Real-time Scalable Video Stream Analysis with Object, Activity and Anomaly Detection.

Joseph Honour - 130291538

Newcastle University

Author Note

Completed as part of my undergraduate BSc in Computer Science with Industrial Placement (G401), supervised by Matt Collison.

Abstract

Computer vision has been a large area of research in recent years, devising methodologies to understand and act on events seen within video streams. A major application of computer vision is to detect anomalies autonomously, and alert users to when they occur. Although industrial technologies exist that are able to do this to a basic standard, they often rely on expensive and exclusive hardware.

This paper proposes an extendable and scalable framework that is able to provide an end-to-end video processing pipeline capable of accurate anomaly detection, in real-time, without complex hardware requirements. The framework will show how the adoption of distributed computing and machine learning enable real-time anomaly detections, without requiring specialized hardware. My design approach is to allow extensibility at every opportunity, so the framework can be adapted for a multitude of use cases, some of which I propose within this paper. Furthermore, the framework will allow horizontal scaling enabling it to handle large volumes of data, while keeping its real-time requirements intact. Finally, the framework will be hosted publicly allowing new avenues to be explored by the community, with avenues of exploration suggested at the end of this paper.

Declaration

“I declare that this dissertation represents my own work, except where otherwise stated.”

Acknowledgments

This is my acknowledgments.

Table of Contents

[Abstract 2](#_Toc507591834)

[Declaration 3](#_Toc507591835)

[Acknowledgments 4](#_Toc507591836)

[1. Introduction 7](#_Toc507591837)

[1.1 Motivation 7](#_Toc507591838)

[1.2 Aim 9](#_Toc507591839)

[1.3 Objectives 9](#_Toc507591840)

[1.4 Paper Structure 9](#_Toc507591841)

[2. Background and Literature Review 9](#_Toc507591842)

[2.1 Video Processing Methodologies and their Adoption 10](#_Toc507591843)

[2.1.1 Object Detection Techniques 11](#_Toc507591844)

[2.1.2 Object Tracking Techniques 13](#_Toc507591845)

[2.1.3 Behavior and Activity Analysis 14](#_Toc507591846)

[2.2 Anomaly Detection with Machine Learning 15](#_Toc507591847)

[2.2.1 Anomaly Detection Models 16](#_Toc507591848)

[2.3 Distributed Computing and the Cloud 17](#_Toc507591849)

[2.3.1 Parallel Computing 17](#_Toc507591850)

[2.3.2 Distributed Messaging 20](#_Toc507591851)

[2.3.3 Cloud Computing 21](#_Toc507591852)

[2.4 Existing Technologies and Approaches 22](#_Toc507591853)

[3. Framework Implementation 23](#_Toc507591854)

[3.1 Architecture 23](#_Toc507591855)

[3.2 Design Principles 24](#_Toc507591856)

[3.3 Language and Tool Adoption 24](#_Toc507591857)

[3.4 Client-Side Video Processing 24](#_Toc507591858)

[3.5 Server-Side Video Processing 24](#_Toc507591859)

[3.6 User Interfacing 24](#_Toc507591860)

[3.7 Deployment 24](#_Toc507591861)

[4. Evaluation 24](#_Toc507591862)

[4.1 Accuracy 24](#_Toc507591863)

[4.2 Performance 24](#_Toc507591864)

[4.3 Extendibility 24](#_Toc507591865)

[5. Conclusion 24](#_Toc507591866)

[5.1 Summary 25](#_Toc507591867)

[5.2 Future Development 25](#_Toc507591868)

[References 26](#_Toc507591869)

[Footnotes 30](#_Toc507591870)

[Tables 31](#_Toc507591871)

[Figures title: 32](#_Toc507591872)

# Introduction

## 1.1 Motivation

In the United Kingdom, an estimated 4.2 million surveillance cameras watch us every day (Norris and McCahill, 2006). With CCTV’s deployment to businesses, homes, shops and high streets we have become one of the most watched nations in the world. Fueling the mass deployment of surveillance equipment is its capability to deter criminals coupled with the recording of evidence of perpetrators if a crime is committed. This can be seen through CCTV’s 96% availability in homicide investigations (Yard, 2010). However, for it to be used an individual must manually watch all relevant footage, highlighting points of interest, which is extremely time consuming and inefficient. A more proactive approach to policing could evolve if the real-time detection of these points of interest within video-streams became available through automation.

This is the key problem with current CCTV; it does not provide real-time analysis or understanding of what is happening within the video stream. This means its adoption can currently has it only in a post-event detection capacity, given a large time input from users. It also means, due to its simple nature, all video footage must be stored for long periods of time, when a user may only be interested in the periods within the video stream containing points of interest. To solve this a system needs to be developed that can work with existing CCTV systems providing understanding of events occurring within a video, detecting and alerting users when points of interest occur.

Systems in the current market that could provide this functionality, only either provide sub-parts of the full desired behavior, or require specialized hardware requiring the replacement of currently deployed CCTV cameras (Table 2).

To enable the demands of intelligent video stream analysis, the development of a base framework is necessary. The framework should provide common functionality for video processing by default, while allowing easy development extension to cater to the broadest range of use cases. Furthermore, horizontally scaling technologies should provide the framework base to meet the demands of even the largest video stream development projects. This would allow an easy transition to developing smart video stream analysis from existing CCTV systems.

## 1.2 Aim

Propose an extendable framework for object detection and tracking, behavior and activity analysis, and anomaly detection, making use of distributed computing and machine learning.

## 1.3 Objectives

1. To research existing video processing techniques and available software packages in order to detect objects and activities within a video stream.
2. To research machine learning techniques with the capability to detect anomalies in time series data produced from objective one.
3. Develop testing scenarios that will allow the evaluation of a proposed real-time video processing framework.
4. Develop a framework that provides a minimum viable product of object, activity and anomaly detection, while being scalable and extensible.
5. Using the test scenarios defined in objective three, evaluate the frameworks ability to analyze real-time video streams in order to detect anomalies and alert users in real-time.
6. Compare and contrast the performance and storage requirements of the proposed framework against existing CCTV technologies and approaches.

## 1.4 Paper Structure

I will describe my paper structure here.

# Background and Literature Review

When developing an extendable framework for video processing, we must consider both software and hardware designs that will allow the framework to be suitable for the largest variety of video stream analysis scenarios. To do this, we must first look at common video processing methodologies and the software techniques adopted in providing them. Further to this, we need to decide how best to scale and extend the framework, making it a contender for high frequency scenarios. This will involve the investigation and adoption of Cloud Computing. The combination of suitable software and hardware decisions will provide the base for implementing a successful framework.

## 2.1 Video Processing Methodologies and their Adoption

In order to develop an effective video processing framework that enables accurate autonomous decisions, we must understand what is occurring within a series of sequential video frames. Existing architecture designs presented (Figure 1) show theoretically the different stages of processing required enabling successful video processing. These stages build upon each other, until we reach a point of desired intelligent understanding of the processed video stream.

The base of these stages is the ability to detect objects within a frame that are of interest to us. From this, tracking can then occur from frame to frame, giving objects persistent identity. Detection and tracking gives our first intelligent understanding of what is occurring in the video and provides the basis for more advanced techniques. Advanced techniques often include facial recognition and person identification, along with behavior and activity analysis, all of which require the location of objects and their previous movements through the video. Given we are able to provide this intelligence over the video stream we can extend analysis to providing anomaly detection on the observed behaviors, allowing real-time alerts to be sent to users.



Figure 1: Adaptation of video processing within a general framework for automated visual surveillance system (Ko, 2008).

### 2.1.1 Object Detection Techniques

The base of any video processing pipeline provides the ability to detect objects accurately within a frame. Detecting objects in images relies on being able to interpret the combinations of pixels correctly to identify the particular object you are looking for. The most common method of doing this is using a Haar feature-based cascade classifier (P. Viola and Jones, 2001). This machine learning approach works by showing a classifier a multitude of images, with some containing the object you wish to detect. Then, by applying features to the image that allow accurate pattern recognition of the object, we can train a classifier. From here, given a trained classifier, we extract the features from the images that identify the object, where each feature is a single value obtained by subtracting the sum of pixels under the white rectangle from the sum of pixels under the black rectangle (Figure 2). The composition of the identifying features of an object enable the final model to identify if, and where, an object is present to a degree of accuracy.Figure 2: Left, an example of Haar cascade features. Right, the adoption of Haar cascade features in detecting a face (OpenCV, 2018).

From this point we are able to improve the performance of detections using an algorithm called Adaboost (P. a Viola and Jones, 2001), which allows ranking of object detection features based on their error rate when attempting to successfully identify the object.

Further work has been proposed in this field using a deep learning based approach (Lecun, Bengio and Hinton, 2015). However, to make this approach feasible within real-time constraints specialized GPU (Graphics Processing Unit) hardware must be used, which would limit the extensibility of any proposed framework making use of these techniques.

Using the foundation of object detection, a video processing pipeline is able to provide the most basic level of analysis into understanding the video stream, enabling the engineering of more advanced processing techniques.

### 2.1.2 Object Tracking Techniques

Building upon object detection is object tracking. Accurate tracking enables the observation of object movement vectors, along with the monitoring of object interaction patterns, providing a more insightful level of analysis over simple object detection.

Tracking works on an online basis, meaning as new sequential data becomes available to the model, it can adapt at runtime to provide improved predictions on future data. Usually, this incorporates identifying an objects appearance and motion pattern allowing the accurate predictions of an objects location as it moves within a series of sequential video frames. When providing object tracking capabilities a variety of algorithms have been developed and made available (Table 1).

|  |  |  |  |
| --- | --- | --- | --- |
| Approach | Overview | Pros | Cons |
| Multiple Instance Learning | Tracks an object by treating its location as a set of positions (‘bags’) that each could contain the objects location based on its previous location. This gives the learning algorithm the responsibility of removing the ambiguity of the exact object location and predicting which instance in each bag is most correct (Baben and Belongie, 2009). | Works well when object partially occluded due to its bag representation. | The tracker cannot handle full occlusion of objects well. |
| Median Flow | Tracks an object by selecting a variety of points within the object space and then computes the trajectory for this object, both forwards and backwards in time. This means the tracker is able to compare both trajectories and make accurate predictions of the objects final location (Kalal, Mikolajczyk and Matas, 2010). | Works well when an objects trajectory is predictable. | The tracker becomes less reliable under fast or unpredictable motion. |
| Tracking-Learning-Detection | Breaks the tracking problem up into: Tracking, learning and detection. The tracker then follows an object between frames with the detector correcting the tracker based on all observed appearances. The learning estimates the error of the detector and updates it to avoid these errors in future (Kalal, Mikolajczyk and Matas, 2012). | Works well under long periods of occlusion. | The tracker does not perform well under large full rotations and can create many false positives. |

Table 1: A comparison of three common approaches to object tracking.

Given the successful employment of object tracking, objects now have identity and persistence as we analyze a video stream in real-time. With the observation of objects over time now enabled, we can observe the behaviors of an object, which provides the capability of further analysis.

### 2.1.3 Behavior and Activity Analysis

Activity recognition allows for the understanding of the full context of the video stream allowing the application of more advanced analysis later in the processing pipeline, such as anomaly detection.

Activity recognition techniques focus on calculating the probability of a performed activity being present, based on a series of observed actions. For instance, a witnessed person tracked between frames, moving at a constant speed and direction, might have a high probability of exhibiting traits associated with walking or running. In order to make distinctions between behaviors, and accurately model what is transpiring within the video stream, a variety of statistical modelling techniques exist.

Highly successful approaches to accurate behavior detection have adopted the Markov Model in various forms (Zin *et al.*, 2010; Kröckel and Bodendorf, 2012). The Markov Model approach works on predicting the next state in time series data by calculating the probability of transitioning to each possible prediction state from the current state. This allows the feeding of object tracking information into a Markov Model, which then calculates the probability of the series of actions corresponding to a given behavior.

Other popular methods of behavior analysis include (Ko, 2008):

* Dynamic Time Warping: a technique for comparing the similarity between two sequences of events. This allows the calculation of the probability that a shown behavior corresponds to a previously known behavior pattern.
* Finite State Machines: models’ behavior as a finite set of states, with transitions between related states. This allows for analysis on real-time video, watching for movements that trigger a tracked object to progress to a different state. A comparison between the transition path through the state machine and known behavior traits enables the identifying of witnessed behaviors.

The accurate analysis of behaviors within real-time video enable a contextual understanding of what is transpiring within the video stream. This understanding of activities within a video stream in itself provides a deep insight into what is transpiring, while permitting the adoption of more advanced video analysis techniques such as anomaly detection.

## 2.2 Anomaly Detection with Machine Learning

Anomaly detection provides a method for identifying unusual or outlying data points within a large set of data points. Within the domain of video processing, the application of these techniques can enable the detection of unusual behavior patterns observed within a video stream or abnormal frequencies of objects detected within a single video frame. Each anomaly detection method uses a model to represent the current state of the data, and then is able to compare new data points to the model and calculate whether it considers the point outside of the normal for the patterns’ it has been shown. When building detection models, it is common to adopt an unsupervised training approach, as it is often very complex to label existing data as being anomalous or not in order to provide a supervised learning data set to train a model.

### 2.2.1 Anomaly Detection Models

For video processing, as we want to detect anomalies over a series of sequential data points, we require models that work over time series data. Historically, models of this nature are cluster based; they are able to detect the density of points within a search space and identify outliers by calculating how far they deviate from a center of mass. Adaptations of this model explore the modelling of densities based on local and global instances, as well as the amount of clustered areas that correspond to normal behavior (Goldstein and Uchida, 2016).

Further to this, the adaptation of One Class Support Vector Machines have provided anomaly detection for short term observed behaviors within video streams (Antonakaki, Kosmopoulos and Perantonis, 2009). They are able to detect whether a short-term series of points is outside of normal observed behaviors, and provide an abnormal threshold capability that can distinguish which data points require more in-depth anomaly analysis.

Recently, the adoption of Neural Networks has provided success in the detection of anomalies, presenting a variety of different methodologies in solving the problem. Based around learning the distribution of features within the data, the models can detect when a data point falls outside of the normal distribution and classify it as anomalous. A Restricted Boltzmann Machine is a Neural Network model for the learning of distributions and, when multiple are stacked on top of each other, build a Deep Belief Network (Längkvist, Karlsson and Loutfi, 2014). However, Deep Belief Networks bring with them a large computational overhead and therefore may not be appropriate for a generic video processing framework.

In the context of anomaly detection, a video processing framework would make use of the proposed unsupervised models when looking at the data produced from the object detection, tracking and behavior analysis stages of the pipeline. The application of varying anomaly detection modes at each stage within the pipeline enable the exploration of multiple avenues of detection.

## 2.3 Distributed Computing and the Cloud

In order to provide video stream analysis at scale a large amount of computing power is required. Processing video with real-time requirements requires keeping the latency of any system under a desired threshold, even when presented with high volumes of data. In recent years, the availability of processing power has become more widely available, and cheap to access, through Cloud computing suppliers. This has enabled the adoption of distributed computing technologies that allow the parallel completion of tasks, enabling video processing to be possible on a large scale. Tools developed in recent years utilize distributed computing, and in this section, we will discuss their use and application in the field of computer vision.

### 2.3.1 Parallel Computing

Within the realm of video processing, there is a need to perform large amounts of calculations on individual frames, in order to provide services ranging from initial object detection to tracking and analysis. As more video streams are added to a system for analysis, there is a need for these expensive operations to scale, therefore we must be able perform these operations in parallel over a distributed set of machines. Two of the most popular frameworks that supply the capability of horizontally scalable distributed computing are Apache Storm (Apache, 2018c) and Apache Flink (Apache, 2018a).

Apache Storm is a real-time distributed stream data processing engine that, when given a source of data events, known as a spout, is able to perform a desired set of transformations to each event in parallel. Apache Storm runs on a distributed cluster of machines, where a topology is defined representing the series of processing stages to apply to each event. A topology is comprised of bolts, with a bolt representing an individual processing stage on an event. The execution of bolts happens within a worker process, with the management of its state controlled by a Zookeeper service. In order to track an events progress through a topology Apache Storm monitors the progress of an event through its transition between bolts, allowing Apache Storm to provide ‘at least once’ processing of events. Apache Storm then uses a service known as Nimbus to schedule and maintain the flow of data through a topology. This architecture gives Apache Storm power to process billions of events across a distributed set of machines in a single day (Toshniwal *et al.*, 2014).

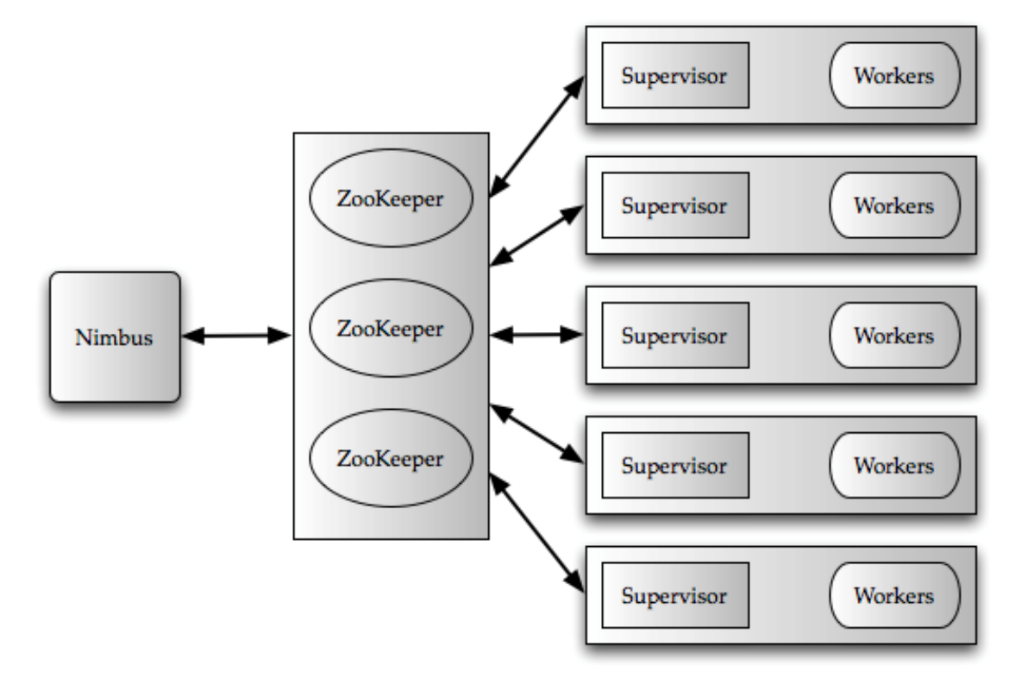


Figure 3: A high level architecture overview of Apache Storm (Toshniwal *et al.*, 2014).

A competitor to Apache Storm is the Apache Flink framework. Apache Flink is a system for processing streaming and batch data. Similar to Apache Storm, it works by expressing a data processing pipeline as a series of tasks to apply to each event as a directed graph. An Apache Flink architecture contains the core components of; a client, a Job Manager and at least one Task Manager (Figure 4). The client takes user code and transforms it into a dataflow graph, similar to Apache Storm’s topology. The Job Manager then coordinates the distributed execution of the dataflow, tracking state and progress of each operation within the stream. This then leaves the actual data processing execution to the Task Managers, reporting the result of any operation to the Job Manager. Differing from Apache Storm, the sending of messages between stages in the dataflow pipeline make use of buffers, with the backpressure from these buffers being used to control the throughput of the pipeline. This allows Apache Flink to provide throughputs of up to 80 million events per second (Carbone *et al.*, 2015).

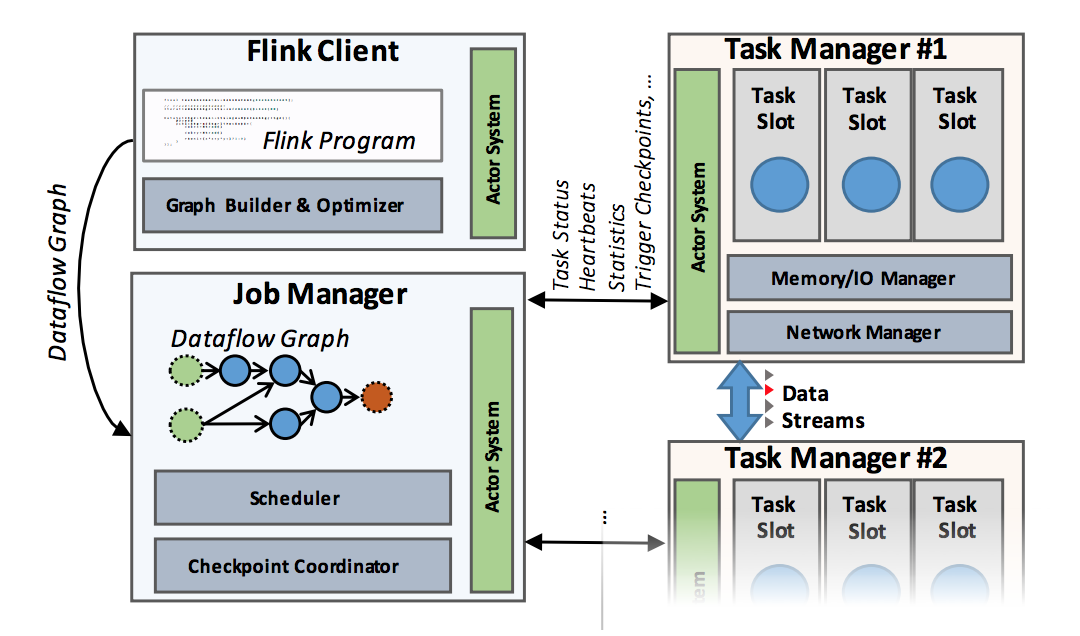


Figure 4: The Apache Flink processing model, showing the architecture adopted from an initial program definition to execution within an Apache Flink cluster (Carbone *et al.*, 2015).

As Apache Storm and Apache Flink offer similar services, performance comparisons between them exist. Under high workloads, it has been found that Apache Flink out performs Apache Storm marginally with regards to latency, with both systems producing similar throughputs (Chintapalli *et al.*, 2016). However, both systems are in their infancy of development, therefore full conclusion about the final performance of each system are hard to make.

The adoption of either of these, or similar, distributed processing frameworks allows for a proposed video processing pipeline to execute analysis on multiple video streams, even when complex calculations are required. The horizontal distribution of work between machines enables the meeting of real-time requirements at scale, a key requirement of a video processing framework.

### 2.3.2 Distributed Messaging

Within any video processing framework, a key success criterion is the ability to decouple the different stages of processing, allowing for the addition of new features without effecting existing stages of the pipeline. As well as decoupling, video processing analysis relies on creating an immutable log of objects, behaviors and anomalies detected within a given video stream. This immutable log enables the understanding of the causality of events; which is a core component of video processing analytics. Further to this, this functionality should be able to scale to multiple video stream inputs without effecting real-time performance requirements. Providing these capabilities is Apache Kafka (Apache, 2018b).

Apache Kafka is a distributed streaming platform, which lets applications publish and subscribe to topics of messages. It stores the messages published in a fault tolerant way, partitioning information across multiple machines, known as brokers, within an Apache Kafka cluster. To communicate with Apache Kafka a client performs a high-performance, language agnostic, TCP protocol, meaning any system can speak with it. Message queues, known as topics, are an immutable log of messages stored over multiple brokers. Clients can subscribe to topics, allowing for real-time processing of messages as they occur.

Apache Kafka’s fault tolerant storage means it is suitable as a storage system as well as a messaging system, allowing the execution of real-time and traditional off-line processing tasks. This enables the execution of a large variety of operations on the data held within Apache Kafka, making it a suitable choice for an extensible video processing framework. Further to its core functionality, Apache Kafka is also extremely fast, outperforming traditional messaging systems when handling the producing and consuming of messages between decoupled systems (Figure 5).

Figure 5: A comparison of Apache Kafka’s producer performance (Left), and its consumer performance (Right) against competitor messaging systems (Kreps, Narkhede and Rao, 2011).

### 2.3.3 Cloud Computing

To enable a horizontally scaling framework for video processing, a range of computing power is required, based on the scale and performance requirements of each system using the proposed framework. For applications working with a single video stream a smaller amount of computational power is required than applications performing analysis on hundreds of concurrent video streams. The adoption of Cloud computing technologies allows this elasticity within a proposed framework.

Cloud computing enables a user to rent the computational power they require, scaling up when demand is high, and reducing when systems become idle. This elasticity is a highly desirable feature when proposing a framework that needs to cater to a large variety of projects and requirements.

The benefits of Cloud computing do not come without risks however; with virtualized hardware exact performance metrics may be unknown until runtime. As a client of a Cloud provider, you are also at the mercy of their availability; if their service becomes unavailable, you are incapable of fixing it yourself. Further considerations to make can include, data confidentiality, software licensing costs and runtime renting costs (Armbrust *et al.*, 2009).

Even with these considerations, Cloud computing offers such elasticity and extensibility that supporting the deployment of a proposed framework to a Cloud Provider can offer great benefits. As Cloud computing becomes more highly adopted within industry the service offered will improve, unlike the physical purchase of hardware to support a product, bringing greater benefit to users of the proposed framework. These benefits therefore make Cloud computing a suitable candidate for the core architecture design of a scalable and extendable video processing framework, allowing users the option to deploy as much or little computing power required to meet their individual needs.

## 2.4 Existing Technologies and Approaches

Existing technologies and frameworks within the domain of video processing are often only designed for a single use case or are closed for development and therefore cannot be built upon. Table 2, shows an overview of recent video analysis frameworks, along with their limitations. It can be seen that the need for a modern, generic, video processing framework is required.

|  |  |  |  |
| --- | --- | --- | --- |
| Product | Overview | Functionality | Limitations |
| Nest (Nest, 2017) | Alerts users in real-time, via an app, to event and motion detection events seen on camera. | Provides event and anomaly detection and real-time alerts when these occur. | Requires specialized hardware, is not extendable for development. |
| A Video Analysis Framework for Surveillance System (Suvonvorn, 2008) | Gives a novel framework approach to online video analysis using .NET 2.0. | Provides object detection, event detection with extensibility to insert new functionality within the application. | Does not make use of distributed computing so will be unable to scale to all user requirements. Further to this, the product is developer in .NET, requiring a Microsoft workstation to run. |
| DiVA (SanMiguel *et al.*, 2008) | Gives a distributed video processing framework that uses a database as a message source allowing components to communicate agnostic of technologies adopted. | Provides object detection with the goal of detecting object abandonment and removal. Extensibility is provided through the communication of modules/algorithms occurring at a database level. | Uses a singular database to communicate, rather than a distributed messaging system. It also requires fixed hardware meaning it currently does not make use of Cloud Computing. |

Table 2: A list of current technologies in the market that aim to provide smart CCTV.

# Framework Implementation

## 3.1 Architecture

This will be my architecture summary.

## 3.2 Design Principles

This will be my design principles.

## 3.3 Language and Tool Adoption

This will be my language and tool adoption.

## 3.4 Client-Side Video Processing

This will be a discussion on what is developed client-side, including testing done.

## 3.5 Server-Side Video Processing

This will be a discussion on what is developed server-side, including testing done.

## 3.6 User Interfacing

This will be a discussion on what is done to notify the user and to explore data, including testing done.

## 3.7 Deployment

This will be a discussion about how to deploy the pipeline.

# Evaluation

Speak about the use case we will evaluate the pipeline under.

## 4.1 Accuracy

This will be about the pipelines accuracy at different stages.

## 4.2 Performance

This will be about the pipelines latency and throughput, along with scalability.

## 4.3 Extendibility

This will be an evaluation of extendibility.

# Conclusion

This will be my conclusion.

## 5.1 Summary

This will be a summary of work completed and its success.

## 5.2 Future Development

This will be a look into areas of future development.

References

Antonakaki, P., Kosmopoulos, D. and Perantonis, S. J. (2009) ‘Detecting abnormal human behaviour using multiple cameras’, *Signal Processing*, 89(9), pp. 1723–1738. doi: 10.1016/j.sigpro.2009.03.016.

Apache (2018a) *Apache Flink*. Available at: https://flink.apache.org/ (Accessed: 19 February 2018).

Apache (2018b) ‘Apache Kafka’. Available at: https://kafka.apache.org/.

Apache (2018c) *Apache Storm*. Available at: http://storm.apache.org/ (Accessed: 19 February 2018).

Armbrust, M. *et al.* (2009) ‘Above the clouds: A Berkeley view of cloud computing’,  *University of California, Berkeley, Tech. Rep. UCB* , pp. 07–013. doi: 10.1145/1721654.1721672.

Baben, B. and Belongie, S. (2009) ‘Visual tracking with online Multiple Instance Learning’, *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 983–990. doi: 10.1109/CVPR.2009.5206737.

Carbone, P. *et al.* (2015) ‘Apache Flink: Unified Stream and Batch Processing in a Single Engine’, *Data Engineering*, 36, pp. 28–38. doi: 10.1109/IC2EW.2016.56.

Chintapalli, S. *et al.* (2016) ‘Benchmarking streaming computation engines: Storm, flink and spark streaming’, in *Proceedings - 2016 IEEE 30th International Parallel and Distributed Processing Symposium, IPDPS 2016*, pp. 1789–1792. doi: 10.1109/IPDPSW.2016.138.

Goldstein, M. and Uchida, S. (2016) ‘A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data’, *PLoS ONE*, 11(4). doi: 10.1371/journal.pone.0152173.

Kalal, Z., Mikolajczyk, K. and Matas, J. (2010) ‘Forward-backward error: Automatic detection of tracking failures’, in *Proceedings - International Conference on Pattern Recognition*, pp. 2756–2759. doi: 10.1109/ICPR.2010.675.

Kalal, Z., Mikolajczyk, K. and Matas, J. (2012) ‘Tracking-learning-detection’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(7), pp. 1409–1422. doi: 10.1109/TPAMI.2011.239.

Ko, T. (2008) ‘A survey on behavior analysis in video surveillance for homeland security applications’, *Applied Imagery Pattern Recognition Workshop, 2008. AIPR ’08. 37th IEEE*, pp. 1–8. doi: 10.1109/AIPR.2008.4906450.

Kreps, J., Narkhede, N. and Rao, J. (2011) ‘Kafka: a Distributed Messaging System for Log Processing’. Available at: http://notes.stephenholiday.com/Kafka.pdf.

Kröckel, J. and Bodendorf, F. (2012) ‘Intelligent Processing of Video Streams for Visual Customer Behavior Analysis’, *ICONS 2012, The Seventh International Conference on Systems*, (c), pp. 163–168. Available at: http://www.thinkmind.org/index.php?view=article&articleid=icons\_2012\_8\_20\_20040.

Längkvist, M., Karlsson, L. and Loutfi, A. (2014) ‘A review of unsupervised feature learning and deep learning for time-series modeling’, *Pattern Recognition Letters*, 42(1), pp. 11–24. doi: 10.1016/j.patrec.2014.01.008.

Lecun, Y., Bengio, Y. and Hinton, G. (2015) ‘Deep learning’, *Nature*, pp. 436–444. doi: 10.1038/nature14539.

Nest (2017) *No Title*. Available at: https://nest.com/uk/cameras/nest-cam-indoor/overview/ (Accessed: 24 November 2017).

Norris, C. and McCahill, M. (2006) ‘CCTV: Beyond penal modernism?’, *British Journal of Criminology*, 46(1), pp. 97–118. doi: 10.1093/bjc/azi047.

OpenCV (2018) *Facial Detection with Haar Cascades in OpenCV*. Available at: https://docs.opencv.org/3.3.0/d7/d8b/tutorial\_py\_face\_detection.html (Accessed: 30 January 2018).

SanMiguel, J. C. *et al.* (2008) ‘DiVA: A Distributed Video Analysis Framework Applied to Video-Surveillance Systems’, in *International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, pp. 207–210. doi: 10.1109/WIAMIS.2008.29.

Suvonvorn, N. (2008) ‘A video analysis framework for surveillance system’, *2008 IEEE 10th Workshop on Multimedia Signal Processing*, pp. 867–871. doi: 10.1109/MMSP.2008.4665195.

Toshniwal, A. *et al.* (2014) ‘Storm@twitter’, in *Proceedings of the 2014 ACM SIGMOD international conference on Management of data - SIGMOD ’14*, pp. 147–156. doi: 10.1145/2588555.2595641.

Viola, P. a and Jones, M. J. (2001) ‘Fast and Robust Classification using Asymmetric AdaBoost and a Detector Cascade’, *Advances in Neural Information Processing System*, (December), pp. 1311–1318. Available at: https://papers.nips.cc/paper/2091-fast-and-robust-classification-using-asymmetric-adaboost-and-a-detector-cascade.pdf.

Viola, P. and Jones, M. (2001) ‘Robust real-time object detection’, *International Journal of Computer Vision*, 57(2), pp. 137–154. doi: http://dx.doi.org/10.1023/B:VISI.0000013087.49260.fb.

Yard, S. (2010) *CCTV in Homicide Investigations*. Available at: https://goo.gl/oS5Tgn (Accessed: 15 November 2017).

Zin, T. T. Z. T. T. *et al.* (2010) ‘A Markov Random Walk Model for Loitering People Detection’, *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on*. doi: 10.1109/IIHMSP.2010.172.

Footnotes

1[Add footnotes, if any, on their own page following references. For APA formatting requirements, it’s easy to just type your own footnote references and notes. To format a footnote reference, select the number and then, on the Home tab, in the Styles gallery, click Footnote Reference. The body of a footnote, such as this example, uses the Normal text style. (Note: If you delete this sample footnote, don’t forget to delete its in-text reference as well. That’s at the end of the sample Heading 2 paragraph on the first page of body content in this template.)]

Tables

Table 1

[Table Title]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Column Head | Column Head | Column Head | Column Head | Column Head |
| Row Head | 123 | 123 | 123 | 123 |
| Row Head | 456 | 456 | 456 | 456 |
| Row Head | 789 | 789 | 789 | 789 |
| Row Head | 123 | 123 | 123 | 123 |
| Row Head | 456 | 456 | 456 | 456 |
| Row Head | 789 | 789 | 789 | 789 |

Note: [Place all tables for your paper in a tables section, following references (and, if applicable, footnotes). Start a new page for each table, include a table number and table title for each, as shown on this page. All explanatory text appears in a table note that follows the table, such as this one. Use the Table/Figure style, available on the Home tab, in the Styles gallery, to get the spacing between table and note. Tables in APA format can use single or 1.5 line spacing. Include a heading for every row and column, even if the content seems obvious. A default table style has been setup for this template that fits APA guidelines. To insert a table, on the Insert tab, click Table.]

Figures title:

Figure 1. [Include all figures in their own section, following references (and footnotes and tables, if applicable). Include a numbered caption for each figure. Use the Table/Figure style for easy spacing between figure and caption.]

For more information about all elements of APA formatting, please consult the APA Style Manual, 6th Edition.