Real Time Video Analytics Based on Deep Learning and Big Data for Smart Station

F Hidayat¹, F Hamami², I A Dahlan³, S H Supangkat⁴, A Fadillah⁵, and A Hidayatuloh⁶
¹,²,³,⁴,⁵,⁶ School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Indonesia

Abstract. CCTV is a sophisticated tool for monitoring and surveillance. It was chosen because of the low price and ease of use. CCTV creates issues in monitoring because officers must always supervise every CCTV monitor to look for abnormal conditions. This research proposes the implementation of smart CCTV for surveillance in Bandung Station. It applied deep learning algorithm to create smart CCTV for sensing and understanding station environment such as passengers’ flow, crowded area, prohibited location and others. Big data technologies are also carried out to solve the complexity of data. The objective of this research is to improve the quality of services in Bandung Station and make the station safer, more secure and more convenient.

1. Introduction

CCTV helps the authority to monitor and supervise environment or specific locations. CCTV is expected to make the officers’ jobs easier because it can reach many locations surveillance centered in one room. Based on CCTV, officer can act in real time action as soon as possible when abnormal conditions happen.

Railway is a mass transportation that is widely used by the public. Railway transportation is cheaper than airplanes for long distance travel. When compared with other land transportation, railway leads than others because it has precision departure and arrival time. People come to railway station to use railway transportation. CCTV cameras were installed to facilitate surveillance in the station area including the passengers. CCTV surveillance is needed to monitor everything in real time to prevent something undesirable happened.

CCTV is great but it is usually used when events have happened. This phenomenon occurs because to supervise the environments with CCTV officers must continuously monitor the CCTV screen. The issue can be solved with automated system seeing abnormal occurrences via CCTV in real time. Automated system is important to be implemented in public facilities at the station. It will help the local authorization to find out quick solution for an ongoing incident. Surveillance in Bandung station is only for monitoring and it cannot recognize objects or activity captured by CCTV.

Automated system is designed for several functions such as restricted area surveillance, object detection, object counting and others. The system process data and it will send real time notification for authority for immediate action.

Deep learning algorithm is used to extract and understand the objects captured by CCTV. Deep learning is widely used to process image data such as license plate recognition, fire detection, traffic indecent, vehicle detection, Fight events and others [1] [2][3][4][5].

Deep learning algorithm is proposed to understand real time video from CCTV. Several famous deep learning models are AE, CNN, DBN, LSTM, RBM, VAE, GAN and RNN [6]. This research implement CNN algorithm with YOLO v3 because of its accuracy and processing rate [7].
CCTV generates huge data in video or image format which is difficult to extract data. Its format is video or image and difficult to extract and understand. The unstructured and large data size makes a big issue and big data technology must be utilized in automated system development.

2. Methods
This research proposes the automated system that able to make CCTV as smart surveillance to do batch and real time analytics. The final output is displayed into dashboard and real time notification. Bandung railway station is the place where this research was conducted.

In the first stage, development of prototypes is done in the laboratory environment. In the next stage it will be deployed in the station environment.

2.1. Bandung Railway Station Profile
Bandung railway station is the area of this research. The station is divided into three zones. Bandung station sketch is shown in Figure 1.

![Figure 1. Bandung Station Sketch](image)

There are several important places in Bandung Station such as parking area, office, commercial area, platform, command center and security department.

2.2. Processing Infrastructure
Hardware specifications used in this research are:
1. Processor Intel i7-9700K
2. Memory 64 GB of RAM
3. Storage SSD and HDD
4. Power supply 1600W
5. GPU Nvidia RTX 2080 Ti,
6. CCTV IP.
Processing infrastructure is shown in Figure 2.

![Processing Infrastructure](image)

**Figure 2.** Processing Infrastructure

### 2.3. System Architecture

System architecture is divided into three big parts. First, deep learning processing to extract important features. Second, big data analytics in real time and batch processing. Third, visualization and real time notification. System architecture is shown in Figure 3.

![System Architecture](image)

**Figure 3.** System Architecture

Studies have been conducted to look for things that can help stations with CCTV. Several potential events that can be solved by smart CCTV are crowd situation, detection dangerous objects, surveillance danger zone, passenger pattern activity and others. These CCTVs generate huge data in video format. Proposed algorithm is developed to extract important features such as *person, datetime, cctv ID* and others. The extracted data from CCTV is called feature.

CNN algorithm extracts raw data from video or images and convert it into semi structured text data. The extracted features is written in JSON format. The following is the example of JSON data generated by CNN algorithm.

```
{"frame":"xxxx","datetime":"year-month-day
 hour:minute:second","object_1":"number-of-object-1","object_2":"number-of-
 object-2","object_n":"number-of-object-n", "cctv":"xx"}
```

Extracted features are *frame_id, datetime, object1, object2, object3 and cctv*. *frame_id* refers to frame ID. *datetime* refers to timestamp. *cctv* refers to ID of CCTV while *object-1 to object-n* refers to captured objects by CCTV. Examples of objects are *person, tvmonitor, chair* etc.
Convolutional neural network (ConvNets or CNNs) is one of the main categories to do images recognition, images classifications. Objects detections, recognition faces etc., are some of the areas where CNNs are widely used. This research also proposes that mechanism to separate which object is dangerous or secure. The dangerous objects are trained with list below and it will be added periodically depend on the needs and situation:

1. Gun
2. Knife
3. Mask
4. Blacklist People

Technically, deep learning using CNN models to train and test, each input image will pass it through a series of convolution layers with filters (Kernels), Pooling, fully connected layers and apply Softmax function to classify an object with probabilistic values between 0 and 1. The below figure is a complete flow of CNN to process an input image and classifies the objects based on values. This training list images will be labelled into defining classes. It means that we collect knife images and it will be labelled into knife classes and respectively for other objects. The next step it will be trained and published weights data. Lastly, this information will be released into the real-time dashboard. The example of CNN object detection is shown in Figure 4.

![Figure 4. CNN algorithm object detection](image)

In the second phase, Apache Spark and Node.js are used to process and analyse both stream or batch data. Apache spark is an open source distributed framework. Spark provides application programming interface to process data. It is fast, fault tolerance, cross platform and easy to use. Spark supports multiple programming languages such as Java, Scala, Python and R. It is powerful for both batch and real time processing. Spark reads semi-structured data in JSON format and process with aggregation and filtering.

Node.js is open source cross platform server. It can execute javascript outside the browser. Features of Node.js are asynchronous, fast execution, highly scalable, no buffering and it is under the MIT license. Node.js is the backend of the system and it was combined with WebSocket. WebSocket is real time communication protocol for web and mobile. WebSocket is needed to support real time response from extracted data generated by deep learning algorithm. WebSocket is able to send responses to clients even though clients do not send requests. The use of WebSocket is good for applications that require sending data in real time.

Raw data is also stored in MongoDB database. MongoDB can save dynamic attributes of data because the object captured by the CCTV is always changing. MongoDB is document-oriented database and classified as NoSQL database. It is a big data solution for storing huge text data. MongoDB schema is flexible not like SQL database [12]. MongoDB is one of the most popular NoSQL database because it is easy to implement and powerful. MongoDB stores data in collections (or similar like table in SQL) which can save different data attributes. Data in MongoDB stored in
Binary JSON (BSON) format. MongoDB can distribute its data into distributed way or horizontal scaling to store and compute huge data faster. The historical data in MongoDB can be analysed to find pattern activity such as the crowded hours, dangerous object or others in the station area. Stored data in MongoDB is shown in Figure 5.

![Figure 5. MongoDB Database](image)

Last phase, Node.js programming language helps to visualize and create real time notification to end-user. In deployment CCTVs are placed in strategic areas. In Figure 6 shows the station dashboard with circle dots refers to CCTV location. There are some CCTVs placed in the front area of the station and inside as well.

![Figure 6. Smart Station Dashboard](image)

### 3. Result and Discussion

Automated system experiment and result is divided into 6 parts.

#### 3.1. Feature extraction

Real time surveillance from CCTV from RTSP protocol is extracted to capture important features. Feature is properties of phenomenon being observed. Feature extraction is shown in Figure 7.
Figure 7. Features extraction from CCTV

Figure 7 shows the CCTV streaming video is processed to detect object in real time. Features from the video above is extracted in JSON format and consists of multiple attributes. The following is JSON data as extracted features.

{"frame":"1124","datetime":"2019-10-10 08:20:22 "person":"2", "cctv":"1"}

3.2. Object counting
System is also able to count objects in real time as shown in Figure 8.

Figure 8. Real time object counting

3.3. Object tracking
Object tracking is used to track unique object during captured by CCTV. It can identify the identity of person although the tagged person is in and out of the camera. Object tracking is shown in Figure 9. First, camera identifies two persons and gives two unique identities as id-1 and id-2. Then person with id-1 is not captured by camera and when that person came into camera system still able to detect him as id-1.

Figure 9. Object tracking
3.4. Unusual event
When system captures unusual event such as a person holding a sharp weapon or using mask, system detects it as a threat as follows in Figure 10.

![Figure 10. Dangerous object detection](image)

3.5. Data storage and batch analysis
Data generated from CCTV is also stored in MongoDB database to gain pattern activities. System does batch analytics from historical data to understand which time is crowded hours is shown in Figure 11.

![Figure 11. Activity pattern with batch analysis](image)

3.6. Real time notification
Both batch and real time processing, it returns output to web application and gives real time notification for specific officer to handle the issue in the station. Real time notification sent to email and telegram for authority is shown in Figure 12 and 13.

![Figure 12. Email Real time notification](image)
4. Conclusion

Smart CCTV as smart surveillance can help to improve the quality of services in station domain. It can do both batch and real time analytics for better decision support. The system can handle up to 30 fps to capture objects from CCTV. The proposed system is able to identify the normal and abnormal situation happened in station. The objective of smart surveillance is to make stations safer, more secure and more convenient.

5. References

[1] B. Sachin Prabhu, S. Kalambur, and D. Sitaram, “Recognition of Indian License Plate number from live stream videos,” 2017 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2017, vol. 2017-January, pp. 2359–2365, 2017.
[2] K. Muhammad, J. Ahmad, I. Mehmood, S. Rho, and S. W. Baik, “Convolutional Neural Networks Based Fire Detection in Surveillance Videos,” IEEE Access, vol. 6, pp. 18174–18183, 2018.
[3] A. P. Shah, J. B. Lamare, T. Nguyen-Anh, and A. Hauptmann, “CADP: A Novel Dataset for CCTV Traffic Camera based Accident Analysis,” Proc. AVSS 2018 - 2018 15th IEEE Int. Conf. Ad. Video Signal-Based Surveill., no. 1, pp. 1–9, 2019.
[4] S. Roy and M. S. Rahman, “Emergency Vehicle Detection on Heavy Traffic Road from CCTV Footage Using Deep Convolutional Neural Network,” 2nd Int. Conf. Electr. Comput. Commun. Eng. ECCE 2019, pp. 1–6, 2019.
[5] A. A. Einstein, “DETECTION OF REAL-WORLD FIGHTS IN SURVEILLANCE VIDEOS Mauricio Perez, Alex C. Kot School of Electrical and Electronic Engineering University of Campinas Institute of Computing,” pp. 2662–2666, 2019.
[6] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, “Deep learning for IoT big data and streaming analytics: A survey,” IEEE Commun. Surv. Tutorials, vol. 20, no. 4, pp. 2923–2960, 2018.
[7] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” 2018.
[8] P. Le Noac’h, A. Costan, and L. Bougé, “A performance evaluation of Apache Kafka in support of big data streaming applications,” Proc. - 2017 IEEE Int. Conf. Big Data, Big Data 2017, vol. 2018-Janua, pp. 4803–4806, 2018.
[9] K. Aziz, D. Zaidouni, and M. Bellafkih, “Real-time data analysis using Spark and Hadoop,” Proc. 2018 Int. Conf. Optim. Appl. ICOA 2018, pp. 1–6, 2018.
[10] D. Surekha, G. Swamy, and S. Venkatramanachikumar, “Real time streaming data storage and processing using storm and analytics with Hive,” Proc. 2016 Int. Conf. Adv. Comput. Control Comput. Technol. ICACCCCT 2016, no. 978, pp. 606–610, 2017.
[11] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” Proc. 2017 Int. Conf. Eng. Technol. IECT 2017, vol. 2018-Janua, pp. 1–6, 2018.
[12] I. Mearaj, P. Maheshwari, and M. J. Kaur, “Data Conversion from Traditional Relational Database to MongoDB using XAMPP and NoSQL,” IIT 2018 - Inf. Technol. Trends Emerg. Technol. Artif. Intell., pp. 94–98, 2019

Acknowledgments

This research is partially funded by the Indonesian Ministry of Research, Technology and Higher Education under WCU Program managed by Bandung Institute of Technology also supported by LPDP RISPRO KOMERSIAL Ministry of Finance, and Smart City Community Lab in Bandung Institute of Technology (ITB) Bandung.