Journal of Big Data

SURVEY PAPER Open Access

Intelligent video surveillance: a review through deep learning techniques for crowd analysis

G. Sreenu* and M. A. Saleem Durai

*Correspondence:
gsreenug@gmail.com
VIT, Vellore 632014, Tamil Nadu, India

Abstract

Big data applications are consuming most of the space in industry and research area. Among the widespread examples of big data, the role of video streams from CCTV cameras is equally important as other sources like social media data, sensor data, agriculture data, medical data and data evolved from space research. Surveillance videos have a major contribution in unstructured big data. CCTV cameras are implemented in all places where security having much importance. Manual surveillance seems tedious and time consuming. Security can be defined in different terms in different contexts like theft identification, violence detection, chances of explosion etc. In crowded public places the term security covers almost all type of abnormal events. Among them violence detection is difficult to handle since it involves group activity. The anomalous or abnormal activity analysis in a crowd video scene is very difficult due to several real world constraints. The paper includes a deep rooted survey which starts from object recognition, action recognition, crowd analysis and finally violence detection in a crowd environment. Majority of the papers reviewed in this survey are based on deep learning technique. Various deep learning methods are compared in terms of their algorithms and models. The main focus of this survey is application of deep learning techniques in detecting the exact count, involved persons and the happened activity in a large crowd at all climate conditions. Paper discusses the underlying deep learning implementation technology involved in various crowd video analysis methods. Real time processing, an important issue which is yet to be explored more in this field is also considered. Not many methods are there in handling all these issues simultaneously. The issues recognized in existing methods are identified and summarized. Also future direction is given to reduce the obstacles identified. The survey provides a bibliographic summary of papers from ScienceDirect, IEEE Xplore and ACM digital library.

Keywords: Big data, Video surveillance, Deep learning, Crowd analysis

Bibliographic Summary of papers in different digital repositories

Bibliographic summary about published papers under the area "Surveillance video analysis through deep learning" in digital repositories like ScienceDirect, IEEExplore and ACM are graphically demonstrated.

ScienceDirect

SceinceDirect lists around 1851 papers. Figure 1 demonstrates the year wise statistics.

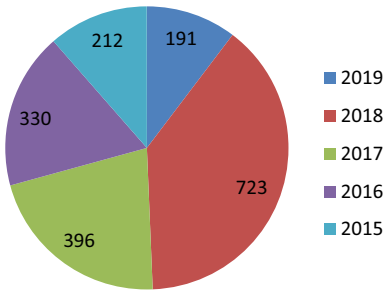

Fig. 1 Year wise paper statistics of "surveillance video analysis by deep learning", in ScienceDirect

Table 1 Title of 25 papers published in ScienceDirect

1	SVAS: Surveillance Video Analysis System [1]
2	Jointly learning perceptually heterogeneous features for blind 3D video quality assessment [2]
3	Learning to detect video events from zero or very few video examples [3]
4	Learning an event-oriented and discriminative dictionary based on an adaptive label-consistent K-SVD method for event detection in soccer videos [4]
5	Towards efficient and objective work sampling: Recognizing workers' activities in site surveillance videos with two-stream convolutional networks [5]
6	Dairy goat detection based on Faster R-CNN from surveillance video [6]
7	Performance evaluation of deep feature learning for RGB-D image/video classification [7]
8	Surveillance scene representation and trajectory abnormality detection using aggregation of multiple concepts [8]
9	Human Action Recognition using 3D convolutional neural networks with 3D Motion Cuboids in Surveillance Videos [9]
10	Neural networks based visual attention model for surveillance videos [10]
11	Application of deep learning for object detection [11]
12	A study of deep convolutional auto-encoders for anomaly detection in videos [12]
13	A novel deep multi-channel residual networks-based metric learning method for moving human localization in video surveillance [13]
14	Video surveillance systems-current status and future trends [14]
15	Enhancing transportation systems via deep learning: a survey [15]
16	Pedestrian tracking by learning deep features [16]
17	Action recognition using spatial-optical data organization and sequential learning framework [17]
18	Video pornography detection through deep learning techniques and motion information [18]
19	Deep learning to frame objects for visual target tracking [19]
20	Boosting deep attribute learning via support vector regression for fast moving crowd counting [20]
21	D-STC: deep learning with spatio-temporal constraints for train drivers detection from videos [21]
22	A robust human activity recognition system using smartphone sensors and deep learning [22]
23	Regional deep learning model for visual tracking [23]
24	Fog computing enabled cost-effective distributed summarization of surveillance videos for smart cities [24]
25	SIFT and tensor based object detection and classification in videos using deep neural networks [25]

Table 1 list title of 25 papers published under same area.

Table 2 gives the list of journals in ScienceDirect where above mentioned papers are published.

Keywords always indicate the main disciplines of the paper. An analysis is conducted through keywords used in published papers. Table 3 list the frequency of most frequently used keywords.

Table 2 List of journals

No: of papers	Journal
19	Neurocomputing
14	Pattern Recognition Letters
11	Pattern Recognition
10	Journal of Visual Communication and Image Representation
7	Expert Systems with Applications
5	Procedia Computer Science

Table 3 Usage frequency of keywords

Frequency	Keywords
41	Deep learning
11	Video surveillance
10	Convolutional neural network
9	Action recognition
7	Computer vision
7	Person re-identification
6	Convolutional neural networks
5	CNN
4	Activity recognition
4	Faster R-CNN
4	Machine learning
4	Surveillance
4	Video

ACM

ACM digital library includes 20,975 papers in the given area. The table below includes most recently published surveillance video analysis papers under deep learning field. Table 4 lists the details of published papers in the area.

IEEE Xplore

Table 5 shows details of published papers in the given area in IEEEXplore digital library.

Violence detection among crowd

The above survey presents the topic surveillance video analysis as a general topic. By going more deeper into the area more focus is given to violence detection in crowd behavior analysis.

Table 6 lists papers specific to "violence detection in crowd behavior" from above mentioned three journals.

Introduction

Artificial intelligence paves the way for computers to think like human. Machine learning makes the way more even by adding training and learning components. The availability of huge dataset and high performance computers lead the light to deep learning concept,

Table 4 Bibliographic summary of papers in ACM digital library

Author	Title	Keywords	Journal	Year
Zeng Yu and Tianrui Li and Ning Yu and Yi Pan and Hongmei Chen and Bing Liu	Reconstruction of hidden representation for robust feature extraction [26]	Deep architectures, auto-encoders, feature representation, reconstruction of hidden representation, unsupervised learning	ACM Trans. Intell. Syst. Technol.	2019
Rahim Mammadli and Felix Wolf and Ali Jannesari	The art of getting deep neural networks in shape [27]	Deep neural networks, computer vision, parallel processing	ACM Trans. Archit. Code Optim.	2019
Tinghui Zhou and Richard Tucker and John Flynn and Graham Fyffe and Noah Snavely	Stereo magnification: learning view synthesis using multiplane images [28]	Deep learning, view extrapolation	ACM Trans. Graph.	2018
Zipei Fan and Xuan Song and Tianqi Xia and Renhe Jiang and Ryosuke Shibasaki and Ritsu Sakuramachi	Online deep ensemble learning for predicting citywide human mobility [29]	Deep learning, ensemble learning, human mobility modeling, intelligent surveillance, urban computing	Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.	2018
Rana Hanocka and Noa Fish and Zhenhua Wang and Raja Giryes and Shachar Fleishman and Daniel Cohen-Or	ALIGNet: partial-shape agnostic alignment via unsupervised learning [30]	Deep learning, self-supervised learning, shape deformation	ACM Trans. Graph.	2018
Mengwei Xu and Feng Qian and Qiaozhu Mei and Kang Huang and Xuanzhe Liu	DeepType: on-device deep learning for input personalization service with minimal privacy concern [31]	Deep Learning, Mobile Computing, Personalization	Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.	2018
Thomas E. Potok and Catherine Schuman and Steven Young and Robert Patton and Federico Spedalieri and Jeremy Liu and Ke-Thia Yao and Garrett Rose and Gangotree Chakma	A study of complex deep learning networks on high-performance, neuromorphic, and quantum computers [32]	Deep learning, high-performance computing, neuromorphic computing, quantum computing	J. Emerg. Technol. Comput. Syst.	2018
Samira Pouyanfar and Saad Sadiq and Yilin Yan and Haiman Tian and Yudong Tao and Maria Presa Reyes and Mei-Ling Shyu and Shu-Ching Chen and S. S. Iyengar	A survey on deep learning: algorithms, techniques, and applications [33]	Deep learning, big data, distributed processing, machine learning, neural networks, survey	ACM Comput. Surv.	2018
Yonglong Tian and Guang-He Lee and Hao He and Chen-Yu Hsu and Dina Katabi	RF-based fall monitoring using convolutional neural networks [34]	Deep learning, Device–free, Fall Detection	Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.	2018
Probir Roy and Shuaiwen Leon Song and Sriram Krishnamoorthy and Abhinav Vishnu and Dipanjan Sengupta and Xu Liu	NUMA-Caffe: NUMA-aware deep learning neural networks [35]	Deep learning, NUMA, neural network, stochastic gradient descent	ACM Trans. Archit. Code Optim.	2018
Charles Lovering and Anqi Lu and Cuong Nguyen and Huyen Nguyen and David Hurley and Emmanuel Agu	Fact or fiction [36]	Deep learning, natural language processing, sentiment analysis, social collaboration, subjectivity classification, text classification, web system	Proc. ACM Hum.-Comput. Interact.	2018

Table 4 (continued)

Author	Title	Keywords	Journal	Year
Heli Ben-Hamu and Haggai Maron and Itay Kezurer and Gal Avineri and Yaron Lipman	Multi-chart generative surface modeling [37]	Deep learning, generative adveserial networks, shape generation	ACM Trans. Graph.	2018
Weifeng Ge and Bingchen Gong and Yizhou Yu	Image super-resolution via deterministic-stochastic synthesis and local statistical rectification [38]	Deep learning, deterministic component, image super-resolution, local correlation matrix, local gram matrix, stochastic component	ACM Trans. Graph.	2018
Peter Hedman and Julien Philip and True Price and Jan-Michael Frahm and George Drettakis and Gabriel Brostow	Deep blending for free-viewpoint image-based rendering [39]	Deep learning, free-viewpoint, image-based rendering	ACM Trans. Graph.	2018
Kalaivani Sundararajan and Damon L. Woodard	Deep learning for biometrics: a survey [40]	Deep learning, autoencoders, convolutional neural networks, deep belief nets, face recognition, feature learning, speaker recognition	ACM Comput. Surv.	2018
Hyungjun Kim and Taesu Kim and Jinseok Kim and Jae-Joon Kim	Deep neural network optimized to resistive memory with nonlinear current–voltage characteristics [41]	Deep neural network, I-V nonlinearity, nonvolatile memory, perceptron	J. Emerg. Technol. Comput. Syst.	2018
Cheng Wang and Haojin Yang and Christoph Meinel	Image captioning with deep bidirectional LSTMs and multi-task learning [42]	Deep learning, LSTM, image captioning, multimodal representations, mutli-task learning	ACM Trans. Multimedia Comput. Commun. Appl.	2018
Shuochao Yao and Yiran Zhao and Huajie Shao and Aston Zhang and Chao Zhang and Shen Li and Tarek Abdelzaher	RDeepSense: reliable deep mobile computing models with uncertainty estimations [43]	Deep Learning, Internet-of-Things, Mobile Computing, Reliability, Uncertainty Estimation	Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.	2018
Dongyu Liu and Weiwei Cui and Kai Jin and Yuxiao Guo and Huamin Qu	DeepTracker: visualizing the training process of convolutional neural networks [44]	Deep learning, correlation analysis, multiple time series, training process, visual analytics	ACM Trans. Intell. Syst. Technol.	2018
Li Yi and Haibin Huang and Difan Liu and Evangelos Kalogerakis and Hao Su and Leonidas Guibas	Deep part induction from articulated object pairs [45]	Deep learning, differentiable sequential RANSAC, motion based part segmentation, shape correspondences	ACM Trans. Graph.	2018
Nanxuan Zhao and Ying Cao and Rynson W. H. Lau	What characterizes personalities of graphic designs? [46]	Deep learning, graphic design, personality	ACM Trans. Graph.	2018
Jiwei Tan and Xiaojun Wan and Hui Liu and Jianguo Xiao	QuoteRec: toward quote recommendation for writing [47]	Deep learning, LSTM, document recommendation, quote recommendation	ACM Trans. Inf. Syst.	2018
Yanru Qu and Bohui Fang and Weinan Zhang and Ruiming Tang and Minzhe Niu and Huifeng Guo and Yong Yu and Xiuqiang He	Product-based neural networks for user response prediction over multi-field categorical data [48]	Deep learning, product-based neural network, recommender system	ACM Trans. Inf. Syst.	2018

Table 4 (continued)

Author	Title	Keywords	Journal	Year
Kangxue Yin and Hui Huang and Daniel Cohen-Or and Hao Zhang	P2P-NET: bidirectional point displacement net for shape transform [49]	Deep neural network, point cloud processing, point set transform, point-wise displacement	ACM Trans. Graph.	2018
Shuochao Yao and Yiran Zhao and Huajie Shao and Chao Zhang and Aston Zhang and Shaohan Hu and Dongxin Liu and Shengzhong Liu and Lu Su and Tarek Abdelzaher	SenseGAN: enabling deep learning for internet of things with a semi-supervised framework [50]	Deep Learning, GAN, Internet-of-Things, Mobile Computing, Semi-Supervised Learning	Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.	2018
Shunsuke Saito and Liwen Hu and Chongyang Ma and Hikaru Ibayashi and Linjie Luo and Hao Li	3D hair synthesis using volumetric variational autoencoders [51]	Deep generative model, hair synthesis, single-view modeling, volumetric variational autoencoder	ACM Trans. Graph.	2018
Anpei Chen and Minye Wu and Yingliang Zhang and Nianyi Li and Jie Lu and Shenghua Gao and Jingyi Yu	Deep surface light fields [52]	Deep Neural Network, Image-based Rendering, Real-time Rendering	Proc. ACM Comput. Graph. Interact. Tech.	2018

Table 5 Bibliographic summary of papers in IEEE Xplore

Document title	Publication_ Year	Funding information
Sparse coding guided spatiotemporal feature learning for abnormal event detection in large videos [53]	2019	National Nature Science Foundation of China; National Youth Top-notch Talent Support Program
Rejecting motion outliers for efficient crowd anomaly detection [54]	2019	Ministry of Science, ICT and Future Planning
Deep multi-view feature learning for person re-identification [55]	2018	National Natural Science Foundation of China; Yunnan Natural Science Funds; Guangdong Natural Science Funds; Yunnan University
Image-to-video person re-identification with temporally memorized similarity learning [56]	2018	National Natural Science Foundation of China; NSFC-Shenzhen Robotics Projects; Natural Science Foundation of Guangdong Province; Fundamental Research Funds for the Central Universities; ZTE Corporation
Fight recognition in video using hough forests and 2D convolutional neural network [57]	2018	Ministerio de EconomÃa y Competitividad
Anomalous sound detection using deep audio representation and a BLSTM network for audio surveillance of roads [58]	2018	National Natural Science Foundation of China; National Laboratory of Pattern Recognition
Convolutional neural networks based fire detection in surveillance videos [59]	2018	National Research Foundation of Korea (NRF); Korea government (MSIP)
Action recognition in video sequences using deep bi-directional LSTM with CNN features [60]	2018	National Research Foundation of Korea Grant; Korea Government (MSIP)
A deep spatiotemporal perspective for understanding crowd behavior [61]	2018	
Road traffic conditions classification based on multilevel filtering of image content using convolutional neural networks [62]	2018	
Indoor person identification using a low-power FMCW radar [63]	2018	Ghent University; imec; Fund for Scientific Research-Flanders (FWO-Flanders)
Support vector machine approach to fall recognition based on simplified expression of human skeleton action and fast detection of start key frame using torso angle [64]	2018	
Person re-identification using hybrid representation reinforced by metric learning [65]	2018	
Evolving head tracking routines with brain programming [66]	2018	Consejo Nacional de Ciencia y Tecnología; https://doi.org/10.13039/501100004963-seventh Framework Programme of the European Union through the Marie Curie International Research Staff Scheme, FP-PEOPLE-2013-IRSES, Project Analysis and Classification of Mental States of Vigilance with Evolutionary Computation; https://doi.org/10.13039/501100003089-centro de Investigación Científica y de Educación Superior de Ensenada, Baja California; TecNM Project 6474.18-P, "Navegación de robots móviles como un sistema adaptativo complejo."
Natural language description of video streams using task-specific feature encoding [67]	2018	Basic Science Research Program through the National Research Foundation of Korea (NRF); Ministry of Education
Background subtraction using multiscale fully convolutional network [68]	2018	National Science Foundation of China
Face verification via learned representation on feature-rich video frames [69]	2017	MEITY, India, NVIDIA GPU grant, and Infosys CAI, IIIT-Delhi; IBM Ph.D. fellowship
Violent activity detection with transfer learning method [70]	2017	

Table 5 (continued)

Document title	Publication_ Year	Funding information
Unsupervised sequential outlier detection with deep architectures [71]	2017	
High-level feature extraction for classification and person re-identification [72]	2017	
An ensemble of invariant features for person reidentification [73]	2017	
Facial expression recognition using salient features and convolutional neural network [74]	2017	Research Council of Norway as a part of the Multimodal Elderly Care Systems Project
Deep head pose: gaze-direction estimation in multimodal video [75]	2015	
Deep reconstruction models for image set classification [76]	2015	SIRF; University of Western Australia; ARC

Table 6 Papers specific to crowd behavior analysis, under deep learning

Title	Year	Digital repository
A review on classifying abnormal behavior in crowd scene [77]	2019	ScienceDirect
Crowd behavior analysis from fixed and moving cameras [78]	2019	
Zero-shot crowd behavior recognition [79]	2017	
The analysis of high density crowds in videos [80]	2017	
Computer vision based crowd disaster avoidance system: a survey [81]	2017	
Deep learning for scene-independent crowd analysis [82]	2017	
Fast face detection in violent video scenes [83]	2016	
Rejecting motion outliers for efficient crowd anomaly detection [54]	2019	IEEEXplore
Deep metric learning for crowdedness regression [84]	2018	
A deep spatiotemporal perspective for understanding crowd behavior [61]	2018	
Crowded scene understanding by deeply learned volumetric slices [85]	2017	
Crowd scene understanding from video: a survey [86]	2017	ACM

which extract automatically features or the factors of variation that distinguishes objects from one another. Among the various data sources which contribute to terabytes of big data, video surveillance data is having much social relevance in today's world. The widespread availability of surveillance data from cameras installed in residential areas, industrial plants, educational institutions and commercial firms contribute towards private data while the cameras placed in public places such as city centers, public conveyances and religious places contribute to public data.

Analysis of surveillance videos involves a series of modules like object recognition, action recognition and classification of identified actions into categories like anomalous or normal. This survey giving specific focus on solutions based on deep learning architectures. Among the various architectures in deep learning, commonly used models for surveillance analysis are CNN, auto-encoders and their combination. The paper Video surveillance systems-current status and future trends [14] compares 20 papers published recently in the area of surveillance video analysis. The paper begins with identifying the main outcomes of video analysis. Application areas where surveillance cameras are unavoidable are discussed. Current status and trends in video analysis are revealed through

literature review. Finally the vital points which need more consideration in near future are explicitly stated.

Surveillance video analysis: relevance in present world

The main objectives identified which illustrate the relevance of the topic are listed out below.

1. Continuous monitoring of videos is difficult and tiresome for humans.
2. Intelligent surveillance video analysis is a solution to laborious human task.
3. Intelligence should be visible in all real world scenarios.
4. Maximum accuracy is needed in object identification and action recognition.
5. Tasks like crowd analysis are still needs lot of improvement.
6. Time taken for response generation is highly important in real world situation.
7. Prediction of certain movement or action or violence is highly useful in emergency situation like stampede.
8. Availability of huge data in video forms.

The majority of papers covered for this survey give importance to object recognition and action detection. Some papers are using procedures similar to a binary classification that whether action is anomalous or not anomalous. Methods for Crowd analysis and violence detection are also included. Application areas identified are included in the next section.

Application areas identified

The contexts identified are listed as application areas. Major part in existing work provides solutions specifically based on the context.

1. Traffic signals and main junctions
2. Residential areas
3. Crowd pulling meetings
4. Festivals as part of religious institutions
5. Inside office buildings

Among the listed contexts crowd analysis is the most difficult part. All type of actions, behavior and movement are needed to be identified.

Surveillance video data as Big Data

Big video data have evolved in the form of increasing number of public cameras situated towards public places. A huge amount of networked public cameras are positioned around worldwide. A heavy data stream is generated from public surveillance cameras that are creatively exploitable for capturing behaviors. Considering the huge amount of data that can be documented over time, a vital scenario is facility for data warehousing and data analysis. Only one high definition video camera can produce around 10 GB of data per day [87].

The space needed for storing large amount of surveillance videos for long time is difficult to allot. Instead of having data, it will be useful to have the analysis result. That will result in reduced storage space. Deep learning techniques are involved with two main components; training and learning. Both can be achieved with highest accuracy through huge amount of data.

Main advantages of training with huge amount of data are listed below. It's possible to adapt variety in data representation and also it can be divided into training and testing equally. Various data sets available for analysis are listed below. The dataset not only includes video sequences but also frames. The analysis part mainly includes analysis of frames which were extracted from videos. So dataset including images are also useful.

The datasets widely used for various kinds of application implementation are listed in below Table 7. The list is not specific to a particular application though it is specified against an application.

Methods identified/reviewed other than deep learning

Methods identified are mainly classified into two categories which are either based on deep learning or not based on deep learning. This section is reviewing methods other than deep learning.

SVAS deals with automatic recognition and deduction of complex events. The event detection procedure consists of mainly two levels, low level and high level. As a result of low level analysis people and objects are detected. The results obtained from low level are used for high level analysis that is event detection. The architecture proposed in the model includes five main modules. The five sections are

Table 7 Various datasets

Dataset	Type/purpose	Model/schema used
ImageNet2012	Images	
PASCAL VOC	Images	
Frames Labeled In Cinema (FLIC)	Popular holywood movies	
Leeds Sports Pose (LSP)	Sports people gathered from FLICKR	
CAVIAR	Used for event detection of surveillance domain	Threshold Model used for spatio temporal motion analysis and Bag of Actions for reducing search space [1]
BEHAVE	Used for event detection of surveillance domain	Threshold Model used for spatio temporal motion analysis and Bag of Actions for reducing search space [1]
YTO	Videos collected from YouTube	
i-LIDS sterile zone	People detection	Intrusion detection system with global features [91]
PETS 2001	Images	Intrusion detection system with global features [91]
MoSIFT	Movie dataset	
STIP	Hockey dataset	
MediaEval 2013 dataset	Collection of movies	
UCSD pedestrian	Pedestrian walkway	Convolutional auto-encoder model [12]

- Event model learning
- Action model learning
- Action detection
- Complex event model learning
- Complex event detection

Interval-based spatio-temporal model (IBSTM) is the proposed model and is a hybrid event model. Other than this methods like Threshold models, Bayesian Networks, Bag of actions and Highly cohesive intervals and Markov logic networks are used.

SVAS method can be improved to deal with moving camera and multi camera data set. Further enhancements are needed in dealing with complex events specifically in areas like calibration and noise elimination.

Multiple anomalous activity detection in videos [88] is a rule based system. The features are identified as motion patterns. Detection of anomalous events are done either by training the system or by following dominant set property.

The concept of dominant set where events are detected as normal based on dominant behavior and anomalous events are decided based on less dominant behavior. The advantage of rule based system is that easy to recognize new events by modifying some rules. The main steps involved in a recognition system are

- Pre processing
- Feature extraction
- Object tracking
- Behavior understanding

As a preprocessing system video segmentation is used. Background modeling is implemented through Gaussian Mixture Model (GMM). For object recognition external rules are required. The system is implemented in Matlab 2014. The areas were more concentration further needed are doubtful activities and situations where multiple object overlapping happens.

Mining anomalous events against frequent sequences in surveillance videos from commercial environments [89] focus on abnormal events linked with frequent chain of events. The main result in identifying such events is early deployment of resources in particular areas. The implementation part is done using Matlab, Inputs are already noticed events and identified frequent series of events. The main investigation under this method is to recognize events which are implausible to chase given sequential pattern by fulfilling the user identified parameters.

The method is giving more focus on event level analysis and it will be interesting if pay attention at entity level and action level. But at the same time going in such granular level make the process costly.

Video feature descriptor combining motion and appearance cues with length invariant characteristics [90] is a feature descriptor. Many trajectory based methods have been used in abundant installations. But those methods have to face problems related with occlusions. As a solution to that, feature descriptor using optical flow based method.

As per the algorithm the training set is divided into snippet set. From each set images are extracted and then optical flow are calculated. The covariance is calculated from optical flow. One class SVM is used for learning samples. For testing also same procedure is performed.

The model can be extended in future by handling local abnormal event detection through proposed feature which is related with objectness method.

Multiple Hierarchical Dirichlet processes for anomaly detection in Traffic [91] is mainly for understanding the situation in real world traffic. The anomalies are mainly due to global patterns instead of local patterns. That include entire frame. Concept of super pixel is included. Super pixels are grouped into regions of interest. Optical flow based method is used for calculating motion in each super pixel. Points of interest are then taken out in active super pixel. Those interested points are then tracked by Kanade–Lucas–Tomasi (KLT) tracker.

The method is better the handle videos involving complex patterns with less cost. But not mentioning about videos taken in rainy season and bad weather conditions.

Intelligent video surveillance beyond robust background modeling [92] handle complex environment with sudden illumination changes. Also the method will reduce false alerts. Mainly two components are there. IDS and PSD are the two components.

First stage intruder detection system will detect object. Classifier will verify the result and identify scenes causing problems. Then in second stage problematic scene descriptor will handle positives generated from IDS. Global features are used to avoid false positives from IDS.

Though the method deals with complex scenes, it does not mentioning about bad weather conditions.

Towards abnormal trajectory and event detection in video surveillance [93] works like an integrated pipeline. Existing methods either use trajectory based approaches or pixel based approaches. But this proposal incorporates both methods. Proposal include components like

- Object and group tracking
- Grid based analysis
- Trajectory filtering
- Abnormal behavior detection using actions descriptors

The method can identify abnormal behavior in both individual and groups. The method can be enhanced by adapting it to work in real time environment.

RIMOC: a feature to discriminate unstructured motions: application to violence detection for video surveillance [94]. There is no unique definition for violent behaviors. Those kind of behaviors show large variances in body poses. The method works by taking the eigen values of histograms of optical flow.

The input video undergoes dense sampling. Local spatio temporal volumes are created around each sampled point. Those frames of STV are coded as histograms of optical flow. Eigen values are computed from this frame. The papers already published in surveillance area span across a large set. Among them methods which are unique in

either implementation method or the application for which it is proposed are listed in the below Table 8.

The methods already described and listed are able to perform following steps

- Object detection
- Object discrimination
- Action recognition

But these methods are not so efficient in selecting good features in general. The lag identified in methods was absence of automatic feature identification. That issue can be solved by applying concepts of deep learning.

The evolution of artificial intelligence from rule based system to automatic feature identification passes machine learning, representation learning and finally deep learning.

Real-time processing in video analysis

Real time Violence Detection Framework for Football Stadium comprising of Big Data Analysis and deep learning through Bidirectional LSTM [103] predicts violent behavior of crowd in real time. The real time processing speed is achieved through SPARK frame work. The model architecture includes Apache spark framework, spark streaming, Histogram of oriented Gradients function and bidirectional LSTM. The model takes

Table 8 Summary of different techniques in video analysis

Title	Method	Tool	Data set
Scenario-based query processing for video-surveillance archives [95]	Query processing system and inverted tracking	VSQL	PETS 2006 and PETS 2007
Activity retrieval in large surveillance videos [96]	Dynamic matching algorithm	Query creation GUI	Pets, Mit traffic
Integrated video object tracking with applications in trajectory-based event detection [97]	Adaptive particle sampling and Kalman filtering	Not mentioned	PETS 2001 test dataset1, camera 1
Evidential event inference in transport video surveillance [98]	Using spatio-temporal correlations for reasoning	Jones and Viola face detector	Own data set
Abnormal event detection based on analysis of movement information of video sequence [99]	Optical flow and Hidden Markov model	Not mentioned	UMN, PETS
Anomalous entities detection and localization in pedestrian flows [100]	Gaussian kernel based feature integration and R-CRF model based classification	Not mentioned	UCSD, UMN, UCD
Snatch theft detection in unconstrained surveillance videos using action attribute modelling [101]	A large GMM called universal attribute model		Own Dataset Snatch 1.0
ArchCam: real time expert system for suspicious behaviour detection in ATM site [102]	Image processing technique	NVIDIA Tegra TX1 SoC 340 with quad core ARM processor and 256 cores GPU	Videos under a mock ATM setup

stream of videos from diverse sources as input. The videos are converted in the form of non overlapping frames. Features are extracted from this group of frames through HOG FUNCTION. The images are manually modeled into different groups. The BDLSTM is trained through all these models. The SPARK framework handles the streaming data in a micro batch mode. Two kinds of processing are there like stream and batch processing.

Intelligent video surveillance for real-time detection of suicide attempts [104] is an effort to prevent suicide by hanging in prisons. The method uses depth streams offered by an RGB-D camera. The body joints' points are analyzed to represent suicidal behavior.

Spatio-temporal texture modeling for real-time crowd anomaly detection [105]. Spatio temporal texture is a combination of spatio temporal slices and spatio temporal volumes. The information present in these slices are abstracted through wavelet transforms. A Gaussian approximation model is applied to texture patterns to distinguish normal behaviors from abnormal behaviors.

Deep learning models in surveillance

Deep convolutional framework for abnormal behavior detection in a smart surveillance system [106] includes three sections.

- Human subject detection and discrimination
- A posture classification module
- An abnormal behavior detection module

The models used for above three sections are, Correspondingly

- You only look once (YOLO) network
- VGG-16 Net
- Long short-term memory (LSTM)

For object discrimination Kalman filter based object entity discrimination algorithm is used. Posture classification study recognizes 10 types of poses. RNN uses back propagation through time (BPTT) to update weight.

The main issue identified in the method is that similar activities like pointing and punching are difficult to distinguish.

Detecting Anomalous events in videos by learning deep representations of appearance and motion [107] proposes a new model named as AMDN. The model automatically learns feature representations. The model uses stacked de-noising auto encoders for learning appearance and motion features separately and jointly. After learning, multiple one class SVM's are trained. These SVM predict anomaly score of each input. Later these scores are combined and detect abnormal event. A double fusion framework is used. The computational overhead in testing time is too high for real time processing.

A study of deep convolutional auto encoders for anomaly detection in videos [12] proposes a structure that is a mixture of auto encoders and CNN. An auto encoder includes an encoder part and decoder part. The encoder part includes convolutional and pooling layers, the decoding part include de convolutional and unpool layers. The architecture

allows a combination of low level frames withs high level appearance and motion features. Anomaly scores are represented through reconstruction errors.

Going deeper with convolutions [108] suggests improvements over traditional neural network. Fully connected layers are replaced by sparse ones by adding sparsity into architecture. The paper suggests for dimensionality reduction which help to reduce the increasing demand for computational resources. Computing reductions happens with 1×1 convolutions before reaching 5×5 convolutions. The method is not mentioning about the execution time. Along with that not able to make conclusion about the crowd size that the method can handle successfully.

Deep learning for visual understanding: a review [109], reviewing the fundamental models in deep learning. Models and technique described were CNN, RBM, Autoencoder and Sparse coding. The paper also mention the drawbacks of deep learning models such as people were not able to understand the underlying theory very well.

Deep learning methods other than the ones discussed above are listed in the following Table 9.

The methods reviewed in above sections are good in automatic feature generation. All methods are good in handling individual entity and group entities with limited size.

Majority of problems in real world arises among crowd. Above mentioned methods are not effective in handling crowd scenes. Next section will review intelligent methods for analyzing crowd video scenes.

Review in the field of crowd analysis

The review include methods which are having deep learning background and methods which are not having that background.

Spatial temporal convolutional neural networks for anomaly detection and localization in crowded scenes [114] shows the problem related with crowd analysis is challenging because of the following reasons

- Large number of pedestrians
- Close proximity
- Volatility of individual appearance
- Frequent partial occlusions

Table 9 Deep learning methods

Title	Deep learning model	Algorithms used
A deep convolutional neural network for video sequence background subtraction [110]	CNN	SuBSENSE algorithm, Flux Tensor algorithm
Tracking people in RGBD videos using deep learning and motion clues [111]	Deep convolutional neural network	Probabilistic tracking algorithm.
Deep CNN based binary hash video representations for face retrieval [112]	Deep CNN	Low-rank discriminative binary hashing, back-propagation (BP) algorithm
DAAL: deep activation-based attribute learning for action recognition in depth videos [113]	1D temporal CNN, 2D spatial CNN, 3D volumetric CNN	Deep activation-based attribute learning algorithm (DAAL)

- Irregular motion pattern in crowd
- Dangerous activities like crowd panic
- Frame level and pixel level detection

The paper suggests optical flow based solution. The CNN is having eight layers. Training is based on BVLC caffe. Random initialization of parameters is done and system is trained through stochastic gradient descent based back propagation. The implementation part is done by considering four different datasets like UCSD, UMN, Subway and finally U-turn. The details of implementation regarding UCSD includes frame level and pixel level criterion. Frame level criterion concentrates on temporal domain and pixel level criterion considers both spatiial and temporal domain. Different metrics to evaluate performance includes EER (Equal Error Rate) and Detection Rate (DR).

Online real time crowd behavior detection in video sequences [115] suggests FSCB, behavior detection through feature tracking and image segmentation. The procedure involves following steps

- Feature detection and temporal filtering
- Image segmentation and blob extraction
- Activity detection
- Activity map
- Activity analysis
- Alarm

The main advantage is no need of training stage for this method. The method is quantitatively analyzed through ROC curve generation. The computational speed is evaluated through frame rate. The data set considered for experiments include UMN, PETS2009, AGORASET and Rome Marathon.

Deep learning for scene independent crowd analysis [82] proposes a scene independent method which include following procedures

- Crowd segmentation and detection
- Crowd tracking
- Crowd counting
- Pedestrian travelling time estimation
- Crowd attribute recognition
- Crowd behavior analysis
- Abnormality detection in a crowd

Attribute recognition is done thorugh a slicing CNN. By using a 2D CNN model learn appearance features then represent it as a cuboid. In the cuboid three temporal filters are identified. Then a classifier is applied on concatenated feature vector extracted from cuboid. Crowd counting and crowd density estimation is treated as a regression problem. Crowd attribute recognition is applied on WWW Crowd dataset. Evaluation metrics used are AUC and AP.

The analysis of High Density Crowds in videos [80] describes methods like data driven crowd analysis and density aware tracking. Data driven analysis learn crowd motion patterns from large collection of crowd videos through an off line manner. Learned pattern can be applied or transferred in applications. The solution includes a two step procedure. Global crowded scene matching and local crowd patch matching. Figure 2 illustrates the two step procedure.

The database selected for experimental evaluation includes 520 unique videos with 720×480 resolutions. The main evaluation is to track unusual and unexpected actions of individuals in a crowd. Through experiments it is proven that data driven tracking is better than batch mode tracking. Density based person detection and tracking include steps like baseline detector, geometric filtering and tracking using density aware detector.

A review on classifying abnormal behavior in crowd scene [77] mainly demonstrates four key approaches such as Hidden Markov Model (HMM), GMM, optical flow and STT. GMM itself is enhanced with different techniques to capture abnormal behaviours. The enhanced versions of GMM are

- GMM
- GMM and Markov random field
- Gaussian poisson mixture model and
- GMM and support vector machine

GMM architecture includes components like local descriptor, global descriptor, classifiers and finally a fusion strategy. The distinction between normal and and abnormal behaviour is evaluated based on Mahalanobis distance method. GMM–MRF model mainly divided into two sections where first section identifies motion pttern through

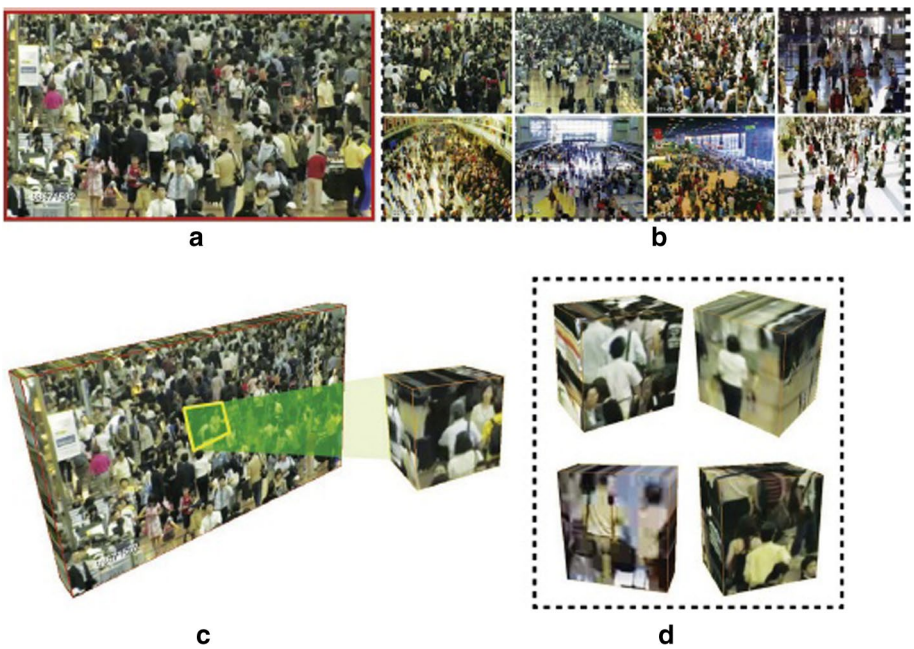

Fig. 2 **a** Test video, **b** results of global matching, **c** a query crowd patch, **d** matching crowd patches [80]

GMM and crowd context modelling is done through MRF. GPMM adds one extra feture such as count of occurrence of observed behaviour. Also EM is used for training at later stage of GPMM. GMM–SVM incorporate features such as crowd collectiveness, crowd density, crowd conflict etc. for abnormality detection.

HMM has also variants like

- GM-HMM
- SLT-HMM
- MOHMM
- HM and OSVMs

Hidden Markov Model is a density aware detection method used to detect motion based abnormality. The method generates foreground mask and perspective mask through ORB detector. GM-HMM involves four major steps. First step GMBM is used for identifying foreground pixels and further lead to development of blobs generation. In second stage PCA–HOG and motion HOG are used for feature extraction. The third stage applies k means clustering to separately cluster features generated through PCA–HOG and motion–HOG. In final stage HMM processes continuous information of moving target through the application of GM. In SLT-HMM short local trajectories are used along with HMM to achieve better localization of moving objects. MOHMM uses KLT in first phase to generate trajectories and clustering is applied on them. Second phase uses MOHMM to represent the trajectories to define usual and unusual frames. OSVM uses kernel functions to solve the nonlinearity problem by mapping high dimensional features in to a linear space by using kernel function.

In optical flow based method the enhancements made are categorized into following techniques such as HOFH, HOFME, HMOFP and MOFE.

In HOFH video frames are divided into several same size patches. Then optical flows are extracted. It is divided into eight directions. Then expectation and variance features are used to calculate optical flow between frames. HOFME descriptor is used at the final stage of abnormal behaviour detection. As the first step frame difference is calculated then extraction of optical flow pattern and finally spatio temporal description using HOFME is completed. HMOFP Extract optical flow from each frame and divided into patches. The optical flows are segmented into number of bins. Maximum amplitude flows are concatenated to form global HMOFP. MOFE method convert frames into blobs and optical flow in all the blobs are extracted. These optical flow are then clustered into different groups. In STT, crowd tracking and abnormal behaviour detection is done through combing spatial and temporal dimensions of features.

Crowd behaviour analysis from fixed and moving cameras [78] covers topics like microscopic and macroscopic crowd modeling, crowd behavior and crowd density analysis and datasets for crowd behavior analysis. Large crowds are handled through macroscopic approaches. Here agents are handled as a whole. In microscopic approaches agents are handled individually. Motion information to represent crowd can be collected through fixed and moving cameras. CNN based methods like end-to-end deep CNN, Hydra-CNN architecture, switching CNN, cascade CNN architecture, 3D CNN and spatio temporal CNN are discussed for crowd behaviour analysis. Different datasets useful

specifically for crowd behaviour analysis are also described in the chapter. The metrics used are MOTA (multiple person tracker accuracy) and MOTP (multiple person tracker precision). These metrics consider multi target scenarios usually present in crowd scenes. The dataset used for experimental evaluation consists of UCSD, Violent-flows, CUHK, UCF50, Rodriguez's, The mall and finally the worldExpo's dataset.

Zero-shot crowd behavior recognition [79] suggests recognizers with no or little training data. The basic idea behind the approach is attribute-context cooccurrence. Prediction of behavioural attribute is done based on their relationship with known attributes. The method encompass different steps like probabilistic zero shot prediction. The method calculates the conditional probability of known to original appropriate attribute relation. The second step includes learning attribute relatedness from Text Corpora and Context learning from visual co-occurrence. Figure 3 shows the illustration of results.

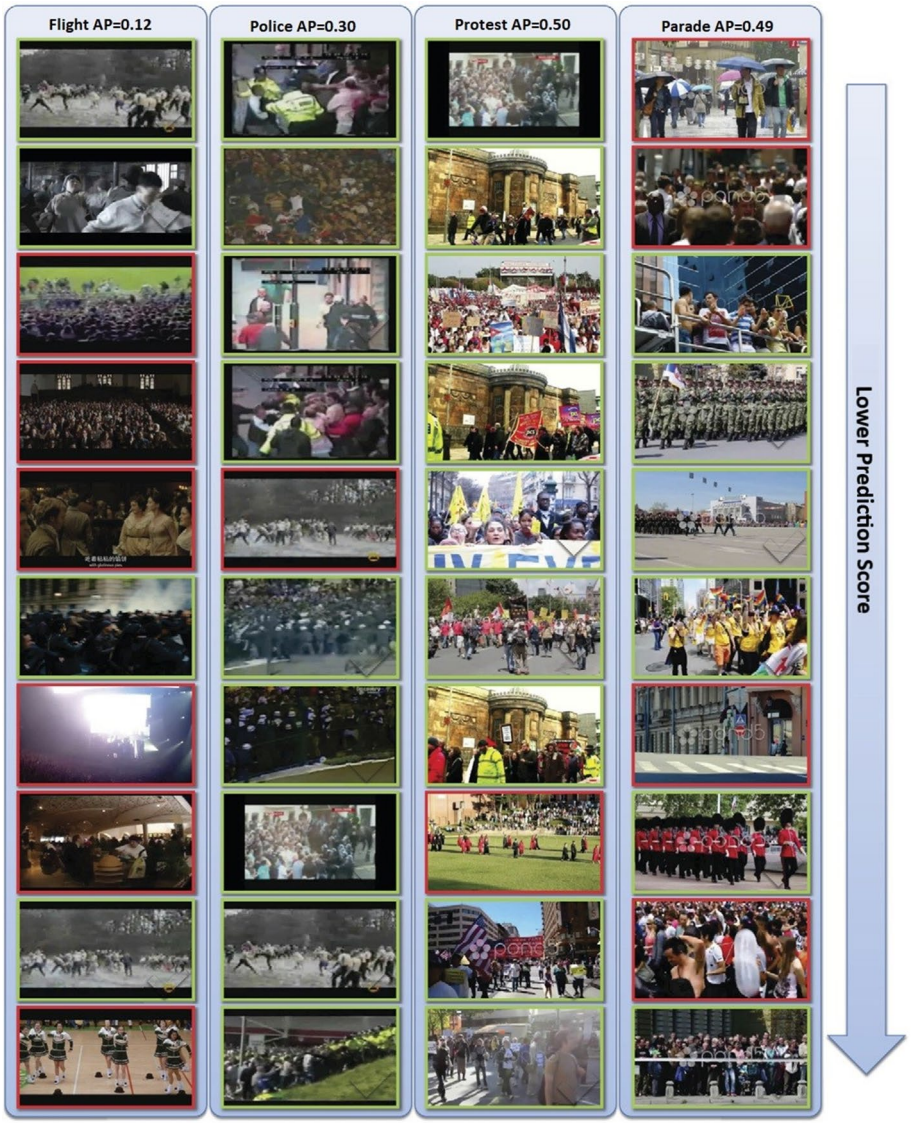

Fig. 3 Demonstration of crowd videos ranked in accordance with prediction values [79]

Computer vision based crowd disaster avoidance system: a survey [81] covers different perspectives of crowd scene analysis such as number of cameras employed and target of interest. Along with that crowd behavior analysis, people count, crowd density estimation, person re identification, crowd evacuation, and forensic analysis on crowd disaster and computations on crowd analysis. A brief summary about benchmarked datasets are also given.

Fast Face Detection in Violent Video Scenes [83] suggests an architecture with three steps such as violent scene detector, a normalization algorithm and finally a face detector. ViF descriptor along with Horn–Schunck is used for violent scene detection, used as optical flow algorithm. Normalization procedure includes gamma intensity correction, difference Gauss, Local Histogram Coincidence and Local Normal Distribution. Face detection involve mainly two stages. First stage is segmenting regions of skin and the second stage check each component of face.

Rejecting Motion Outliers for Efficient Crowd Anomaly Detection [54] provides a solution which consists of two phases. Feature extraction and anomaly classification. Feature extraction is based on flow. Different steps involved in the pipeline are input video is divided into frames, frames are divided into super pixels, extracting histogram for each super pixel, aggregating histograms spatially and finally concatenation of combined histograms from consecutive frames for taking out final feature. Anomaly can be detected through existing classification algorithms. The implementation is done through UCSD dataset. Two subsets with resolution 158×238 and 240×360 are present. The normal behavior was used to train k means and KUGDA. The normal and abnormal behavior is used to train linear SVM. The hardware part includes Artix 7 xc7a200t FPGA from Xilinx, Xilinx IST and XPower Analyzer.

Deep Metric Learning for Crowdedness Regression [84] includes deep network model where learning of features and distance measurements are done concurrently. Metric learning is used to study a fine distance measurement. The proposed model is implemented through Tensorflow package. Rectified linear unit is used as an activation function. The training method applied is gradient descent. Performance is evaluated through mean squared error and mean absolute error. The WorldExpo dataset and the Shanghai Tech dataset are used for experimental evaluation.

A Deep Spatiotemporal Perspective for Understanding Crowd Behavior [61] is a combination of convolution layer and long short-term memory. Spatial informations are captured through convolution layer and temporal motion dynamics are confined through LSTM. The method forecasts the pedestrian path, estimate the destination and finally categorize the behavior of individuals according to motion pattern. Path forecasting technique includes two stacked ConvLSTM layers by 128 hidden states. Kernel of ConvLSTM size is 3×3, with a stride of 1 and zeropadding. Model takes up a single convolution layer with a 1×1 kernel size. Crowd behavior classification is achieved through a combination of three layers namely an average spatial pooling layer, a fully connected layer and a softmax layer.

Crowded Scene Understanding by Deeply Learned Volumetric Slices [85] suggests a deep model and different fusion approaches. The architecture involves convolution layers, global sum pooling layer and fully connected layers. Slice fusion and weight sharing schemes are required by the architecture. A new multitask learning deep model is

Table 10 Crowd analysis methods

Title	Method	Tool	Data set
Measurement of congestion and intrinsic risk in pedestrian crowds [116]	Use computational mesh	Not mentioned	Not mentioned
A classification method based on streak flow for abnormal crowd behaviors [117]	Streak flow based on fluid mechanics, ViBe algorithm, classification method,	Streakline	ViF
An intelligent decision computing paradigm for crowd monitoring in the smart city [118]	Extended Kalman filtering approach, Agent motion-based learning model, SIFT feature descriptor, EM algorithm	Not mentioned	The dataset is prepared with surveillance cameras using 60 mm × 120 mm lens from Puri rath yatra festival
Learning deep event models for crowd anomaly detection [119]	Deep neural network, PCANet, deep GMM	Not mentioned	UCSD Ped1 Dataset, Avenue Dataset

projected to equally study motion features and appearance features and successfully join them. A new concept of crowd motion channels are designed as input to the model. The motion channel analyzes the temporal progress of contents in crowd videos. The motion channels are stirred by temporal slices that clearly demonstrate the temporal growth of contents in crowd videos. In addition, we also conduct wide-ranging evaluations by multiple deep structures with various data fusion and weights sharing schemes to find out temporal features. The network is configured with convlutional layer, pooling layer and fully connected layer with activation functions such as rectified linear unit and sigmoid function. Three different kinds of slice fusion techniques are applied to measure the efficiency of proposed input channels.

Crowd Scene Understanding from Video A survey [86] mainly deals with crowd counting. Different approaches for crowd counting are categorized into six. Pixel level analysis, texture level analysis, object level analysis, line counting, density mapping and joint detection and counting. Edge features are analyzed through pixel level analysis. Image patches are analysed through texture level analysis. Object level analysis is more accurate compared to pixel and texture analysis. The method identifies individual subjects in a scene. Line counting is used to take the count of people crossed a particular line.

Table 10 will discuss some more crowd analysis methods.

Results observed from the survey and future directions

The accuracy analysis conducted for some of the above discussed methods based on various evaluation criteria like AUC, precision and recall are discussed below.

Rejecting Motion Outliers for Efficient Crowd Anomaly Detection [54] compare different methods as shown in Fig. 4. KUGDA is a classifier proposed in Rejecting Motion Outliers for Efficient Crowd Anomaly Detection [54].

Fast Face Detection in Violent Video Scenes [83] uses a ViF descriptor for violence scene detection. Figure 5 shows the evaluation of an SVM classifier using ROC curve.

Figure 6 represents a comparison of detection performance which is conducted by different methods [80]. The comparison shows the improvement of density aware detector over other methods.

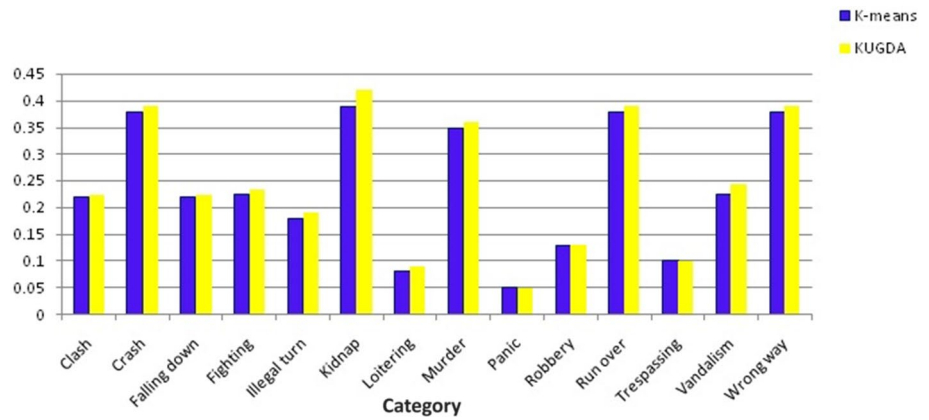

Fig. 4 Comparing KUGDA with K-means [54]

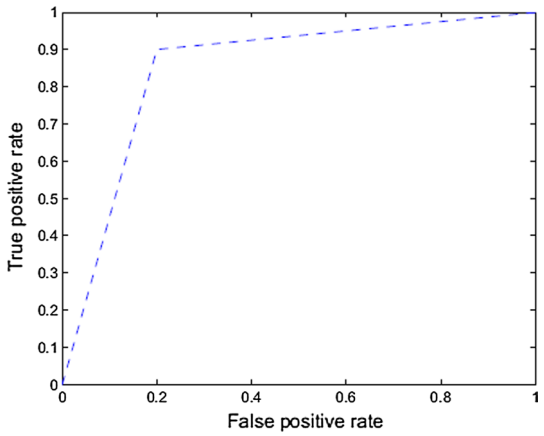

Fig. 5 Receiver operating characteristics of a classifier with ViF descriptor [83]

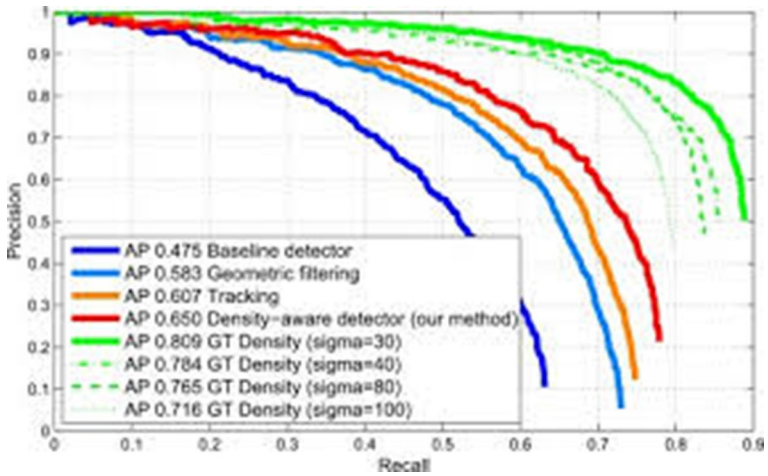

Fig. 6 Comparing detection performance of density aware detector with different methods [80]

As an analysis of existing methods the following shortcomings were identified. Real world problems are having following objectives like

- Time complexity
- Bad weather conditions
- Real world dynamics
- Occulsions
- Overlapping of objects

Existing methods were handling the problems separately. No method handles all the objectives as features in a single proposal.

To handle effective intelligent crowd video analysis in real time the method should be able to provide solutions to all these problems. Traditional methods are not able to generate efficient economic solution in a time bounded manner.

The availability of high performance computational resource like GPU allows implementation of deep learning based solutions for fast processing of big data. Existing deep learning architectures or models can be combined by including good features and removing unwanted features.

Conclusion

The paper reviews intelligent surveillance video analysis techniques. Reviewed papers cover wide variety of applications. The techniques, tools and dataset identified were listed in form of tables. Survey begins with video surveillance analysis in general perspective, and then finally moves towards crowd analysis. Crowd analysis is difficult in such a way that crowd size is large and dynamic in real world scenarios. Identifying each entity and their behavior is a difficult task. Methods analyzing crowd behavior were discussed. The issues identified in existing methods were listed as future directions to provide efficient solution.

Abbreviations

SVAS: Surveillance Video Analysis System; IBSTM: Interval-Based Spatio-Temporal Model; KLT: Kanade–Lucas–Tomasi; GMM: Gaussian Mixture Model; SVM: Support Vector Machine; DAAL: Deep activation-based attribute learning; HMM: Hidden Markov Model; YOLO: You only look once; LSTM: Long short-term memory; AUC: Area under the curve; ViF: Violent flow descriptor.

Acknowledgements

Not applicable.

Authors' contributions

GS and MASD selected and analyzed different papers for getting more in depth view about current scenarios of the problem and its solutions. Both authors read and approved the final manuscript.

Funding

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Received: 7 December 2018 Accepted: 28 May 2019
Published online: 06 June 2019

References

1. Kardas K, Cicekli NK. SVAS: surveillance video analysis system. Expert Syst Appl. 2017;89:343–61.
2. Wang Y, Shuai Y, Zhu Y, Zhang J. An P Jointly learning perceptually heterogeneous features for blind 3D video quality assessment. Neurocomputing. 2019;332:298–304 **(ISSN 0925-2312)**.
3. Tzelepis C, Galanopoulos D, Mezaris V, Patras I. Learning to detect video events from zero or very few video examples. Image Vis Comput. 2016;53:35–44 **(ISSN 0262-8856)**.
4. Fakhar B, Kanan HR, Behrad A. Learning an event-oriented and discriminative dictionary based on an adaptive label-consistent K-SVD method for event detection in soccer videos. J Vis Commun Image Represent. 2018;55:489–503 **(ISSN 1047-3203)**.
5. Luo X, Li H, Cao D, Yu Y, Yang X, Huang T. Towards efficient and objective work sampling: recognizing workers' activities in site surveillance videos with two-stream convolutional networks. Autom Constr. 2018;94:360–70 **(ISSN 0926-5805)**.
6. Wang D, Tang J, Zhu W, Li H, Xin J, He D. Dairy goat detection based on Faster R-CNN from surveillance video. Comput Electron Agric. 2018;154:443–9 **(ISSN 0168-1699)**.
7. Shao L, Cai Z, Liu L, Lu K. Performance evaluation of deep feature learning for RGB-D image/video classification. Inf Sci. 2017;385:266–83 **(ISSN 0020-0255)**.
8. Ahmed SA, Dogra DP, Kar S, Roy PP. Surveillance scene representation and trajectory abnormality detection using aggregation of multiple concepts. Expert Syst Appl. 2018;101:43–55 **(ISSN 0957-4174)**.
9. Arunnehru J, Chamundeeswari G, Prasanna Bharathi S. Human action recognition using 3D convolutional neural networks with 3D motion cuboids in surveillance videos. Procedia Comput Sci. 2018;133:471–7 **(ISSN 1877-0509)**.
10. Guraya FF, Cheikh FA. Neural networks based visual attention model for surveillance videos. Neurocomputing. 2015;149(Part C):1348–59 **(ISSN 0925-2312)**.
11. Pathak AR, Pandey M, Rautaray S. Application of deep learning for object detection. Procedia Comput Sci. 2018;132:1706–17 **(ISSN 1877-0509)**.
12. Ribeiro M, Lazzaretti AE, Lopes HS. A study of deep convolutional auto-encoders for anomaly detection in videos. Pattern Recogn Lett. 2018;105:13–22.
13. Huang W, Ding H, Chen G. A novel deep multi-channel residual networks-based metric learning method for moving human localization in video surveillance. Signal Process. 2018;142:104–13 **(ISSN 0165-1684)**.
14. Tsakanikas V, Dagiuklas T. Video surveillance systems-current status and future trends. Comput Electr Eng. In press, corrected proof, Available online 14 November 2017.
15. Wang Y, Zhang D, Liu Y, Dai B, Lee LH. Enhancing transportation systems via deep learning: a survey. Transport Res Part C Emerg Technol. 2018. https://doi.org/10.1016/j.trc.2018.12.004 **(ISSN 0968-090X)**.
16. Huang H, Xu Y, Huang Y, Yang Q, Zhou Z. Pedestrian tracking by learning deep features. J Vis Commun Image Represent. 2018;57:172–5 **(ISSN 1047-3203)**.
17. Yuan Y, Zhao Y, Wang Q. Action recognition using spatial-optical data organization and sequential learning framework. Neurocomputing. 2018;315:221–33 **(ISSN 0925-2312)**.
18. Perez M, Avila S, Moreira D, Moraes D, Testoni V, Valle E, Goldenstein S, Rocha A. Video pornography detection through deep learning techniques and motion information. Neurocomputing. 2017;230:279–93 **(ISSN 0925-2312)**.
19. Pang S, del Coz JJ, Yu Z, Luaces O, Díez J. Deep learning to frame objects for visual target tracking. Eng Appl Artif Intell. 2017;65:406–20 **(ISSN 0952-1976)**.
20. Wei X, Du J, Liang M, Ye L. Boosting deep attribute learning via support vector regression for fast moving crowd counting. Pattern Recogn Lett. 2017. https://doi.org/10.1016/j.patrec.2017.12.002.
21. Xu M, Fang H, Lv P, Cui L, Zhang S, Zhou B. D-stc: deep learning with spatio-temporal constraints for train drivers detection from videos. Pattern Recogn Lett. 2017. https://doi.org/10.1016/j.patrec.2017.09.040 **(ISSN 0167-8655)**.
22. Hassan MM, Uddin MZ, Mohamed A, Almogren A. A robust human activity recognition system using smartphone sensors and deep learning. Future Gener Comput Syst. 2018;81:307–13 **(ISSN 0167-739X)**.
23. Wu G, Lu W, Gao G, Zhao C, Liu J. Regional deep learning model for visual tracking. Neurocomputing. 2016;175:310–23 **(ISSN 0925-2312)**.
24. Nasir M, Muhammad K, Lloret J, Sangaiah AK, Sajjad M. Fog computing enabled cost-effective distributed summarization of surveillance videos for smart cities. J Parallel Comput. 2018. https://doi.org/10.1016/j.jpdc.2018.11.004 **(ISSN 0743-7315)**.
25. Najva N, Bijoy KE. SIFT and tensor based object detection and classification in videos using deep neural networks. Procedia Comput Sci. 2016;93:351–8 **(ISSN 1877-0509)**.
26. Yu Z, Li T, Yu N, Pan Y, Chen H, Liu B. Reconstruction of hidden representation for Robust feature extraction. ACM Trans Intell Syst Technol. 2019;10(2):18.
27. Mammadli R, Wolf F, Jannesari A. The art of getting deep neural networks in shape. ACM Trans Archit Code Optim. 2019;15:62.
28. Zhou T, Tucker R, Flynn J, Fyffe G, Snavely N. Stereo magnification: learning view synthesis using multiplane images. ACM Trans Graph. 2018;37:65
29. Fan Z, Song X, Xia T, Jiang R, Shibasaki R, Sakuramachi R. Online Deep Ensemble Learning for Predicting Citywide Human Mobility. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2018;2:105.
30. Hanocka R, Fish N, Wang Z, Giryes R, Fleishman S, Cohen-Or D. ALIGNet: partial-shape agnostic alignment via unsupervised learning. ACM Trans Graph. 2018;38:1.
31. Xu M, Qian F, Mei Q, Huang K, Liu X. DeepType: on-device deep learning for input personalization service with minimal privacy concern. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2018;2:197.
32. Potok TE, Schuman C, Young S, Patton R, Spedalieri F, Liu J, Yao KT, Rose G, Chakma G. A study of complex deep learning networks on high-performance, neuromorphic, and quantum computers. J Emerg Technol Comput Syst. 2018;14:19.

33. Pouyanfar S, Sadiq S, Yan Y, Tian H, Tao Y, Reyes MP, Shyu ML, Chen SC, Iyengar SS. A survey on deep learning: algorithms, techniques, and applications. ACM Comput Surv. 2018;51:92.
34. Tian Y, Lee GH, He H, Hsu CY, Katabi D. RF-based fall monitoring using convolutional neural networks. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2018;2:137.
35. Roy P, Song SL, Krishnamoorthy S, Vishnu A, Sengupta D, Liu X. NUMA-Caffe: NUMA-aware deep learning neural networks. ACM Trans Archit Code Optim. 2018;15:24.
36. Lovering C, Lu A, Nguyen C, Nguyen H, Hurley D, Agu E. Fact or fiction. Proc ACM Hum-Comput Interact. 2018;2:111.
37. Ben-Hamu H, Maron H, Kezurer I, Avineri G, Lipman Y. Multi-chart generative surface modeling. ACM Trans Graph. 2018;37:215
38. Ge W, Gong B, Yu Y. Image super-resolution via deterministic-stochastic synthesis and local statistical rectification. ACM Trans Graph. 2018;37:260
39. Hedman P, Philip J, Price T, Frahm JM, Drettakis G, Brostow G. Deep blending for free-viewpoint image-based rendering. ACM Trans Graph. 2018;37:257
40. Sundararajan K, Woodard DL. Deep learning for biometrics: a survey. ACM Comput Surv. 2018;51:65.
41. Kim H, Kim T, Kim J, Kim JJ. Deep neural network optimized to resistive memory with nonlinear current–voltage characteristics. J Emerg Technol Comput Syst. 2018;14:15.
42. Wang C, Yang H, Bartz C, Meinel C. Image captioning with deep bidirectional LSTMs and multi-task learning. ACM Trans Multimedia Comput Commun Appl. 2018;14:40.
43. Yao S, Zhao Y, Shao H, Zhang A, Zhang C, Li S, Abdelzaher T. RDeepSense: Reliable Deep Mobile Computing Models with Uncertainty Estimations. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2018;1:173.
44. Liu D, Cui W, Jin K, Guo Y, Qu H. DeepTracker: visualizing the training process of convolutional neural networks. ACM Trans Intell Syst Technol. 2018;10:6.
45. Yi L, Huang H, Liu D, Kalogerakis E, Su H, Guibas L. Deep part induction from articulated object pairs. ACM Trans Graph. 2018. https://doi.org/10.1145/3272127.3275027.
46. Zhao N, Cao Y, Lau RW. What characterizes personalities of graphic designs? ACM Trans Graph. 2018;37:116.
47. Tan J, Wan X, Liu H, Xiao J. QuoteRec: toward quote recommendation for writing. ACM Trans Inf Syst. 2018;36:34.
48. Qu Y, Fang B, Zhang W, Tang R, Niu M, Guo H, Yu Y, He X. Product-based neural networks for user response prediction over multi-field categorical data. ACM Trans Inf Syst. 2018;37:5.
49. Yin K, Huang H, Cohen-Or D, Zhang H. P2P-NET: bidirectional point displacement net for shape transform. ACM Trans Graph. 2018;37:152.
50. Yao S, Zhao Y, Shao H, Zhang C, Zhang A, Hu S, Liu D, Liu S, Su L, Abdelzaher T. SenseGAN: enabling deep learning for internet of things with a semi-supervised framework. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2018;2:144.
51. Saito S, Hu L, Ma C, Ibayashi H, Luo L, Li H. 3D hair synthesis using volumetric variational autoencoders. ACM Trans Graph. 2018. https://doi.org/10.1145/3272127.3275019.
52. Chen A, Wu M, Zhang Y, Li N, Lu J, Gao S, Yu J. Deep surface light fields. Proc ACM Comput Graph Interact Tech. 2018;1:14.
53. Chu W, Xue H, Yao C, Cai D. Sparse coding guided spatiotemporal feature learning for abnormal event detection in large videos. IEEE Trans Multimedia. 2019;21(1):246–55.
54. Khan MUK, Park H, Kyung C. Rejecting motion outliers for efficient crowd anomaly detection. IEEE Trans Inf Forensics Secur. 2019;14(2):541–56.
55. Tao D, Guo Y, Yu B, Pang J, Yu Z. Deep multi-view feature learning for person re-identification. IEEE Trans Circuits Syst Video Technol. 2018;28(10):2657–66.
56. Zhang D, Wu W, Cheng H, Zhang R, Dong Z, Cai Z. Image-to-video person re-identification with temporally memorized similarity learning. IEEE Trans Circuits Syst Video Technol. 2018;28(10):2622–32.
57. Serrano I, Deniz O, Espinosa-Aranda JL, Bueno G. Fight recognition in video using hough forests and 2D convolutional neural network. IEEE Trans Image Process. 2018;27(10):4787–97. https://doi.org/10.1109/tip.2018.2845742.
58. Li Y, Li X, Zhang Y, Liu M, Wang W. Anomalous sound detection using deep audio representation and a blstm network for audio surveillance of roads. IEEE Access. 2018;6:58043–55.
59. Muhammad K, Ahmad J, Mehmood I, Rho S, Baik SW. Convolutional neural networks based fire detection in surveillance videos. IEEE Access. 2018;6:18174–83.
60. Ullah A, Ahmad J, Muhammad K, Sajjad M, Baik SW. Action recognition in video sequences using deep bi-directional LSTM with CNN features. IEEE Access. 2018;6:1155–66.
61. Li Y. A deep spatiotemporal perspective for understanding crowd behavior. IEEE Trans Multimedia. 2018;20(12):3289–97.
62. Pamula T. Road traffic conditions classification based on multilevel filtering of image content using convolutional neural networks. IEEE Intell Transp Syst Mag. 2018;10(3):11–21.
63. Vandersmissen B, et al. indoor person identification using a low-power FMCW radar. IEEE Trans Geosci Remote Sens. 2018;56(7):3941–52.
64. Min W, Yao L, Lin Z, Liu L. Support vector machine approach to fall recognition based on simplified expression of human skeleton action and fast detection of start key frame using torso angle. IET Comput Vision. 2018;12(8):1133–40.
65. Perwaiz N, Fraz MM, Shahzad M. Person re-identification using hybrid representation reinforced by metric learning. IEEE Access. 2018;6:77334–49.
66. Olague G, Hernández DE, Clemente E, Chan-Ley M. Evolving head tracking routines with brain programming. IEEE Access. 2018;6:26254–70.
67. Dilawari A, Khan MUG, Farooq A, Rehman Z, Rho S, Mehmood I. Natural language description of video streams using task-specific feature encoding. IEEE Access. 2018;6:16639–45.
68. Zeng D, Zhu M. Background subtraction using multiscale fully convolutional network. IEEE Access. 2018;6:16010–21.

69. Goswami G, Vatsa M, Singh R. Face verification via learned representation on feature-rich video frames. IEEE Trans Inf Forensics Secur. 2017;12(7):1686–98.
70. Keçeli AS, Kaya A. Violent activity detection with transfer learning method. Electron Lett. 2017;53(15):1047–8.
71. Lu W, et al. Unsupervised sequential outlier detection with deep architectures. IEEE Trans Image Process. 2017;26(9):4321–30.
72. Feizi A. High-level feature extraction for classification and person re-identification. IEEE Sens J. 2017;17(21):7064–73.
73. Lee Y, Chen S, Hwang J, Hung Y. An ensemble of invariant features for person reidentification. IEEE Trans Circuits Syst Video Technol. 2017;27(3):470–83.
74. Uddin MZ, Khaksar W, Torresen J. Facial expression recognition using salient features and convolutional neural network. IEEE Access. 2017;5:26146–61.
75. Mukherjee SS, Robertson NM. Deep head pose: Gaze-direction estimation in multimodal video. IEEE Trans Multimedia. 2015;17(11):2094–107.
76. Hayat M, Bennamoun M, An S. Deep reconstruction models for image set classification. IEEE Trans Pattern Anal Mach Intell. 2015;37(4):713–27.
77. Afiq AA, Zakariya MA, Saad MN, Nurfarzana AA, Khir MHM, Fadzil AF, Jale A, Gunawan W, Izuddin ZAA, Faizari M. A review on classifying abnormal behavior in crowd scene. J Vis Commun Image Represent. 2019;58:285–303.
78. Bour P, Cribelier E, Argyriou V. Chapter 14—Crowd behavior analysis from fixed and moving cameras. In: Computer vision and pattern recognition, multimodal behavior analysis in the wild. Cambridge: Academic Press; 2019. pp. 289–322.
79. Xu X, Gong S, Hospedales TM. Chapter 15—Zero-shot crowd behavior recognition. In: Group and crowd behavior for computer vision. Cambridge: Academic Press; 2017:341–369.
80. Rodriguez M, Sivic J, Laptev I. Chapter 5—The analysis of high density crowds in videos. In: Group and crowd behavior for computer vision. Cambridge: Academic Press. 2017. pp. 89–113.
81. Yogameena B, Nagananthini C. Computer vision based crowd disaster avoidance system: a survey. Int J Disaster Risk Reduct. 2017;22:95–129.
82. Wang X, Loy CC. Chapter 10—Deep learning for scene-independent crowd analysis. In: Group and crowd behavior for computer vision. Cambridge: Academic Press; 2017. pp. 209–52.
83. Arceda VM, Fabián KF, Laura PL, Tito JR, Cáceres JG. Fast face detection in violent video scenes. Electron Notes Theor Comput Sci. 2016;329:5–26.
84. Wang Q, Wan J, Yuan Y. Deep metric learning for crowdedness regression. IEEE Trans Circuits Syst Video Technol. 2018;28(10):2633–43.
85. Shao J, Loy CC, Kang K, Wang X. Crowded scene understanding by deeply learned volumetric slices. IEEE Trans Circuits Syst Video Technol. 2017;27(3):613–23.
86. Grant JM, Flynn PJ. Crowd scene understanding from video: a survey. ACM Trans Multimedia Comput Commun Appl. 2017;13(2):19.
87. Tay L, Jebb AT, Woo SE. Video capture of human behaviors: toward a Big Data approach. Curr Opin Behav Sci. 2017;18:17–22 **(ISSN 2352-1546)**.
88. Chaudhary S, Khan MA, Bhatnagar C. Multiple anomalous activity detection in videos. Procedia Comput Sci. 2018;125:336–45.
89. Anwar F, Petrounias I, Morris T, Kodogiannis V. Mining anomalous events against frequent sequences in surveillance videos from commercial environments. Expert Syst Appl. 2012;39(4):4511–31.
90. Wang T, Qiao M, Chen Y, Chen J, Snoussi H. Video feature descriptor combining motion and appearance cues with length-invariant characteristics. Optik. 2018;157:1143–54.
91. Kaltsa V, Briassouli A, Kompatsiaris I, Strintzis MG. Multiple Hierarchical Dirichlet Processes for anomaly detection in traffic. Comput Vis Image Underst. 2018;169:28–39.
92. Cermeño E, Pérez A, Sigüenza JA. Intelligent video surveillance beyond robust background modeling. Expert Syst Appl. 2018;91:138–49.
93. Coşar S, Donatiello G, Bogorny V, Garate C, Alvares LO, Brémond F. Toward abnormal trajectory and event detection in video surveillance. IEEE Trans Circuits Syst Video Technol. 2017;27(3):683–95.
94. Ribeiro PC, Audigier R, Pham QC. Romaric Audigier, Quoc Cuong Pham, RIMOC, a feature to discriminate unstructured motions: application to violence detection for video-surveillance. Comput Vis Image Underst. 2016;144:121–43.
95. Şaykol E, Güdükbay U, Ulusoy Ö. Scenario-based query processing for video-surveillance archives. Eng Appl Artif Intell. 2010;23(3):331–45.
96. Castanon G, Jodoin PM, Saligrama V, Caron A. Activity retrieval in large surveillance videos. In: Academic Press library in signal processing. Vol. 4. London: Elsevier; 2014.
97. Cheng HY, Hwang JN. Integrated video object tracking with applications in trajectory-based event detection. J Vis Commun Image Represent. 2011;22(7):673–85.
98. Hong X, Huang Y, Ma W, Varadarajan S, Miller P, Liu W, Romero MJ, del Rincon JM, Zhou H. Evidential event inference in transport video surveillance. Comput Vis Image Underst. 2016;144:276–97.
99. Wang T, Qiao M, Deng Y, Zhou Y, Wang H, Lyu Q, Snoussi H. Abnormal event detection based on analysis of movement information of video sequence. Optik. 2018;152:50–60.
100. Ullah H, Altamimi AB, Uzair M, Ullah M. Anomalous entities detection and localization in pedestrian flows. Neurocomputing. 2018;290:74–86.
101. Roy D, Mohan CK. Snatch theft detection in unconstrained surveillance videos using action attribute modelling. Pattern Recogn Lett. 2018;108:56–61.
102. Lee WK, Leong CF, Lai WK, Leow LK, Yap TH. ArchCam: real time expert system for suspicious behaviour detection in ATM site. Expert Syst Appl. 2018;109:12–24.

103. Dinesh Jackson Samuel R, Fenil E, Manogaran G, Vivekananda GN, Thanjaivadivel T, Jeeva S, Ahilan A. Real time violence detection framework for football stadium comprising of big data analysis and deep learning through bidirectional LSTM. Comput Netw. 2019;151:191–200 **(ISSN 1389-1286)**.
104. Bouachir W, Gouiaa R, Li B, Noumeir R. Intelligent video surveillance for real-time detection of suicide attempts. Pattern Recogn Lett. 2018;110:1–7 **(ISSN 0167-8655)**.
105. Wang J, Xu Z. Spatio-temporal texture modelling for real-time crowd anomaly detection. Comput Vis Image Underst. 2016;144:177–87 **(ISSN 1077-3142)**.
106. Ko KE, Sim KB. Deep convolutional framework for abnormal behavior detection in a smart surveillance system. Eng Appl Artif Intell. 2018;67:226–34.
107. Dan X, Yan Y, Ricci E, Sebe N. Detecting anomalous events in videos by learning deep representations of appearance and motion. Comput Vis Image Underst. 2017;156:117–27.
108. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR). 2015.
109. Guo Y, Liu Y, Oerlemans A, Lao S, Lew MS. Deep learning for visual understanding: a review. Neurocomputing. 2016;187(26):27–48.
110. Babaee M, Dinh DT, Rigoll G. A deep convolutional neural network for video sequence background subtraction. Pattern Recogn. 2018;76:635–49.
111. Xue H, Liu Y, Cai D, He X. Tracking people in RGBD videos using deep learning and motion clues. Neurocomputing. 2016;204:70–6.
112. Dong Z, Jing C, Pei M, Jia Y. Deep CNN based binary hash video representations for face retrieval. Pattern Recogn. 2018;81:357–69.
113. Zhang C, Tian Y, Guo X, Liu J. DAAL: deep activation-based attribute learning for action recognition in depth videos. Comput Vis Image Underst. 2018;167:37–49.
114. Zhou S, Shen W, Zeng D, Fang M, Zhang Z. Spatial–temporal convolutional neural networks for anomaly detection and localization in crowded scenes. Signal Process Image Commun. 2016;47:358–68.
115. Pennisi A, Bloisi DD, Iocchi L. Online real-time crowd behavior detection in video sequences. Comput Vis Image Underst. 2016;144:166–76.
116. Feliciani C, Nishinari K. Measurement of congestion and intrinsic risk in pedestrian crowds. Transp Res Part C Emerg Technol. 2018;91:124–55.
117. Wang X, He X, Wu X, Xie C, Li Y. A classification method based on streak flow for abnormal crowd behaviors. Optik Int J Light Electron Optics. 2016;127(4):2386–92.
118. Kumar S, Datta D, Singh SK, Sangaiah AK. An intelligent decision computing paradigm for crowd monitoring in the smart city. J Parallel Distrib Comput. 2018;118(2):344–58.
119. Feng Y, Yuan Y, Lu X. Learning deep event models for crowd anomaly detection. Neurocomputing. 2017;219:548–56.

Publisher's Note