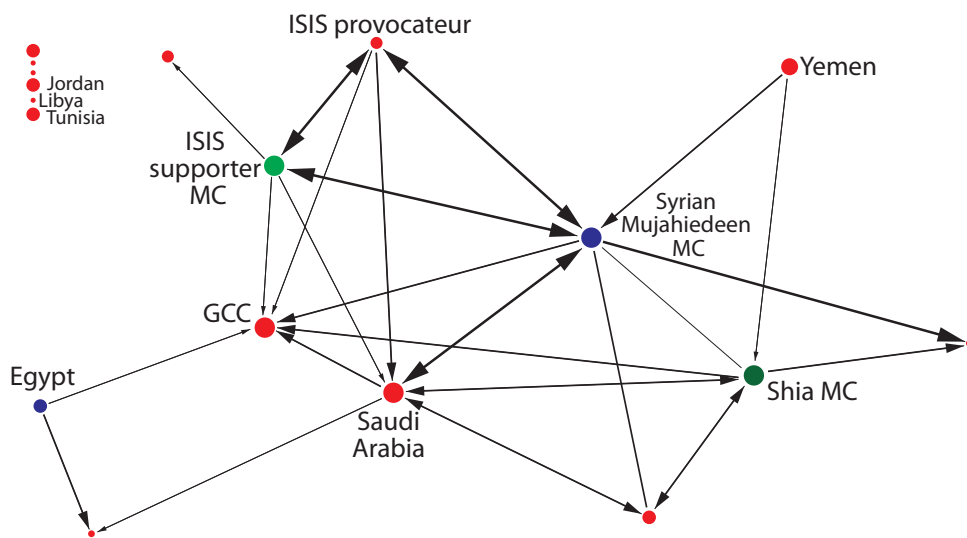

Figure 7

Pro- and anti-ISIL metacommunities and their interactions on Twitter

Source: W. Marcellino et alii, *Monitoring social media,* RAND Corporation, 2017, 34.

- Document clustering: Document clustering involves the use of descriptors and descriptor extraction. Descriptors are sets of words that describe the contents within the cluster. Document clustering is generally considered to be a centralised process. Examples of document clustering include web document clustering for search users.

It can be clearly seen that digital image processing and text mining plays an important role to extract meaning from the vast amount of surveillance data that the defence sector must deal with. Of course, the algorithms cannot work properly without the appropriate underlying infrastructure (database, hardware, analytical tools) that is capable of processing Big Data. A summary of processing techniques for imagery and textual data can be seen in Figure 8.
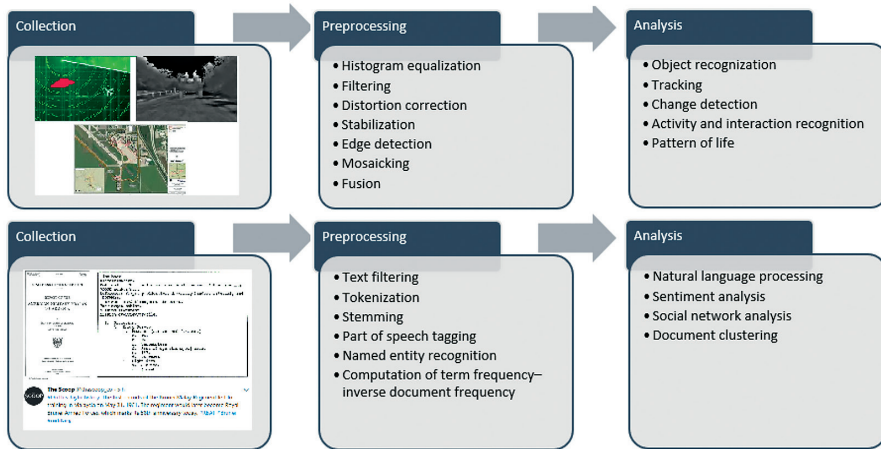
Figure 8
Processing imagery and textual data
Source: edited by the author

4. Summary

Big Data is creating the military of the future. In modern warfare, the most remarkable change is the appearance of information as a weapon. The basis of military operations is the ability to assess a situation made possible by accurate collected information. In today's military operations, the primary challenge is not the collection of data. It has become more important to retrieve relevant information from the collected data, which then can be converted to helpful reconnaissance data, knowledge for the command decision-making process.

With the ever-increasing amount of surveillance data that comes from the variety of sources in mostly unstructured formats, it becomes necessary to leverage the latest advances in information technology to successfully handle the vast amount of data. Without Big-Data analytic solutions, it would be impossible for analysts to sort through the billions of data points available (volume, variety, and velocity), identify the relevant and irrelevant pieces of data (veracity), safeguard the rights of citizens and follow other applicable laws and regulations, and discover relevant intelligence insights.

Big Data analytic technologies are a fast-developing field, and these tools are crucial components in dealing with the information overload they are facing. In these usage areas, security and real-time processing of data are extremely important. That is why the military sector is acutely aware of the importance of the different Big Data solutions and intends to access and implement these new technologies in their existing processes.

Most of the intelligence data is collected from high resolution satellites, large manned aircraft, smaller unmanned platforms, tiny handheld devices, unattended ground sensors and from different social media sources: online news, posts on micro-blogging sites like Twitter and social media platforms like Facebook.

The study presented the definition of Big Data and its application to military reconnaissance data and its focus on the processing of imagery and textual data, revealing modern data processing and analytics methods which help process them effectively.

The most often used image processing techniques were introduced, beginning with the different preprocessing techniques such as filtering, stabilisation and so on, to enhance image quality, and also the most sophisticated techniques to interpret the images and get final accurate results from the actual battlefield. In the case of textual data, the most important text mining techniques were discussed, including information retrieval, lexical analysis to study word frequency distributions, pattern recognition, tagging/annotation, information extraction, data mining techniques (including link and association analysis), visualisation, and predictive analytics.

References

'Big Data for defence and security.' Royal United Services Institute, ocassional paper, 2013. Available: www.slideshare.net/emcacademics/big-datafordefenceandsecurityreportfinal (30. 05. 2020.)

'Big Data for the Enterprise.' Oracle White Paper, June 2013. Available: www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf (30. 05. 2020.)

Dijk, Judith – Eekeren, Adam W. M. van – Rojas, Olga Rajadell – Burghouts, Gertjan J. – Schutte, Klamer: 'Image processing in aerial surveillance and reconnaissance: From pixels to understanding.' Proceedings of SPIE (The International Society for Optical Engineering), Electro-Optical Remote Sensing, Photonic Technologies, and Applications VII; and Military Applications in Hyperspectral Imaging and High Spatial Resolution Sensing, 88970A (15 October 2013). DOI: https://doi.org/10.1117/12.2029591

Erturk, S.: 'Improved Region of Interest for Infrared Images Using Rayleigh Contrast-Limited Adaptive Histogram Equalization.' 2013. Available: http://world-comp-proceedings.com/proc/p2013/IPC2477.pdf (30. 05. 2020.)

'Expanded Top Ten Big Data Security and Privacy Challenges.' Cloud Security Alliance, April 2013. Available: https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Expanded_Top_Ten_Big_Data_Security_and_Privacy_Challenges.pdf (30. 05. 2020.)

Haig, Zs.: Információs műveletek a kibertérben. Budapest, Dialóg Campus, 2018.

Hamilton, P. S. – Kreuzer P. M.: 'The Big Data Imperative. Air Force Intelligence for the Information Age.' Air & Space Power Journal 32, no 1 (2018), 4–20.

Harb, Johathas D. G. – Becker, Karin: 'Emotion analysis of reaction to Terrorism on Twitter.' Proceedings of the SBC 33rd Brazilian Symposium on Databases, 2018, Rio de Janeiro, Brazil, 97–108. Available: http://sbbd.org.br/2018/wp-content/uploads/sites/5/2018/08/097-sbbd_2018-fp.pdf (30. 10. 2020.)

Haridas, M.: 'Redefining military intelligence using big data.' *Scholar Warrior*, Autumn 2015, 72–78.

Loaiza, F. – Shah, J. – Rolfe, R.: 'Real-Time Information Extraction from Big Data.' Institute for Defence Analysis, Virginia, 2015. Available: www.researchgate.net/publication/305827265_Real-Time_Information_Extraction_from_Big_Data (30. 10. 2020.)

Marcellino, W. et alii: *Monitoring social media*. RAND Corporation, 2017. Available: www.rand.org/content/dam/rand/pubs/research_reports/RR1700/RR1742/RAND_RR1742.pdf (04. 02. 2021.)

ME4 Toh, Bao En: 'Swimming in Sensors, Drowning in Data – Big Data Analytics for Military Intelligence.' *Pointer, Journal of The Singapore Armed Forces* 42, no 1 (2016), 51–65.

National and Transnational Security Implications of Big Data in the Life Sciences. American Association for the Advancement of Science, 2014. Available: www.aaas.org/sites/default/files/AAAS-FBI-UNICRI_Big_Data_Report_111014.pdf (30. 05. 2020.)

Parekh, Deven – Amarasingam, Amarnath – Dawson, Lorne – Ruths, Derek: 'Studying Jihadists on Social Media: A Critique of Data Collection Methodologies.' *Perspectives on Terrorism* 12, no 3 (2018), 3–21. Available: www.universiteitleiden.nl/binaries/content/assets/customsites/perspectives-on-terrorism/2018/issue-3/01–-studying-jihadists-on-social-media-a-critique-of-data-collection-methodologies.pdf (30. 10. 2020.)

Shah, V. V. D.: 'Image Processing and its Military Applications.' *Defence Science Journal* 37, no 4 (1987), 457–468. DOI: https://doi.org/10.14429/dsj.37.5932

Terzi, Duygu Sinanc – Terzi, Ramazan – Sagiroglu, Seref: 'A survey on security and privacy issues in big data.' *Proceedings of the 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST)*, London, 2015. 202–207. DOI: https://doi.org/10.1109/icitst.2015.7412089

Thiele, R. D.: 'Mit Daten siegen – Big Data verändert Wirtschaft und Streitkräfte.' ISPSW Strategy Series: Focus on Defense and International Security, No. 393 (2015), 1–7. Available: www.files.ethz.ch/isn/195104/393_Thiele.pdf (30. 05. 2020.)

Tikk, D.: *Az informatika alkalmazásai: Szövegbányászat.* Budapest, Typotex, 2007.

'Use real-time HD image processing for military and civilian surveillance.' EDN, 2013. Available: www.edn.com/Pdf/ViewPdf?contentItemId=4424095 (30. 05. 2020.)

Vasileios, L.: *Comparative analysis of the hashtags #RefugeesWelcome and #StopRefugees*. Master thesis, Korinth, 2017.