A Framework for Big data Analytics in Video Surveillance

Dr. C ANITHA*, S GNANA NITHIS RAJ

*Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, SIMATS, Chennai, Tamilnadu, India, Email: anithac.sse@saveetha.com
19BCE2541, Department of Computer Science and Engineering, VIT, Vellore Campus, Vellore, Tamil Nadu, India, Email: gnananithis.raj2019@vitstudent.ac.in
*Corresponding Author

Abstract:
This article discusses the breadth and usage of big data analytics in video surveillance as well as potential future applications. The new technological era provides users with numerous opportunities to produce data at any given time. As a result, a massive amount of data is produced every second all over the world. A major percentage of this is accounted for by the amount of video data created. Education, science, geographical exploration, agriculture, safety and security, entertainment, etc. are the primary fields that create a vast amount of video data daily. A substantial chunk of this comes from the daily reports of surveillance data collected from the security camera that is utilised often. Such voluminous information calls for a dedicated infrastructure for archiving, retrieval, processing, and analysis. In this regard, Big Data Analytics is a useful tool. This piece's focus is on how Big Data Analytics is being used in the modern world of video surveillance. It also plans to zero in on Big Data research possibilities related to visual surveillance. The video surveillance tools used in the Big Data framework are described. This paper's primary objective is to investigate the state of video surveillance in the modern day through the lens of Big Data Analytics and to predict its potential future developments.

Keywords: Big Data Analytics, video surveillance, video analysis.
DOI Number: 10.14704/NQ.2022.20.13.NQ88043 **NeuroQuantology 2022; 20(13): 323-328**

1. Introduction

The term "big data" refers to data volumes that are too large for typical relational databases to effectively store, manage, and handle in a timely manner. Big data is characterised by a large amount, a fast rate of change, and a wide diversity of sources. Data sources are growing more complex than those for conventional data because of the impact of AI, social media, mobile devices, and the Internet of Things (IoT).
Devices, videos/audio, sensors, networks, log files, transactional apps, websites, and social media are all examples of sources that create data at an unprecedented velocity and volume. Using Big Data analytics, you may glean insights into things like previously unseen patterns, previously unknown relationships, emerging industry trends, and customer preferences from massive volumes of data. Analyzing large amounts of data may help with many things, including making better decisions and spotting and preventing fraud. Large amounts of unstructured big data now include abundant footage from surveillance cameras. When safety is a concern, CCTV (closed circuit television) cameras are set up. Manual monitoring appears to be a time-consuming and laborious process. Depending on the specifics of the scenario, security can be defined in a number

of ways, including the ability to prevent theft, detect violence, and estimate the chance of an explosion. "Security" has come to mean almost anything out of the ordinary in crowded public places. Since these events typically include large groups of people, spotting violent behaviour among them might be difficult. Due to many real-world restrictions, detecting anomalous or aberrant behaviour in a crowd video scene is extremely challenging. Security surveillance systems are used to provide secure and safe transportation, entertainment, and supported living in our daily lives.

2. Big Data

Big Data refers to a large quantity of data that has expanded exponentially over time. Because of its size and complexity, this data set defies the capabilities of standard data management systems. Although big data is equivalent to traditional data, it is far larger. There are three distinct forms of "big data," namely, structured, semi-structured, and unstructured. Structured data refers to information that can be stored, retrieved, and analysed in a predefined format. When the format is completely understood beforehand, computer science expertise has gotten better at coming up with new ways to interact with and gain benefit from such content. However, today's usual sizes range from multiple zetta bytes due to concerns that were anticipated when the size of such data increased to a great level. An illustration of structured data is a database table of the student. Any type of data that lacks a clear shape or arrangement is considered unstructured. Unstructured data, in addition to its massive size, also offers a variety of processing obstacles that must be addressed in order to extract value from it. A good example of a heterogeneous data source is one that combines unstructured data types with basic text files, photos, and videos. Today's businesses have access to vast amounts of data, but they often lack the expertise to get value from it since information is often unstructured or in its raw form. You may find both structured and unstructured data in the semi-structured category. The definition of a table in a relational database management system is different from the definition of semi-structured data, which appears to be structured [1]. The XML file format is an example of data that is semi-structured. the four V's of big data: volume, variety, velocity, and variability.

Volume: The term "Big Data" is used to describe a tremendous quantity of data. The size of the data set is a major factor in evaluating its significance. The quantity of information also plays a role in determining whether or not a collection of data may be considered "big data." Therefore, when using Big Data solutions, "Volume" must be taken into account.

Variety: Variety is the term for a broad range of data types and sources, both structured and unstructured. For the most part, apps only consider databases and spreadsheets as sources of data. Emails, photos, videos, audio recordings, data from monitoring devices, and portable document formats (PDFs) are all being included into data analysis programmes. Variation in unstructured data makes it more challenging to store, mine, and analyse.

Velocity: The term "velocity" is used to describe the rate at which new information is produced. How rapidly data can be generated and analysed to meet demands is what will define its actual potential. Big Data Velocity refers to the rate at which information is collected from many sources, including but not limited to business processes, application logs, networks, social media sites, sensors, mobile devices, and so on. There is a constant deluge of data.

Variability: This refers to the fact that there are instances when data displays inconsistency, which makes it difficult to manage and handle.

Veracity: It is crucial to take the reliability of the data source into account in order to properly assess the data.

Viscosity: It calculates the data's resistance to flow. It might relate to the inertia experienced while travelling through a data set. Due to the diversity of data sources, data flows may move more slowly and require more

complicated processing. Viscosity may be controlled with the use of enhanced streaming, a flexible integration bus, and smart event processing.

Virality: The term "virality" refers to the transfer of data through a network. It determines how quickly data may flow across networks. With the rate of spread, time is the most important quantity to consider.

1.1 The Big Data Sources

Black Box Data: This information is produced by all types of aircraft, including helicopters and jets. Voices of the flight crew, microphone recordings, and details on aircraft performance are all included in black box data.

Social Media Data: Social media platforms including Twitter, Facebook, Instagram, Pinterest, and Google+ have created this data.

Stock Exchange Data: This information comes from stock exchanges and relates to client sales and purchases of shares.

Power Grid Data: These are power grid data. It stores data on specific nodes, such as use details.

Transport Data: This covers a vehicle's potential capacity, model, availability, and travel distance.

Search Engine Data: One of the most important sources of large data is this. Search engines obtain their information from big databases.

Surveillance Data: A live video data set, or a collection of video clips recorded in difficult circumstances.

1.2 Challenges of Big Data

- One issue with big data is the exponential growth of raw data. Massive volumes of data are stored in data centres and databases, and the amount of data is continuously rising. With the exponential rise of data, it's becoming increasingly challenging for businesses to properly store it.
- The second issue is deciding which Big Data tool to use. There are many Big Data tools available, but choosing the wrong one might cost you effort, time, and money.
- The next issue with big data is its protection. Organizations typically overlook to safeguard data because they are so focused on understanding and analysing it. Unprotected data eventually becomes a haven for hackers.

3. Big Data Analytics

Big Data analytics is the process of acquiring, organising, and analysing enormous amounts of data in order to discover patterns and other important information. Big Data analytics may help businesses better comprehend the information included in their data and pinpoint the data that is most important to their operations and upcoming business decisions. Big Data analysts are frequently curious about the information discovered through data analysis.

3.1 Big Data Life Cycle

A business case, which outlines the rationale and objective of the study, is the first stage in the Big Data analytics lifecycle. Stage 2: Data Source Identification: In this stage, a wide range of data sources are identified. Data filtering in Stage 3 involves removing faulty data from all of the recognised data from Stage 2. Data that is incompatible with the tool is extracted at stage 4's "data extraction" stage and then changed into a format that is compatible. Data from several datasets with the same fields are combined in stage five, known as data aggregation. Stage 6: Data analysis. To find relevant information, data is assessed using analytical and statistical methods.

Stage 7: Data visualisation. Big Data analysts may create visual representations of the analysis using programmes like Tableau and Power BI. The last stage of the Big Data analytics life cycle, stage 8, is when the analysis's final findings are made available to corporate stakeholders so they may proceed with taking appropriate action. Figure 1 shows the life cycle of big data.

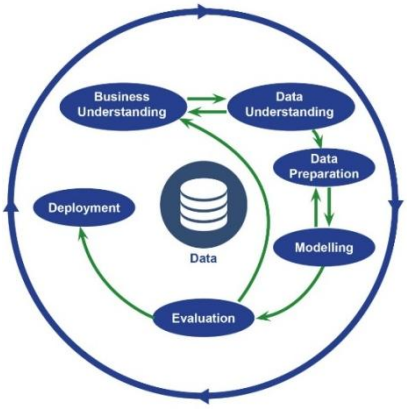

Figure 1 Big Data Life Cycle

3.2 Analytics Tools for Big Data

Various tools and technologies enable big data analytics processes. Some of the most popular technologies and tools for supporting big data analytics are included in Table 1 below.

Table 1 Big Data Analytics Tools

Tool	Description
R-Programming	For statistical research, scientific computing, and data visualisation utilising R Programming, a domain-specific programming language was created.
Altamira LUMIFY	Because it enables users to access analytics choices including graph visualisations, full-text faceted search, dynamic histograms, interactive geographic views, and collaborative workspaces that can be shared in real-time, Lumify is regarded as a strong big data analytics solution.
Apache Hadoop	On clusters of affordable hardware, Apache Hadoop is an open-source software platform for data storage and application execution.
MongoDB	A NoSQL document-oriented database called MongoDB is used to store large amounts of data. MongoDB is distinguished from Hadoop by its reputation for resilience.
RapidMiner	The most effective and simple graphical user interface for the design of the analytical process is provided by RapidMiner. It is a piece of open-source software.
Apache Spark	One of the most potent open source big data analytics tools is Apache Spark. It is a framework for data processing that has the ability to access very big data sets fast. Additionally, it may divide up data processing responsibilities across several machines, either independently or in cooperation with other distributed computing tools.
Microsoft Azure	Microsoft Azure, originally Windows Azure, is a platform for public cloud computing that it manages. It offers a variety of services, including networking, computing, analytics, and storage.
Xplenty	Simple graphical data pipelines are offered by the cloud-based ETL (Extract, Transform, and Load) system Xplenty. These pipelines provide automated data transfer between sources and destinations.
Zoho Analytics	A cloud-based reporting and analysis tool geared for business customers, Zoho Analytics is also offered for on-premises deployments. One of the product's main selling points is how simple it is to use. It uses a simple drag-and-drop user interface.

4. Role of Video Surveillance

Because cameras may capture criminals in the act, giving concrete proof for property owners and deterring potential offenders, video surveillance is an essential component of security. It's crucial to remember, though, that video surveillance is not the most effective security measure. A team of security guards must continuously watch the camera stream for video surveillance to be effective. Surveying video surveillance footage for hours, on end is not the best way to use somebody's time [2]. Therefore, businesses need solutions that can monitor the surveillance feed and alert security when something occurs – the best solution is video analytics. Video analytics generates a lot of ring in the surveillance industry due to functionalities that improve security. Here are a few examples of said functionality.

Tripwire detection: Through security cameras [3], video analytics can detect if someone crosses over a virtual boundary. Once the detection is noted, an instant alert is sent to security personnel.

People counting: Video analytics counts the number of people in a particular area [4], which is especially useful in large venues, like sports arenas. If there is a significant volume of foot traffic in one area, then the system can alert security personnel who can moderate the situation.

Motion detection: With the help of video analytics, security can set up zones where motion is not allowed [5]. The zones can be of any size or shape, and if someone enters the zone, security will pick up on it immediately via cameras.

Abnormal event detection: The identification of unusual events in video surveillance is a

difficult problem, but it is critical to the early-warning security and protection system [6].

Face Detection: Security cameras can scan someone's face and compare it against a list of public and private databases [7].

Automatic Number Plate Recognition (ANPR): To recognise and scan car licence plates from Charged CCTV [8], video cameras, law enforcement cameras, or specialised high-speed ANPR cameras deployed on roadside infrastructures, ANPR employs Optical Character Recognition (OCR) technology.

Crowd detection: A real-time monitoring system called crowd detection [9] enables the assessment of capacity or occupancy concerns within a specific region by detecting crowd density. Applications of crowd detection include population counts, managing large public events, managing disasters, monitoring safety, managing military operations, and detecting suspicious activities.

Surveillance video enhancement: Video Captured at low illumination can be enhanced to get high quality data [10].

Pan-tilt-zoom (PTZ) Auto-Tracking: Surveillance cameras [11] can follow and focus on persons within a greater field of view thanks to PTZ auto-tracking video analytics. Additionally, auto PTZ enables a security guard to multitask.

COVID-19 impact analysis: An active surveillance system [12] that can keep an eye on individuals and impose social distance between them can successfully restrict the spread of this lethal virus.

5. Conclusion

This article goes into depth on the development of big data analysis techniques for video surveillance. It examines a wide range of research opportunities in the area of managing, storing, and analysing video data. Many disciplines of information processing, including video image analysis, have given consideration to the use of big data analytic methodologies. Due to the numerous challenges, using big data for video analysis will be one of the most fascinating research areas in the upcoming years.

6. Reference

[1] Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. In 2013 international conference on collaboration technologies and systems (CTS) (pp. 42-47). IEEE.

[2] Subudhi, B. N., Rout, D. K., & Ghosh, A. (2019). Big data analytics for video surveillance. Multimedia Tools and Applications, 78(18), 26129-26162.

[3] Hasan, M., Mohan, S., Pellizzoni, R., & Bobba, R. B. (2020, March). Period adaptation for continuous security monitoring in multicore real-time systems. In 2020 Design, Automation & Test in Europe Conference & Exhibition (DATE) (pp. 430-435). IEEE.

[4] Khan, N., Ullah, A., Haq, I. U., Menon, V. G., & Baik, S. W. (2020). SD-Net: Understanding overcrowded scenes in real-time via an efficient dilated convolutional neural network. Journal of Real-Time Image Processing, 1-15.

[5] Salau, J., & Krieter, J. (2020). Analysing the Space-Usage-Pattern of a cow herd using video surveillance and automated motion detection. Biosystems Engineering, 197, 122-134.

[6] Leyva, R., Sanchez, V., & Li, C. T. (2017, April). The LV dataset: A realistic surveillance video dataset for abnormal event detection. In 2017 5th International Workshop on Biometrics and Forensics (IWBF) (pp. 1-6). IEEE.

[7] Saadat, M. N., Kabir, H., Long, Z. A., Sofian, H., & Zuhairi, M. F. A. (2020, January). Efficient Face Detection And Identification In Networked Video Surveillance Systems. In 2020 14th International Conference on Ubiquitous Information Management and Communication (IMCOM) (pp. 1-9). IEEE.

[8] Wang, R., Sang, N., Wang, R., & Jiang, L. (2014). Detection and tracking strategy for license plate detection in video. Optik, 125(10), 2283-2288.

[9] Zhan, B., Monekosso, D. N., Remagnino, P., Velastin, S. A., & Xu, L. Q. (2008). Crowd analysis: a survey. Machine Vision and Applications, 19(5), 345-357.

[10] Anitha, C., & Kumar, R.M.S. (2019).Naturalness Preserved Extremely Low Light Video Frame Enhancement by Correcting Illumination and Reducing Fixed Pattern Noise, Journal of Advanced Research in Dynamical and Control Systems,11(7 Special Issue), pp. 1043–1051.

[11] Ghosh, A., Subudhi, B. N., & Ghosh, S. (2012). Object detection from videos captured by moving camera by fuzzy edge incorporated Markov random field and local histogram matching. IEEE Transactions on Circuits and Systems for Video Technology, 22(8), 1127-1135.

[12] Haghighi, M. S., Sheikhjafari, A., Jolfaei, A., Farivar, F., & Ahmadzadeh, S. (2020). Automation of Recording in Smart Classrooms via Deep Learning and Bayesian Maximum a Posteriori Estimation of Instructor's Pose. IEEE Transactions on Industrial Informatics, 17(4), 2813-2820.

[13] Shorfuzzaman, M., Hossain, M. S., & Alhamid, M. F. (2021). Towards the sustainable development of smart cities through mass video surveillance: A response to the COVID-19 pandemic. Sustainable cities and society, 64, 102582.