International Journal of Scientific Engineering and Technology Research

ISSN 2319-8885
Volume.08,
Jan-Dec-2019,
Pages:166-170

www.ijsetr.com

An Object Detection and Classification Framework for High Performance Video Analytics

DIVYA NUNE
Assistant Professor, Dept of ECE, SITS, Tirupati, Andhrapradesh, India, Email: divyapsdv1995@gmail.com.

Abstract: Recent past has observed a rapid increase in the availability of inexpensive video cameras producing good quality videos. The video streams coming from these cameras need to be analyzed for extracting useful information such as object detection and object classification. Object detection from these video streams is one of the important applications of video analysis and becomes a starting point for other complex video analytics applications. Traditional video analysis approaches for object detection and classification such as color based, statistical background suppression, adaptive background; template matching and Guassian are subjective, inaccurate and at times may provide incomplete monitoring results. There is also a lack of object classification in these approaches. These approaches do not automatically produce color, size and object type information. Moreover, these approaches are costly and time consuming to such an extent that their usefulness is sometimes questionable. To overcome these challenges, we present a cloud based video stream analysis framework for object detection and classification. The framework focuses on building a scalable and robust cloud computing platform for performing automated analysis of thousands of recorded video streams with high detection and classification accuracy. An operator using this framework will only specify the analysis criteria and the duration of video streams to analyze. The analysis criteria define parameters for detecting objects of interests (face, car, van or truck) and size/color based classification of the detected objects. The recorded video streams are then automatically fetched from the cloud storage, decoded and analyzed on cloud resources. The operator is notified after completion of the video analysis and the analysis results can be accessed from the cloud storage.

Keywords: Adaptive Background, Guassian, Cloud Storage.

I. INTRODUCTION

Visual surveillance in dynamic business environments attempts to detect, track, and recognize objects of interest from multiple videos, and more generally to interpret object behaviors and actions. For instance, it aims to automatically compute the flux of people at public areas such as stores and travel sites, and then attain congestion and demographic analysis to assist in crowd traffic management and targeted advertisement. Such intelligent systems would replace the traditional surveillance setups where the number of cameras exceeds the capacity of costly human operators to monitor them. Proceeding with a low-level image features to high-level event understanding approach, there are three main steps of visual analytics: detection of objects and agents, tracking of such objects and indicators from frame to frame, and evaluating tracking results to describe and infer semantic events and latent phenomena. This analogy can be extended to other applications including motion-based recognition, access control, video indexing, human-computer interaction, and vehicle traffic monitoring and navigation. This chapter reviews fundamental aspects of the detection and tracking steps to support a deeper appreciation of many applications presented in the rest of the book. Imagine waiting for your turn in a shopping line at a busy department store. Your visual system can easily sense humans and identify different layers of their interactions.

As with other tasks that our brain does effortlessly, visual analytics has turned long out to be entangled for machines. Not surprisingly, this is also an open problem for visual perception. There are approximately 6 million cameras in the UK alone [1]. Camera based traffic monitoring and enforcement of speed restrictions have increased from just over 300,000 in 1996 to over 2 million in 2004 [2]. In a traditional video analysis approach, a video stream coming from a monitoring camera is either viewed live or is recorded on a bank of DVRs or computer HDD for later processing. Depending upon the needs, the recorded video stream is retrospectively analyzed by the operators. Manual analysis of the recorded video streams is an expensive undertaking. It is not only time consuming, but also requires a large number of staff, office work place and resources. A human operator loses concentration from video monitors only after 20 minutes [3]; making it impractical to go through the recorded videos in a time constrained scenario. In real life, an operator may have to juggle between viewing live and recorded video contents while searching for an object of interest, making the situation a lot worse especially when resources are scarce and decisions need to be made relatively quicker. Traditional video analysis approaches for object detection and classification such as color based [4], adaptive background [5], template matching [6] and Guassian [7] are subjective, inaccurate and at times may

provide incomplete monitoring results. There is also a lack of object classification in these approaches [4], [7]. These approaches do not automatically produce colour, size and object type information [5].

Moreover, these approaches are costly and time consuming to such an extent that their usefulness is sometimes questionable [6]. To overcome these challenges, we present a cloud based video stream analysis framework for object detection and classification. The framework focuses on building a scalable and robust cloud computing platform for performing automated analysis of thousands of recorded video streams with high detection and classification accuracy. An operator using this framework, will only specify the analysis criteria and the duration of video streams to analyse. The analysis criteria defines parameters for detecting objects of interests (face, car, van or truck) and size/colour based classification of the detected objects. The recorded video streams are then automatically fetched from the cloud storage, decoded and analysed on cloud resources. The operator is notified after completion of the video analysis and the analysis results can be accessed from the cloud storage. The framework reduces latencies in the video analysis pro- 2 cess by using GPUs mounted on computer servers in the cloud. This cloud based solution offers the capability to analyse video streams for on-demand and on-the-fly monitoring and analysis of the events.

The framework is evaluated with two case studies. The first case study is for vehicle detection and classification from the recorded video streams and the second one is for face detection from the video streams. We have selected these case studies for their wide spread applicability in the video analysis domain. The following are the main contributions of this paper: Firstly, to build a scalable and robust cloud solution that can perform quick analysis on thousands of stored/recorded video streams. Secondly, to automate the video analysis process so that no or minimal manual intervention is needed. Thirdly, to achieve high accuracy in object detection and classification during the video analysis process. This work is an extended version of our previous work [8]. The rest of the paper is organized as followed: The related work and state of the art are described in Section II. Our proposed video analysis framework is explained in Section III. This section also explains different components of our framework and their interaction with each other. Porting the framework to a public cloud is also discussed in this section. The video analysis approach used for detecting objects of interest from the recorded video streams is explained in Section IV. Section V explains the experimental setup and Section VI describes the evaluation of the framework in great detail.

II. LITERATURE REVIEW

Video analytics have also been the focus of commercial vendors. Vi-System [16] offers an intelligent surveillance system with real time monitoring, tracking of an object within a crowd using analytical rules. Vi-System does not work for recorded videos, analytics rules are limited and need to be defined in advance. Project BESAFE [17] aimed for automatic surveillance of people and tracking their abnormal behaviour using trajectories approach. It lacks scalability to a large number of streams and requires high bandwidth for video stream transmission. 3 IVA 5.60 [18] is an embedded video analysis system and is capable of detecting, tracking and analyzing moving objects in a video stream. EptaCloud [19] extends the functionality provided by IVA 5.60 in a scalable environment. The video analysis system is built into the cameras of IVA 5.60 that increases its installation cost. Intelligent Vision is not scalable and does not serve our requirements. Because of abundant computational power and extensive support on multi-threading, GPUs have become an active research area to improve performance of video processing algorithms. For example, Lui et. al. [20] proposed a hybrid parallel computing framework based on the MapReduce [21] .

The results suggest that such a model will be hugely beneficial for video processing and real time video analytics systems. We aim to use a similar approach in this research. Existing cloud based video analytics approaches do not support recorded video streams [16] and lack scalability [17]. GPU based approaches are still experimental [20]. IVA 5.60 [18] supports only embedded video analytics, otherwise their approaches are close to the approach presented in this research. The framework being reported in this paper uses GPU mounted servers in the cloud to capture and record video streams and to analyse the recorded video streams using a cascade of classifiers for object detection. Supported Video Formats: CIF, QCIF, 4CIF and Full HD video formats are supported for video stream recording in the presented framework. The resolution of a video stream in CIF format is 352x288 and each video frame has 99k pixels. QCIF (Quarter CIF) is a low resolution video format and is used in setups with limited network bandwidth. Video stream resolution in QCIF format is 176x144 and each video frame has 24.8k pixels. The resolution of a video stream in 4CIF format is 704x576 and each frame has 396k pixels. CIF and 4CIF formats have been used for acquiring video streams from the camera sources for traffic/object monitoring in our framework. Full HD (Full High Definition) video format captures video streams with 1920x1080 resolution. The video streams are captured at a constant bitrate of 200kbps and at 25 fps in the results reported in this paper. Table I summarizes the supported video formats and their parameters.

III. PROPOSED FRAME WORK

This section outlines the proposed framework, its different components and the interaction between them (Figure 1). The proposed framework provides a scalable and automated solution for video stream analysis with minimum latencies and user intervention. It also provides capability for video stream capture, storage and retrieval. This framework makes the video stream analysis process efficient and reduces the processing latencies by using GPU mounted servers in the cloud. It empowers a user by automating the process of identifying and finding objects and events of interest. Video stream are captured and stored in a local storage from a

cluster of cameras that have been installed on roads/buildings for the experiments being reported in this paper. The video streams are then transferred to cloud storage for further analysis and processing. The system architecture of the video analysis framework is depicted in Figure1 and the video streams analysis process on an individual compute node is depicted in Figure 2a.

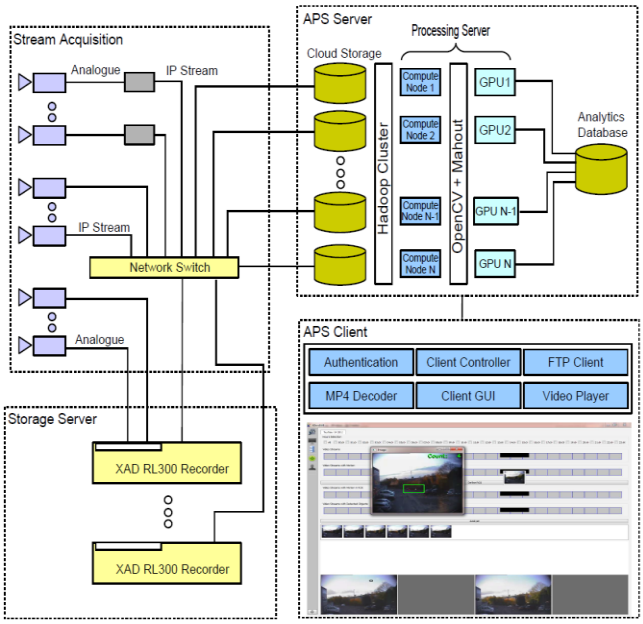

Fig1. System Architecture of the Video Analysis Framework.

Our framework employs a modular approach in its design. At the top level, it is divided into client and server components (Figure 1). The server component runs as a daemon on the cloud nodes and performs the main task of video stream analysis. Whereas, the client component supports multi-user environment and runs on the client machines (operators in our case). The control/data flow in the framework is divided into the following three stages:

- Video stream acquisition and storage
- Video stream analysis
- Storing analysis results and informing operator

A. Video Stream Acquisition

The Video Stream Acquisition component captures video streams from the monitoring cameras and transmits to the requesting clients for relaying in a control room and/or for storing these video streams in the cloud data center. The captured video streams are encoded using H.264 encoder.

B. Storage Server

The scale and management of the data coming from hundreds or thousands of cameras will be in exabytes, let alone all of the more than 4 million cameras in UK. Therefore, storage of these video streams is a real challenge. To address this issue, H.264 encoded video streams received from the video sources, via video stream acquisition, are recorded as MP4 files on storage servers in the cloud. The storage server has RL300 recorders for real time recording of video streams. It stores video streams on disk drives and meta-data about the video streams is recorded in a database (Figure 1). The received video streams are stored as 120 seconds long video files. These files can be stored in QCIF, CIF, 4CIF or in Full HD video format.

C. Analytics Processing Server (APS)

The APS server sits at the core of our framework and performs the video stream analysis. It uses the cloud storage for retrieving the recorded video streams and implements a processing server as compute nodes in a Hadoop cluster in the cloud data center (Figure 1). The analysis of the recorded video streams is performed on the compute nodes by applying the selected video analysis approach. The selection of a video analysis approach varies according to the intended video analysis purpose. The analytics results and meta-data about the video streams is stored in the Analytics Database.

D. APS Client

The APS Client is responsible for the end-user/operator interaction with the APS Server. The APS Client is deployed at the client sites such as police traffic control rooms or city council monitoring centers. It supports multi-user interaction and different users may initiate the analysis process for their specific requirements, such as object identification, object classification, or the region of interest analysis. These operators can select the duration of recorded video streams for analysis and can specify the analysis parameters. The analysis results are presented to the end-users after an analysis is completed. The analysed video streams along with the analysis results are accessible to the operator over 1/10 Gbps LAN connection from the cloud storage.

E. Porting the Video Analytics Framework to a Public Cloud

The presented video analysis framework is evaluated on the private cloud at the University of Derby. Porting the framework to a public cloud such as Amazon EC2, Google Compute Engine or Microsoft Azure will be a straight forward process.

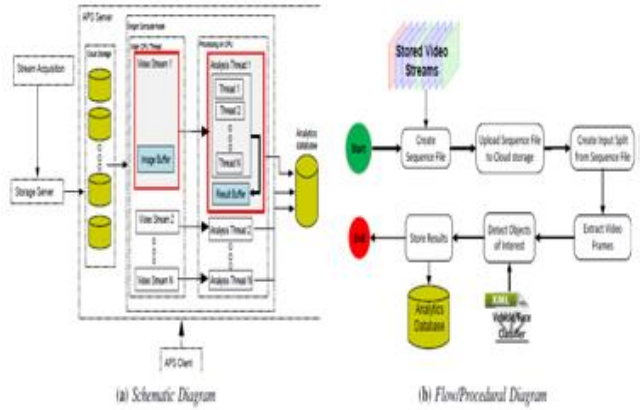

Fig 2: Video stream analysis on a compute node.

IV. EXPERIMENTAL RESULTS

We present and discuss the results obtained from the two configurations detailed in Section V. These results focus on

evaluating the framework for object detection accuracy, performance and scalability of the framework. The execution of the framework on the cloud nodes with GPUs evaluates the performance and detection accuracy of the video analysis approach for object detection and classification. It also evaluates the performance of the framework for video stream decoding, video stream data transfer between CPU and GPU and the performance gains by porting the compute intensive parts of the algorithm to the GPUs.

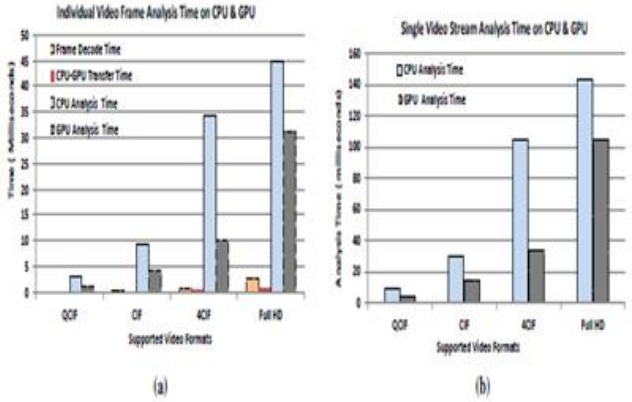

Fig 3: (a) Frame Decode, Transfer and Analysis Time, (b) Total Analysis Time of One Video Stream on CPU & GPU.

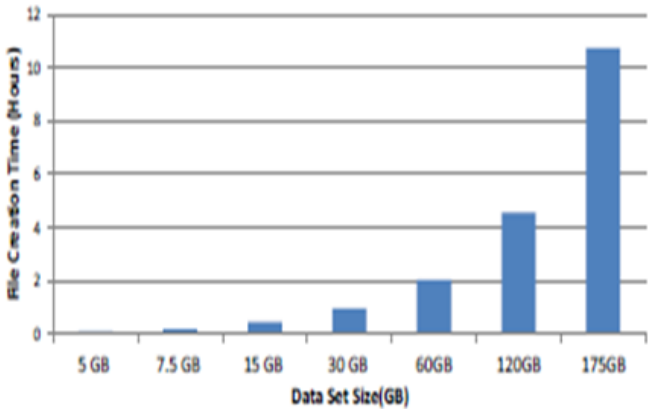

Fig 4: Sequence File Creation Time with Varying Data Sets.

The video stream decoding time for extracting one video frame for the supported video formats varied from between 0.11 to 2.78 milliseconds. The total time for decoding a video stream of 120 seconds duration varied between 330 milliseconds to 8.34 seconds for the supported video formats. It can be observed from Figure 5a that less time is taken to decode a lower resolution video format and more time to decode higher resolution video formats. The video stream decoding time is same for both CPU and GPU implementations as the video stream decoding is only done on CPU. Figure 4 depicts the time needed to convert input data sets into a sequence file. The time needed to create a sequence file increases with the increasing size of the data set. However, this is a one off process and the resulting file remains stored in cloud data storage for all future analysis purposes.

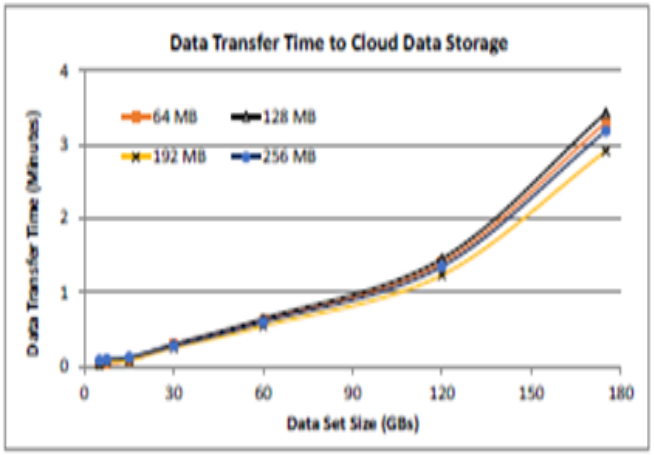

Fig 5: Data Transfer Time to Cloud Data Storage.

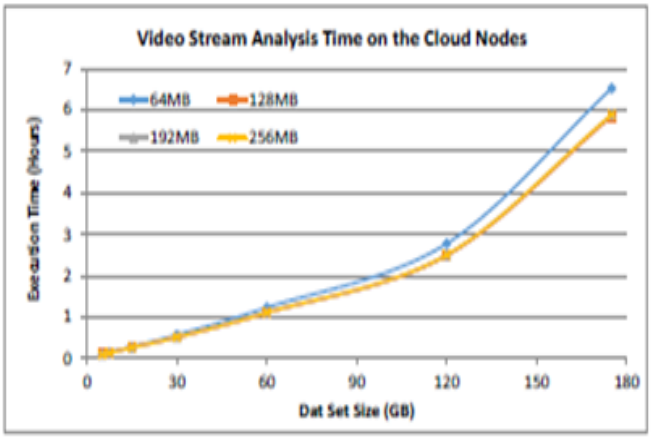

Fig 6: Video Stream Analysis Time on Cloud Nodes.

V.CONCLUSION

The cloud based video analytics framework for automated object detection and classification is presented and evaluated in this paper. The framework automated the video stream analysis process by using a cascade classifier and laid the foundation for the experimentation of a wide variety of video analytics algorithms. The video analytics framework is robust and can cope with varying number of nodes or increased volumes of data. The time to analyse one month of video data depicted a decreasing trend with the increasing number of nodes in the cloud.

VI. FUTURE SCOPE

In future, we would also like to extend our framework for processing the live data coming directly from the camera sources. This data will be directly written into data pipeline by converting into sequence files. We would also extend our framework by making it more subjective. It will enable us to perform logical queries, such as, "How many cars of a specific colour passed yesterday" on video streams.

VII. REFERENCES

[1] "The picture in not clear: How many surveillance cameras are there in the UK?" Research Report, July 2013.

[2] K. Ball, D. Lyon, D. M. Wood, C. Norris, and C. Raab, "A report on the surveillance society," Report, September 2006.
[3] M. Gill and A. Spriggs, "Assessing the impact of CCTV," London Home Office Research, Development and Statistics Directorate, February 2005.
[4] S. J. McKenna and S. Gong, "Tracking colour objects using adaptive mixture models," Image Vision Computing, vol. 17, pp. 225–231, 1999.
[5] D. Koller, J. W. W. Haung, J. Malik, G. Ogasawara, B. Rao, and S. Russel, "Towards robust automatic traffic scene analysis in real-time,"in International conference on Pattern recognition, 1994, pp. 126–131.
[6] J. S. Bae and T. L. Song, "Image tracking algorithm using template matching and PSNF-m," International Journal of Control, Automation, and Systems, vol. 6, no. 3, pp. 413–423, June 2008.
[7] J. Hsieh, W. Hu, C. Chang, and Y. Chen, "Shadow elimination for effective moving object detection by gaussian shadow modeling," Image and Vision Computing, vol. 21, no. 3, pp. 505–516, 2003.
[8] T. Abdullah, A. Anjum, M. Tariq, Y. Baltaci, and N. Antonopoulos, "Traffic monitoring using video analytics in clouds," in 7th IEEE/ACM Intl. Conf. on Utility and Cloud Computing, 2014, pp. 39–48.
[9] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 747–757, August 2000. [10] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the CVPR, 2001, pp. 511–518.
[11] Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, L. Cao, and T. Huang, "Large-scale image classification: Fast feature extraction and svm training," in Proceedings of the CVPR, 2011.
[12] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," in Proceedings of COLT'98, 1998.
[13] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in Proceedings of 26th IEEE MSST), 2010.
[14] A. Ishii and T. Suzumura, "Elastic stream computing with clouds," in 4th IEEE Intl. Conference on Cloud Computing, 2011, pp. 195–202.
[15] Y. Wu, C. Wu, B. Li, X. Qiu, and F. Lau, "CloudMedia: When cloud on demand meets video on demand," in 31st International Conference on Distributed Computing Systems, 2011, pp. 268–277.
[16] "Vi-system," http://www.agentvi.com/.
[17] "Project BESAFE," http://imagelab.ing.unimore.it/besafe/.
[18] "IVA 5.60 intelligent video analysis," Bosch, Tech. Rep., 2014.
[19] http://www.eptascape.com/products/eptaCloud.html
[20] K.-Y. Liu, T. Zhang, and L. Wang, "A new parallel video understanding and retrieval system," in Proceedings of the ICME, 2010, pp. 679–684.
[21] J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing on large clusters," Comm. of the ACM, vol. 51, no. 1, pp. 107–113, 2008.
[22] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Trans. on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560–576, July 2003.
[23] H. Schulzrinne, A. Rao, and R. Lanphier, "Real time streaming protocol (RTSP)," Internet RFC 2326, April 1996.
[24] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A transport protocol for real-time applications," Internet RFC 3550, 2203.
[25] "OpenCV," http://opencv.org/.