

PAPER • OPEN ACCESS

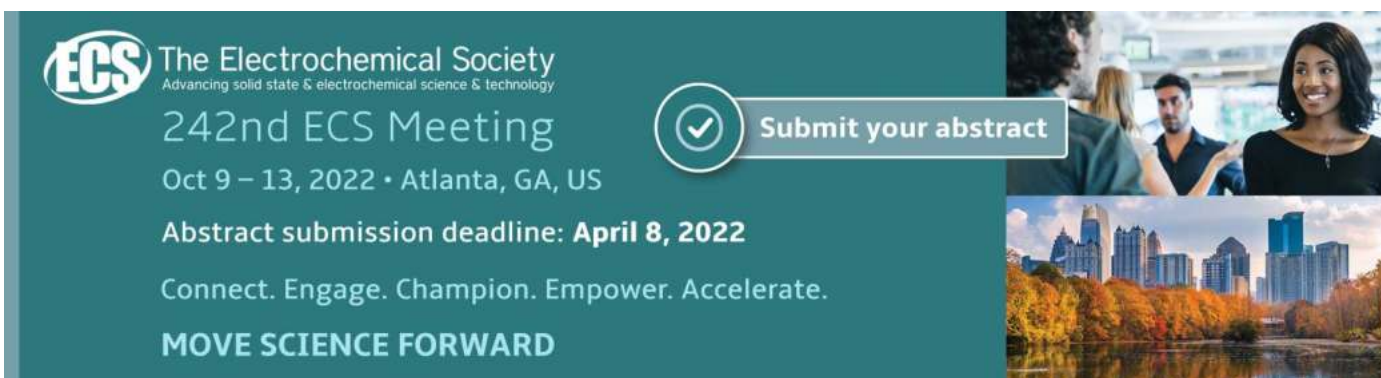
Automatic Detection of Anomalies in Video Surveillance using Artificial Intelligence

To cite this article: Sreedevi R Krishnan *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1085** 012020

View the [article online](#) for updates and enhancements.

You may also like

- [SRP Meeting: Social and Political Implications of Communicating Radiation Risk Daresbury, Warrington, 20 June 2001](#)
Karen Davies
- [The Role of Cash Flow in Financial Early Warning of Agricultural Enterprises Based on Logistic Model](#)
Fengru Sun
- [SRP Annual General Meeting: Health Physics Instrumentation and Analytical Techniques Edinburgh, 24-26 April 2001](#)
Barbara Gallani, Dave Drury and Steve Gower



ECS The Electrochemical Society
Advancing solid state & electrochemical science & technology

242nd ECS Meeting
Oct 9 – 13, 2022 • Atlanta, GA, US

Abstract submission deadline: **April 8, 2022**

Connect. Engage. Champion. Empower. Accelerate.

MOVE SCIENCE FORWARD

Submit your abstract

The banner features a teal background with white and light blue text. On the right side, there are two images: the top one shows a group of people in a meeting, and the bottom one shows a city skyline with a lake and trees in autumn.

Automatic Detection of Anomalies in Video Surveillance using Artificial Intelligence

Sreedevi R Krishnan^{1*}, P Amudha¹, and S Sivakumari¹

¹Department of Computer Science & Engineering, School of Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, Tamil Nadu, India.

*Email id sreedevirkrishnan@gmail.com

Abstract The significance of security in day to day life is increasing, and hence the use of a video surveillance system is becoming commonly accepted in almost every public places. Even though by placing surveillance system in public places will help to trace out the culprits of the anomalous situation, it is hard to trace it out. The knowledge of the incidence time is needed to identify the event, and also, it requires an enormous data handling task. To identify the anomalous situation in real-time will help the authorities to decrease the consequences and loss during the anomalous event. The paper proposes an in-depth study of various automatic anomaly detection techniques which helps to reduce the loss occurred of the anomalous situation. The advancements in Artificial Intelligence help in the quick and automatic identification of nominal and anomalous events. The sequential and incremental learning approach in feature extraction will help to generate a model that will provide much accurate classifications and predictions of anomalies.

1. Introduction

The need for security to the life and property of ordinary people is increasing. As the violence during theft increases, for guaranteeing the security of citizens in public places need an authorised, and the behaviour of the people needs to be analysed. Nowadays, most public areas are under Closed Circuit Television (CCTV) surveillance for providing security to people and property. Using CCTV surveillance, a large amount of video data needs to be analysed and to handle this Big Data is tedious.

A smart CCTV is needed for the identification and behaviour analysis of the people in public places. Various Neural Networks algorithms are useful for the automatic identification of abnormal situations in a security system. Using Artificial Intelligence, we can identify the features and frequency of occurrence of abnormal events. Besides, the study provides techniques to reduce the amount of data storage by removing the videos, which is under normal conditions, and this will solve the main problem related to Big Data management. The motivation to select this topic is to ensure the safety and security of people in the public area in a fast and effective way. The immediate identification of abnormal events will help to reduce the casualty of the situation or even avoiding the situation, thereby enforcing the security of people.

2. Automatic Anomaly Detection in Video Surveillance



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Through Artificial intelligence, the machine can mimic a human being and also may have human cognisance without being manually coded. Artificial Intelligence performs a systematic study of algorithms and statistical analysis to build a model using artificial neural network. It relies on data patterns and deductions for making predictions using computers. AI algorithms build a model based on the information known as *training data*; this model can predict the future without explicitly coded to execute out the task. During the Learning procedure, certain information which is known from previous experience is given to the machine called training. During this training, the machine will consequently distinguish the knowledge. Using the model built for classifying nominal and anomalous events using Artificial Intelligence, the various anomalies occurring in the real-life scenario can be detected. By providing proper alerts, the after-effects of abnormal events can be reduced to an extent. Some of the anomaly detection methods are discussed in the following subsections.

2.1. Smart Processing and Storage Utilisation method in Surveillance System

Z. Shao, J. Cai, and Z. Wang [2] present an innovative smart processing and storage utilisation method, in a video surveillance system for event identification and alert message. The strategy incorporates three sections: smart pre-alert for rare events, sufficient storage for surveillance video, and quick access to suspected videos, which ultimately discovers the temporal-spatial relation analysis of rare events in various checkpoints. Initially, an irregular behavioural database is established for a smart surveillance system that stores and manages alerts in case of crisis. Then, selected videos of the surveillance system are stored according to the alert information and avoid other videos; this significantly decreases the storage space consumption. Finally, when video evidence is required, the videos related to irregular behaviours are outlined and retrieved accordingly.

The characteristics of criminal behaviour are studied from previous experiences and recorded, to distinguish the anomalous practices from ordinary conduct. Anomalous behaviours can be decided by the frequency of occurrence, the identity of the target, and the surrounding area of an event. Anomalous activity database is initially created by storing the vivid data of anomalous activities captured from various video cameras. This metadata comprises of several types of anomalous conduct, name of checking premises, the details of video camera including device ID and location, starting and finishing time of the event, Etc. The anomalous behaviour database gives the metadata for further examination.

The Event- Abnormal Behaviour Correlation Model [2] is used to identify and co-relate the events. The model can identify abnormal activities with temporal auto-correlation greater than the previously set observational limit as an unsafe one. When an anomalous behaviour alert is activated from a specific video surveillance device, the smart surveillance system will analyse the temporal correlation regarding previous records in the database. A multi-site relation is examined regarding the activity of the criminal suspect.

The criminals usually follow a practice of studying the premises by roaming before doing a criminal activity; this spatial association can be identified from the various surveillance cameras placed in the smart surveillance system. The surveillance system is shown in figure 1, in which people who are commonly roaming during an event can be recognised and re-identified [4]. If the person is identified using the face recognition system [1] and found as walking suspiciously, it will be marked as a notified event. If the roaming occurs in various destinations at the same time: the smart observing camera tries to identify the persons by checking the multi-site checkpoints using spatial relevance and then raise the alert. The reoccurrence at a specific time interval and temporal auto-correlation is taken for the significance of the conduct. If two or three roaming occasions are recognised at a time by the surveillance system in various areas, the roaming conduct will be fixed as a high-risk one. If the same person visits multi-site and reports events of wandering: then identify that he is part of an activity that is closely related to time and space risk. In the event, if the roaming reoccurrence in the examination

time frame is deeply found a lot higher than the ordinarily reported information at the hour, then this event will be a suspicious one.

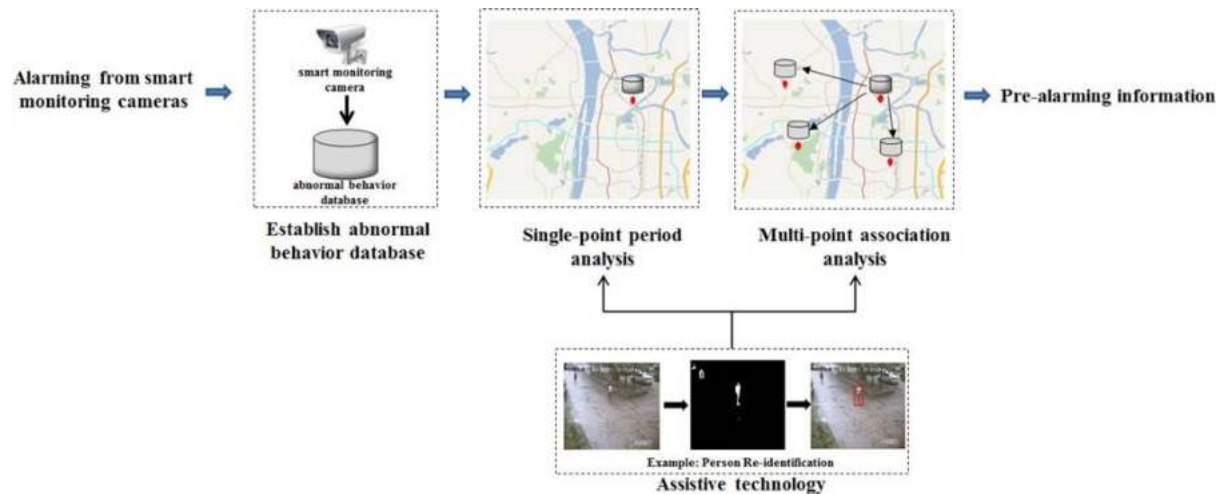


Figure 1. Pre-alarming based on abnormal behaviour association analysis

When an accident or abnormal condition occurs, the model will be refreshed and updated. From the abnormal characteristics database, recurrence, and various abnormal behaviour are considered for the last seven days, and the security system can be implemented more or less accordingly. Sometimes the same reoccurrence may happen for abnormal behaviours, and then the model features out the more significant weight. The study of the above paper gives knowledge about the smart monitoring of Big Data in Surveillance Video and memory utilisation. Even though the system can quickly identify and update the anomalous situation, this model will be more useful for identification of culprits of an event occurred than avoiding the abnormal condition.

2.2. Unsupervised learning and Knowledge Attaining in Surveillance System

Nawaratne, Bandaragoda, Adikari, Alahakoon, De Silva, and Yu [3] present a methodology that consolidates significant level video investigation with self-learning, incremental knowledge attaining, and self-governing smart surveillance system. This methodology is fit for identifying variations that happen after some time and isolating anomalies from re-events, without the essential of a labelled data set. The analysis of the behaviour is a beneficial methodology for finding out the primary usage of irregular and regular behaviour patterns. A constant irregular movement detection method is introduced in [3] for the analysis of a group of people, which helps to analyse regular, irregular, or emergency in the group stream. In this methodology, the behaviour of movement is taken from movement vectors in the video clips. Depending on the world-wide group movement behaviour, and movement vectors, the unexpected conduct in social groups can be identified based on the approach given in [6]. The blend of an optical stream and background elimination procedures are utilised to extract movement at each picture outline and discard small movement vectors having noise, issues of compression, and non-stationary foundation pixels.

Using Microsoft Cognitive Service [7] the characterisation and automatic labelling of videos based on semantic matters can be performed. For example, the activities of a person, objects, or complex events in the video frame can be characterised and labelled as such. The framework is given by Google Cloud Vision [8], and the Tensorflow open-source provides a very significant level of video analysis like emotion, face, optical character recognition, logo identification, object detection, and unsuitable content identification [10].

The learning and thinking in human cognisance is a progressive method. A human builds up fundamental knowledge and information through real-time surroundings and gradually refines the understanding as new data shows up. Learning, revising, and unlearning is a usual and constant procedure in human cognisance [11]. The unsupervised-learning algorithm, Incremental Knowledge Acquisition and Self-Learning (IKASL) tend to update the labelled dataset by continuous learning, thereby keeps up reliability and versatility across the iterations of learning [11]. The IKASL algorithm has procedures for generalisation and learning which help in knowledge acquisition and self-study. By the unsupervised learning method, the data set and features learned from the IKASL algorithm will be exploiting the video exploration capacity of Microsoft Cognitive Services API. The algorithms working process is explained in the following subsections.

The recorded labels created by Microsoft Cognitive Services API for frames of videos are used to build the feature space. From the video frames, the details of feature space comprise of keywords for identifying the activities, behaviours, and various objects are present. Figure 2 shows two video frames taken at time intervals from a typical checkpoint. According to [3], the keywords are identified in figure 2(a) using Activity Recognition [12]. Figure 2(b) is having the video frame taken from the same video checkpoint after 8 seconds, and the keywords are identified. Different methods are used for identifying this feature space like object recognition method [13] to infer the idea of the video frame, computer visual service provider [8] that study the video frames to generate a description or its combinations.



Figure 2. Two video frames are taken at time intervals from a typical checkpoint

IKASL algorithm uses a layered structure of n layers, and each layer consists of learning and generalisation sub-layers. The clustering of the input feature vector is included as a Growing Self-Organising Map (GSOM) [13] component in the sub-layer called the learning layer. The encoded summarised representation of the next learning layer is done in the generalisation sub-layer, and this layer will become the initial layer of the following learning layer, as shown in figure 3.

In the basic structure and model of the IKASL algorithm [3], the initial layer consists of 4 nodes with the random initialisation of weights. Initially, the sequences of data sets are given as tuples, in which the selection of a successful node is decided based on distance measurements like Euclidean or Cosine distance. The weight adjustment of a successful node is made according to the input tuple, which will be further adjusted by the nearby nodes. The learning in every layer is done by the GSOM algorithm [13]. After an iteration of learning, every successful node in a layer is identified and generalised up to an extent, and the outcome will return two significant results in the learning

procedure. The learning knowledge is passed to neighbouring nodes as successful nodes of weight vectors. In the generalisation sub-layer, the fuzzy integral will consolidate the weight vectors into a single one. The knowledge obtained from the learning and self-organisation process is provided in the generalisation sub-layer. Several iterations of learning are performed in the data set using learning and generalisation sub-layers.

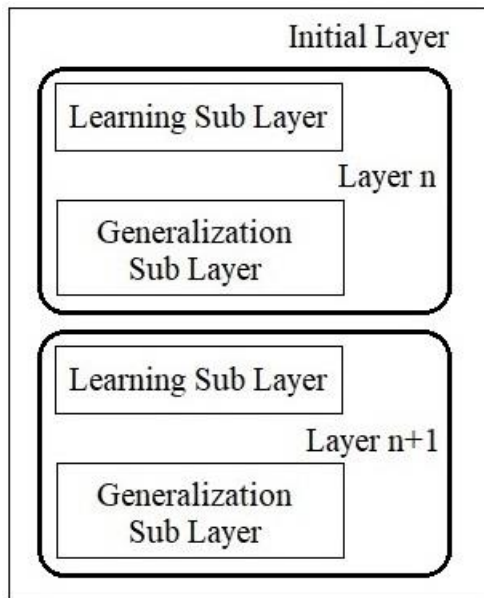


Figure 3. IKASL Model

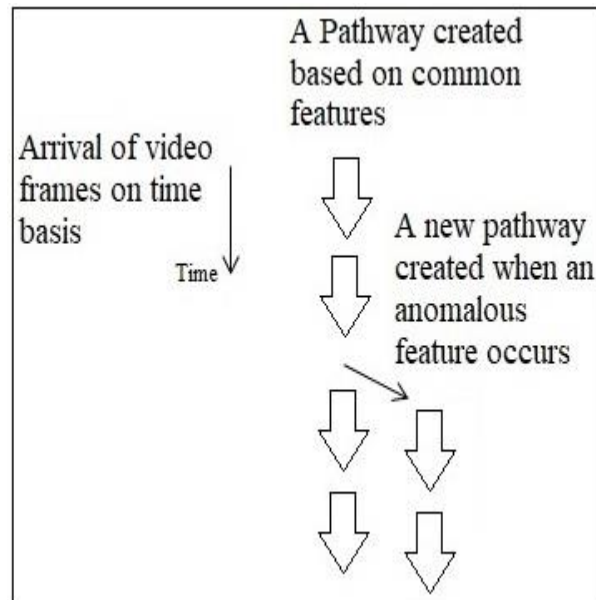


Figure 4. IKASL Pathway Generation

Using the IKASL algorithm, incremental learning of a new pathway is dependent on clusters recognised in learning sub-layers as in Figure 4. New pathways might be produced if there are new features in the input given to the algorithm and to classify the cluster according to the input tuples or feature vectors. For instance, consider figure 2, in which video frames are given to the IKASL algorithm in the order of occurrence, and each pathway represents the feature vector based on the keywords which form a cluster. If the feature vector has common features, then all the video frames will be in the same cluster, and if there is a substantial change in a feature vector, the new pathway will be created. For instance, if a video frame consists of danger features like fire, the IKASL algorithm will create a new pathway with new feature vectors and keyword sets. If the danger feature is cleared in the subsequent frame, the feature vector will move back to the original pathway using the IKASL algorithm.

The new pathway creation using IKASL algorithm can be used for various video surveillance applications like anomaly detection, identification of unknown objects, intrusion detection, observation of road accidents, and safety alerts generation when needed. The model can be extended by adding more dataset to extract more feature vectors.

2.3 Continual Learning for Anomaly Detection

Keval Doshi and Yasin Yilmaz in [5] propose an Anomaly detection method using transfer learning [14] and continual learning [15] which helps to reduce the training complexity. The model uses feature extraction based on neural networks for transfer learning, and for decision making continual learning statistical k-nearest neighbour (kNN). In transfer learning, a model which is developed for one task may be reused for another related task. Whereas Continuous learning with automatic incremental development of models is termed as Continuous Learning, and this helps to solve complex problems.

This Machine Learning model will update the Prediction model smoothly using task and data thereby enabling reuse and retrain useful knowledge and skills. The use of transfer learning will considerably reduce the training complexity. The statistical framework for sequential anomaly detection can perform continual and few-shot learning of video data, which help to evaluate video anomaly detection datasets and real surveillance video inputs.

The anomalous events are usually circumstantial, for example, a problem in the practical situation of riding a bike through the road is a normal situation whereas riding through a pedestrian way will be an anomalous situation. So in practical, the nominal event needs to be updated continually, and the decisions are updated using Deep Neural Networks (DNNs) [16], which uses the sequential nature of video anomaly detection.

The features from the video are evaluated, and feature vectors of each object having attributes like motion, location and appearance are created. The weights are applied to features which adjust according to the relative importance of each feature category during feature selection. To obtain the features from a real-time video stream, a pre-trained real-time object detection system like You Only Look Once (YOLO) [17] is used. The proposed model uses YOLOv3, which gives a bounding box (location) and class probabilities (appearance) for each video frame. The centre of the box is monitored for obtaining the area and location features. From the test videos, the objects deviating from nominal paths or new classes will help to identify anomalies related to spatial information. For temporal information, the hypothesis used is, for any motion anomaly the probability distribution of the optical flow of the frame would alter. For this, the mean, variance and high order statistics skewness, and kurtosis is extracted, which is used to represent irregularity and sharpness of the probability distribution. A feature vector is constructed for each time by combining the motion, location and appearance features [5]. For Anomaly Detection with minimum delays, a nonparametric sequential anomaly detection algorithm [5] is used. The Euclidean k Nearest Neighbor (kNN) distance is used to identify the local interactions between nominal data points. The hypothesis of the statistical distance between the nominal features and anomalous instances will be further away from the k nearest neighbour is used here.

During the testing phase, the feature vectors are extracted for each object, and kNN distance is computed. The instantaneous frame-level anomaly is evaluated, and running decision statistics for each time is calculated. For nominal data, the value of decision statistics will always be a negative value, and anomalous data will get a positive value. If the anomalous data persists to exist for n frames for successive time periods the value of the decision statistics will grow. Furthermore, if it exceeds the threshold, the frames will be labelled as anomalous. The test will continue with another set of frames and label the frames.

The continual learning is used to test statistical results and labelling the frames. The model proposes a human-in-the-loop approach to classify the label as true or false alarms from time to time. This sequential update will continually learn new data and prevent from wrong classification of abnormal alarms. This model avoids retraining of deep learning methods from scratch and disastrous forgetting. The model can be extended to work effectively on challenging situations like various weather conditions occurring dynamically.

2.4 Ensemble Learning using Bagging and Inception-v3 for Anomaly Detection

Yumna Zahid, Muhammad Atif Tahir and Muhammad Nouman Durrani in [9] propose an Anomaly Detection method by extracting video features using Inception-v3 deep learning network which avoids segmentation. The model is implemented by transfer learning the Convolution Neural Network (CNN) [19]. The Homogeneous bagging ensemble of 3-Layer Fully Connected (FC) Network is used for classification of anomalous events.

For unsupervised feature extraction of video, a widely used pre-trained Inception-v3 image recognition model is used. The Inception-v3 network provides video level and frame-level feature representations. Bagging Ensemble is used for randomly selecting data points from the training data, and to build the training model. A Fully Connected 3-Layer Neural Network is used as the base model for training. The Fully connected base classifier consists of ReLu function and Sigmoid function in the first and last layer respectively for the prediction score calculation. The confidence score from various models is integrated using an average operation called Bagging Ensemble algorithm [9].

The Fully connected 3- Layer Neural Network [9] is used for the calculation of the prediction score with an Adagrad optimizer [9]. The model will binary classify the video input into nominal or anomalous according to the probability score. A high probability score indicates the anomalous events and a low score indicates nominal events. The model also eliminates the necessity of segment videos for feature extraction. The videos are randomly selected, and each iteration is having a batch size of 60. The learning rate is kept as 0.001, and the hinge loss function is used in binary Support Vector Machine (SVM) classification. The dataset used for training consists of real world anomalous and nominal data. This model provides a good accuracy level for real time video anomaly detection and can be extended for fine-grained classification.

3. Conclusion

This paper studies some of the novel techniques used for automatic Anomaly Detection in a video surveillance system. In this paper, a comprehensive description of various Anomaly Detection Methods that exists in real video streaming is discussed. Anomaly detection in real video streaming needs a fast and novel model, which uses the previously learned knowledge and continual learned knowledge. For developing a fast and automatic model, novel techniques in Neural Networks is an elegant solution. Convolution Neural Networks [19] and Convolutional Long Short Term Memory (LSTM) [18] will provide motion and appearance feature identification effectively. The Convolutional LSTM can be used for sequence prediction problems with spatial inputs like images or video. Generative Adversarial Networks can generate internal scene representation from a given frame. Convolutional Autoencoder (CAE) [21] is another widely used neural network used in real-time video anomaly detection. The CAE apply the reconstruction error of each frame with an anomaly score.

By implementing a reliable model for automatic Anomaly Detection, the smart security system can work more efficiently and ensure the safety of people effectively. The applications of Neural Networks are developing rapidly and are yet to explore more in the area of computer vision and automatic anomaly detection in real-time videos.

4. References

- [1] W Heng, T Jiang, and W Gao 2019 How to Assess the Quality of Compressed Surveillance Videos Using Face Recognition *IEEE Transactions on Circuits and Systems for Video Technology* **29** 2229-43
- [2] Z Shao, J Cai, and Z Wang 2018 Smart Monitoring Cameras Driven Intelligent Processing to Big Surveillance Video Data *IEEE Transactions on Big Data* **4** 105-16
- [3] R Nawaratne, T Bandaragoda, A Adikari, D Alahakoon, D De Silva, and X Yu 2017 Incremental knowledge acquisition and self-learning for autonomous video surveillance *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society* 4790-95
- [4] L Bazzani, M Cristani, A Perina, and V Murino 2012 Multiple-shot person re-identification by chromatic and epitomic analyses *Pattern Recognit. Lett.*, **33** 898–903
- [5] K Doshi and Y Yilmaz 2020 Continual Learning for Anomaly Detection in Surveillance Videos *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* **10** 1025-34

- [6] I R de Almeida and C R Jung 2013 Change Detection in Human Crowds Proceedings of the 2013 XXVI Conference on Graphics, Patterns and Images 63–69
- [7] Cognitive Services—Intelligence Applications —2020 *Microsoft Azure [Online]* Available link <https://azure.microsoft.com/enin/services/cognitive-services> [last Accessed 14-Feb-2020]
- [8] Vision API - Image Content Analysis 2020 *Google Cloud Platform [Online]* Available link <https://cloud.google.com/vision/> [last Accessed 14- Feb-2020]
- [9] Y Zahid, M A Tahir and M N Durrani 2020 Ensemble Learning Using Bagging And Inception-V3 For Anomaly Detection In Surveillance Videos *IEEE International Conference on Image Processing* 588-92
- [10] Comparing image tagging services 2020 *Google Vision, Microsoft Cognitive Services, Amazon Rekognition, and Clarifai, Filestack Blog, [Online]* Available link <https://blog.filestack.com/thoughts-andknowledge/comparing-google-vision-microsoft-cognitive-amazonrekognition-clarifai/> [Last Accessed 15-Feb-2020]
- [11] D D Silva and D Alahakoon 2010 Incremental knowledge acquisition and self-learning from text *The 2010 International Joint Conference on Neural Networks (IJCNN)* 1–8
- [12] Collective Activity Dataset 2020 [Online] Available link <http://vhosts.eecs.umich.edu/vision/activity-dataset.html> [Last Accessed: 15-Feb-2020]
- [13] D Alahakoon, S K Halgamuge, and B Srinivasan 2000 Dynamic self organising maps with controlled growth for knowledge discovery *IEEE Trans. Neural Networks* **11** 601–14
- [14] Jason Brownlee 2017 [Online] A Gentle Introduction to Transfer Learning for Deep Learning *Deep Learning for Computer Vision* Available link <https://machinelearningmastery.com/transfer-learning-for-deep-learning/> [Last Accessed 29-Nov-2020]
- [15] Vincenzo Lomonaco 2017 [Online] Why Continual Learning is the key towards Machine Intelligence *Continual AI* Available link <https://medium.com/continual-ai/why-continuous-learning-is-the-key-towards-machine-intelligence-1851cb57c308> [Last Accessed 29-Nov-2020]
- [16] Michael Nielsen 2015 *Neural Networks and Deep Learning* (Determination Press) Available link <http://neuralnetworksanddeeplearning.com/faq.html> [Last Accessed 30-Nov-2020]
- [17] Redmon, Joseph and Farhadi, Ali 2018 YOLOv3: An Incremental Improvement *arXiv* Available link <https://pjreddie.com/darknet/yolo/> [Last Accessed 30-Nov-2020]
- [18] Jason Brownlee 2017 CNN Long Short-Term Memory Networks *Machine Learning Mastery - Long Short-Term Memory Networks* Available link <https://machinelearningmastery.com/cnn-long-short-term-memory-networks/> [Last Accessed 29- Nov-2020]
- [19] Brownlee 2017 How Do Convolutional Layers Work in Deep Learning Neural Jason Networks? *Machine Learning Mastery- Deep Learning for Computer Vision* Available link <https://machinelearningmastery.com/convolutional-layers-for-deep-learning-neural-networks/> [Last Accessed 29- Nov-2020]
- [20] Jason Brownlee 2019 A Gentle Introduction to Generative Adversarial Networks (GANs) *Machine Learning Mastery-Generative Adversarial Networks* Available link <https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/> [Last Accessed 29- Nov-2020]
- [21] Manasses Ribeiro, Andre Eugenio Lazzaretti and Heitor Silverio Lopes 2018 A study of deep convolutional auto-encoders for anomaly detection in videos *Pattern Recognition Letters* **105** 13-22