Informatics Bulletin, Faculty of Computers and Artificial Intelligence, Helwan University
Published Online *Vol 7 Issue 2, July 2025*
(https://fcihib.journals.ekb.eg)

A Comparative Study for Video Analytics Techniques based on Video Surveillance Cameras

Doaa Mabrouk [1,*], Manal A. Abdel-Fattah [2], Ahmed Taha [3]

[1] Software Engineering Department, Faculty of Engineering & Technology, Egyptian Chinese University, Cairo, Egypt
[2] Information Systems Department, Faculty of Computers and Artificial Intelligence, Helwan University, Helwan, Egypt
[3] Computer Science Department, Faculty of Computers and Artificial Intelligence, Benha University, Benha, Egypt
* dmabrouk@ecu.edu.eg

***Abstract*— Video analytics has recently become a fast-growing trend due to the heavy reliance on video surveillance cameras in many aspects that offer different applications in various domains, such as asset protection, violence detection, traffic monitoring, .etc. Several studies have addressed a variety of video analytic techniques from different perspectives. However, many challenges have been encountered, and more investigations are still required. This paper provides an exhaustive comparative study to investigate the diverse research efforts presented in video analytics based on surveillance cameras, with a detailed discussion and analysis of their main benefits and challenges. Our findings reveal that deep learning methods outperform traditional approaches in crowd anomaly detection, while real-time performance remains a critical challenge for many advanced techniques. Furthermore, this analytical study presents a taxonomy of video analytics applications and techniques and the primary research gaps and limitations that have been concluded, proposing promising directions for future research in video analytics.**

***Index Terms*— Crowd Behavior Analysis, Data Analysis, Video Analytics, Video Surveillance Cameras, State-of-Art, Survey**

I. INTRODUCTION

The persistent evolution of surveillance cameras has allowed them to penetrate various environments and systems [1] widely. With this affordable expanding technology, multiple applications have emerged, such as security and safety control, traffic management, healthcare, weather monitoring, and education [2]. While functioning all day, surveillance cameras generate vast videos daily [3]. These videos need to be analyzed to extract valuable insights and patterns that would either anticipate an upcoming serious event or learn from the past for better decisions in the future [4]. At the former stage, the applications relying on surveillance cameras were traditional and autonomous, with no analytical capabilities. It was insufficient for detection action prevention and required humans to continuously monitor and manually analyze cameras [5] [6]. Automated software has been made possible because of the scalable number of cameras and sensors, enhanced infrastructure, and advanced techniques for processing videos for image sequence analysis, video classification, object identification and tracking, and activity analysis [7]. Thus, video analytics based on surveillance cameras has recently drawn the attention of many researchers to optimize storage and better analyze behaviors [8],[9]. Video analytics combines computer vision and pattern recognition to extract contextual information and knowledge from the scene to understand the actions [10], [11]. It makes the surveillance system more efficient, reduces the workload, and helps to capture the full value of security video, i.e., monitoring abandoned objects, loitering and traffic flow, etc. Many organizations use it to increase business intelligence and provide a wealth of information [13]. Video analytics uses different algorithms and techniques to compare object types and determine behaviors or actions in real-time [14], which were considered for numerous surveillance camera-based applications as categorized in Fig. 1. The taxonomy presented above shows the very diverse range of applications of video analytics within a video surveillance context. It categorizes these into four main classes, namely, Human-Related, Nature-Related, Social-Related, and Object-Tracking and Detection. For instance, human-related applications include tasks such as detection and recognition, violence detection, shopping behavior analysis, some medical and educational applications, and behavior analysis, which has several types such as (Person, Group, and Crowd), Nature-related applications are concentrated on natural events, including fall detection, lava flow, fire detection. Applications that are social related include sports analysis, music recognition, and those within smart homes. Finally,

Object-Tracking and Detection applications will include areas such as traffic analysis, road safety, tracking, anomaly detection, and vehicle identification. This taxonomy gives a wide view of the different domains in which video analytics can be effectively utilized and shows the potential to address a wide range of challenges and opportunities.

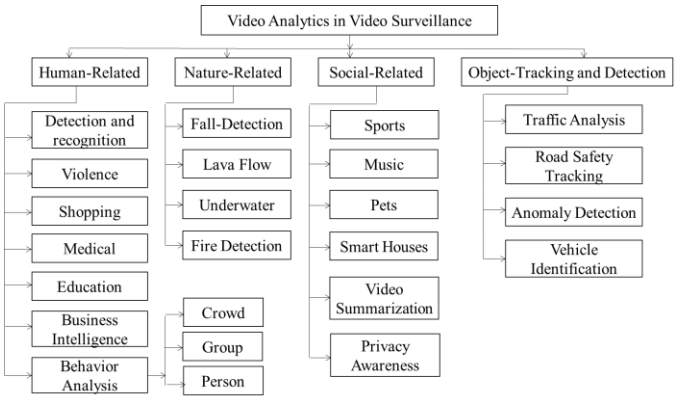

Fig.1. The taxonomy of applications based on video surveillance cameras applying video analysis.

This paper will be dedicated to a comprehensive investigation and comparison of different techniques of video analytics applied in surveillance systems. More Precisely, we aim to: (1) determine the strengths and weaknesses of various approaches, including traditional methods and deep learning algorithms; (2) Analyze the performance and limitations of existing video analytics systems in real-world scenarios; and (3) Identify main research gaps and propose promising future directions toward improving the accuracy, efficiency, and robustness of video analytics in surveillance applications.

The large-scale surveillance data should satisfy some properties, such as low false rate, high true alarm rates, and low computation cost [9]. Furthermore, the nature of these video sequences differs from conventional data, as they consume exponentially increasing storage, are heterogeneous since generated from various data sources, and are highly distributed, in which the same scene can be scattered on several cameras from different views with an extra computation cost [9], [15], [16]. Accordingly, storage optimization is considered by recording only video scenes that contain motion rather than static scenes [17]. Another type of application does not consider storing what is presented on cameras; rather, it extensively depends on human intervention to keep watching and acting when needed [14]. This represents another dilemma, as analyzing such massive streams in real-time is still challenging. Most video analysis is performed in the pixel domain to extract robust and meaningful features from the visual data [18]. However, analyzing video sequences using pixel algorithms takes an enormous time to compute such a massive amount of data, allowing compressed video data analytics techniques to be considered to reduce the computation power [19] and super-resolving of one frame from multiple neighboring frames using 3D convolutional in a spatio-temporal network [20]. Other studies used video object supervised/unsupervised segmentation methods [21]. Such video segmentation approaches contain foreground and background methods, in which video analysis can better understand the foreground and background and their relationship. In addition, video motion analysis has emerged in sports and some medical applications (i.e., catheters), which capture movie images from movie cameras and then apply frame-by-frame video playback to be analyzed with better performance [22]. However, many challenges arise when analyzing such a special nature of data, which promotes more studies to overcome such obstacles.

To make an effective and unbiased comparison, a selection of techniques is necessary in the realm of video analytics, which satisfies some selection criteria such as the following. Prevalence and Impact: We focused on techniques enjoying an important volume of research and significant practical applications on surveillance systems, considering the general impact they represent in the field and the conditions that allow its real-world implementation. Diversity of approaches: the selection includes a range of approaches, from traditional computer vision -textural background subtraction, optical flow- up to machine-learning algorithms such as Support Vector Machines, Hidden Markov Models, or deep learning architectures such as Convolutional Neural Networks and Recurrent Neural Networks. Availability and ease of access: We focus on such techniques for which public implementation exists, making possible easy reproducibility and follow-up research.

This paper comprehensively studies the different video analytics applications that rely on surveillance cameras. A detailed comparative evaluation of the primary applied algorithms and their associated challenges is discussed. Thus, we present the gaps currently concluded in this field, highlighting the pivotal research directions that can be adopted in the future. The rest of the paper is structured as follows. Section II discusses the video analytics conducted in human-related applications based on surveillance cameras. Section III provides datasets, and Section IV provides a detailed discussion and evaluation of the current research gaps in this field. Section V includes our conclusion and the future directions to consider.

II. VIDEO ANALYSIS IN HUMAN-RELATED APPLICATIONS

Human monitoring using surveillance cameras has widely been applied for different privacy and security purposes, ranging from simple individual detection and recognition applications to more sophisticated applications like behavior analysis, violence detection, shopping trends analysis, and sign language recognition, in addition to some medical and educational applications. Several research

studies were conducted for video analytics, where different algorithms and techniques have been considered to analyze those generated streams from surveillance cameras. Human activity can be categorized into gestures, actions, interactions, and behavior [23]. A comparison of human gesture recognition data mining classification methods by Kinect camera, including Backpropagation Neural Network (BPNN), Support Vector Machine (SVM), Decision Tree, and Naïve Bayes, was investigated in [24]. The performance of these classifiers is measured in terms of accuracy, precision, recall, and F1-score. The results obtained after the test showed that SVM had an accuracy of 92% in recognizing hand gestures. However, these classifiers were not applied in time series analysis to detect motion.

On the other hand, different background modeling methods were investigated in [25] for video analysis, comparing their performance and computation cost. GMM, AGMM, Vibe, PBAS, SOBS, SACON, KDE, and Codebook methods were considered to overcome illumination changes and dynamic background problems. The study evaluated the performance of these methods using metrics such as frame rate, detection rate, and false alarm rate. Statistical analysis, including ANOVA and t-tests, was performed to determine statistically significant differences in performance between the methods. The study claimed that AGMM, SOBS, Vibe, and PBAS were better, while others did not give accurate results. Further details are provided in the following sub-sections.

A. Detection and Recognition

A deep learning method for abnormal detection was presented in [26] to deal with abnormal behaviors under different conditions, such as background variations and several subjects (individual, two persons, and crowd). The characteristics concerned with abnormal detections were automatically learned through a CNN consisting of three stages: an input layer for images, a middle layer for detection, and a final layer for classification. However, the abnormal behavior detection for different subjects under more diverse conditions was not addressed to help design a robust intelligent surveillance system that could tackle various practical situations, i.e., indoor contexts, lawns, sports fields, and pedestrian crossing. In [27], a Bayesian Risk Kernel Density Estimation (BRKDE) was introduced to find the attractive region in video surveillance. The region was determined based on the Kernel Density Estimation (KDE) to create a map for crowd density. Different loss functions were used to find the distribution's peak, but finding peaks in low-density areas and crowd density of humans were not investigated, in which the distribution's peak is dramatically affected by outliers for large distances. A semi-autonomous system for tracking and fast, interactive retrieval was proposed in [28] to track and trace people over multiple real static cameras in a shopping mall. The system mainly consisted of three components: tracklet generation (track and detect), re-identification, and GUI. All videos have been processed in parallel on a distributed system, where tracking and detecting were continuously stored in a database. Yet, behavioral profiling and automatic action recognition were not considered. Another Gaussian Mixture Model (GMM) and Kalman Filter (KF) were proposed in [29] to detect and track persons in a video. GMM detected a moving person in an area, while a super-resolution technique was used to reduce the time for detection. Yet, the system was not applied in a real-time application, in which filters should be considered to overcome incomplete shapes and noise of the post-processing step in the proposed system.

In [30], a machine learning framework based on an Intrusion Detection System (IDS) was presented to improve the detection accuracy of both anomaly and misuse detections. Different machine learning algorithms were applied, such as inductive learning, case learning, and genetic learning, in which the detection was generally classified as normal or abnormal for each frame in the video. However, the system's speed, real-time, and robustness were not discussed. A covariance and automatic artificial neural network were used [31] to detect anomalies and abnormal events in crowd behaviors in dangerous situations. Crowd behaviors were classified to estimate the crowd's density, extract motives for movements, and detect abnormal events from the unidirectional flow of the crowd. The video frames were labeled as normal and abnormal based on the distance between the covariance matrix and the variance of optical flow. However, the real-time performance of the method, as well as its association with the linear discriminate analysis (LDA) technique for covariance matrix, were not investigated. A method for moving object detection, tracking, and classification from a video captured by a moving camera without additional sensors was proposed in [32] for real-time applications. Nevertheless, the heavy crowd of moving objects with overlapping different features was not investigated. In [33], a 3D human body detection, tracking, and recognition framework from depth video sequences were proposed using spatiotemporal features and the Modified-Hidden Markov Model (M-HMM). It is considered better activity recognition, as it overcame the problems of missing joint information, unclear human silhouettes, and large distance subjects that cause low recognition. However, the authors [34] introduced a method incorporating the histogram of oriented gradient (HOG), the theory of visual saliency, the saliency prediction model, and deep multi-level networks to detect humans in video sequences. The HOG features were trained on an SVM to detect humans in any frame. In addition, the K-means algorithm was applied for HOG features clustering to find movement patterns of humans in the frames. Yet,

incorporating motion tracking on an optical flow in conjunction with saliency windowing to detect motion patterns of humans for video surveillance was not considered. Another system considers background modeling, background subtraction, and foregrounds and backgrounds to detect humans and human tracking, respectively. However, the system did not investigate activity recognition in video surveillance. In [35], multiple action detection, recognition [36], [37], and summarization approach was presented. The background subtraction approach was used to extract human bodies to make action detection. After that, motion detection and tracking methods determined a set of actions in the generated sequence. These sequences have been divided into shots representing homogenous actions using similarity between frames. There are two methods used to classify action. Cosine similarity of the HOGs of Temporal Difference Map (TDMap) and the second one is a CNN classification of actions from TDMap images. Finally, the summarization of the video was crucial to some challenges like illumination and overlapping. Summarizations depend on the type of scene (private or public), whether the scene is (static or dynamic), and whether the scene is (crowded or uncrowded). In [38], [39], the need for more reliable detection and recognition of activities and hence the use of advanced techniques like Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Support Vector Machine (SVM) in Human Activity Recognition (HAR) systems. The article discusses several datasets, including Opportunity, Pamap2, VanKasteren, Ordonez, WISDM, and UCI-HAR, that consist of a diverse set of sensor-based and vision-based data. Some of the presented HAR techniques include the EnsemHAR model, which presents high accuracy rates for all the tested datasets. Some identified gaps include limited literature on real-time activities, sparse coverage of diverse activity domains, and challenges related to data preprocessing, hardware constraints, and activity misalignment. Future work should explore new applications, improve data collection and preprocessing methods, develop tailored algorithms and hardware solutions in collaboration with experts, and leverage the latest technologies to bridge this gap. The detection of abnormal crowd-level behaviors (CABs) in very complex crowd interaction cases, trying to mitigate threatening forces like excessive contact forces and turbulence, was discussed in [40]. It suggests applying the Multi-Scale Motion Consistency Network (MSMC-Net) as a solution because it uses a graph-based representation to capture spatial and temporal motion consistency information. MSMC-Net uses multiple feature graphs and a network attention mechanism for adaptive feature fusion to detect CABs. Unlike existing methodologies focusing on local anomalies, this paper focuses on global patterns in collective crowd motion. It introduces a multi-scale motion learning framework to cover different types of CABs. Experimental results show that MSMC-Net significantly outperforms all related baselines across various operational scenarios. To detect abnormal behavior in massive crowd videos during the Hajj pilgrimage, the existing limitations in current solutions are considered while targeting only small-scale crowds with simplistic abnormalities. The proposed approach adopts a Generative Adversarial Network (GAN)-based methodology [41], employed in optical flow capturing motion information within the videos. Optical flow extraction, GAN-based model training, a comprehensive loss function, and transfer learning mechanisms are some of the components involved. The "Massive Crowd Abnormal Behaviors HAJJ Dataset" features a greater depth of analysis, enhanced by its diverse abnormal behavior videos, surpassing the capabilities of existing datasets. For instance, the use of the proposed framework turns out to be effective by observing impressive accuracy rates. However, problems still exist, especially in further improving accuracy. Further study will require a greater collection of datasets, intricate feature extraction, and refining models to detect abnormal behaviors in videos of massive crowds accurately. [42], Identifying anomalous behavior in crowd videos, understanding how to handle occlusion, crowd density variations, or varying contextual understanding. It proposes solutions with pre-trained convolutional neural networks (CNNs) such as GoogLeNet, VGGNet, or AlexNet to take out features from video frames that are further fed into the different machine learning classifiers for anomaly detection. The three datasets for evaluation are CUHK Avenue, UCSD Ped-1, and UCSD Ped-2, which hold video clips depicting walkways crowded with people with normal and abnormal events. Despite the presentation of figures demonstrating the anomaly detection performance, this work lacks detailed quantitative results or comparisons with existing methods. A critique noted in this context is that blurriness may be introduced into the denoising step during frame preprocessing, prompting the suggestion to explore better denoising techniques for future research. TABLE 1 summarizes the studies that consider video analysis using surveillance cameras for detection and recognition.

TABLE 1 Summarization of the state-of-the-art techniques for Object detection and recognition.

Ref #	Video Analysis Purpose	Applied Techniques	Context	Dataset	Evaluation Criteria	No. of Cameras	Analysis Type
[26]	Abnormal detection under different conditions	CNN based abnormal detection	N/A	CMU, CMU, UTI, PEL, HOF, WED	Performance, time	Single	Offline
[27]	Find attractive regions	BRKDE	Human detection	9 real-world datasets (PETS2009, Mall dataset)	Accuracy	Single	Offline
[28]	Track and trace people	Semi-autonomous system for tracking and fast interactive retrieval	Shopping Mall	VIPeR dataset		Multi-camera	Real-time
[29]	Detect and track persons in videos	GMM and KF	N/A	Benchmark video sequences for Shopping center, Buffet, campus Restaurant, lobby	Performance	Single	Offline
[30]	Improve detection accuracy	Machine learning based on IDS	N/A	Shell commands with different lengths	accuracy	Single	Offline
[31]	Model crowd behavior in dangerous situations	Covariance and automatic ANN to detect several anomalies	Complex (indoor and outdoor)	UMN 2006, PET2009, BMVA	Performance	Single	Offline
[32]	Detect and track moving objects	A method based on moving cameras without additional sensors	N/A	Self-made datasets for Plaza, playground, and library entrance	Performance Recall, Precision, F-measure	Single	Real-time
[33]	Detect, track and recognize human body	Spatiotemporal features and M-HMM	N/A	MSRDaily Activity 3D, IM-Daily Depth Activity, SMMC-10	Accuracy	Single	Offline
[34]	Detect humans in video sequences	HOG, visual saliency, and saliency prediction model deep multi-level NW	N/A	OSU color-thermal pedestrian	Precision Recall Time	Single	Offline
[43]	Detect and track human	Background, background subtraction and KF	N/A	WEIZMANN		Single	Offline
[35]	Detect, recognize, and summarization	HOG of Temporal Difference Map (TDMap), CNN	Human detection, recognition, and summarization	Weizmann, KTH, UCF-ARG, UT-Interaction, IXMAS and MHAD	performance	N/A	Offline and recommend using it in real time because of its simplicity
[41]	Detect abnormal behavior in crowd	Generative Adversarial Network (GAN)-based	Hajj	UMN, UCSD, and the HAJJ dataset	Accuracy	N/A	offline

B. Behavior Analysis

Behavior analysis is the formal study of behavior patterns to comprehend the varied contexts in which humans behave or perform activities[44]. This includes gathering data from many sources, including surveillance cameras, online activities, social interactions, IoT sensors, and those obtained from manufacturing machine behaviors. Behavior analysis is applied in many fields like education, medicine, sports, festivals, transportation, etc. In educational settings, the goal is to address the student's behavior problems using evidence-based strategies with the anticipation of enhancing academic results and creating a conducive learning

environment[45]. Besides, in privacy-based applications [46], behavior analysis is realized by applying techniques like local object tracking (LOT) utilizing infrared sensor arrays at a 99% correct recognition rate for bedside activities and extremely high user satisfaction. Generally, behavior analysis is the systematic observation and modification of behavior through empirical evidence in psychology, public policy, education, and mental health. Human behavior analysis, group behavior analysis, and crowd behavior analysis could be some of the classifications for behavior analysis.

1) Human behavior analysis

An expert system for real-time detection of suspicious behaviors in malls was proposed in [47]. It allowed automated risk situation detection to help security officers enhance asset protection. The proposed work provided monocular vision and tried to decrease the number of cameras in real time, but it did not care about processing time. Besides, collaborative, and stereo cameras with machine learning algorithms were not considered to facilitate the extraction of 3D data and automate all surveillance tasks. In [48], a multi-model human behavior analysis based on Kernel Canonical Correlation Analysis (KCCA) and Multi-View Hidden Conditional Random Field (MV-HCRF) was presented to capture dynamic hidden nonlinear correlations and interactions to recognize agreements and disagreements from nonverbal audio and video of spontaneous political debates. However, the proposed methodology was not applied in different multi-model human behavior analysis applications. Besides, neither the metrics used to measure performance nor performance results were mentioned. Infinite Hidden Conditional Random Fields (IHCRF) based on the Hierarchical Dirichlet Process (HDP) that was capable of automatically learning the optimal number of hidden states for classification was introduced in [49] to recognize instances of agreements and disagreements in recorded spontaneous human behavior. The proposed approach was used to find the label of a new observation sequence and the correct number of hidden states rather than HCRF, which was concerned with gesture recognition [50] and speech recognition [51]. It handled complex features without any changes in any training procedures. Yet the proposed model was not applied to various datasets and variational inference; different approaches for learning IHCRF were not investigated. In [52], an embedded intelligent surveillance system was proposed to detect, track, and analyze human behavior scenes and signal when dangerous behavior occurs indoors. The proposed system was presented to overcome strong lightness variation and crossing people. The mean-shift method and Oriented Fast and Rotated BRIEF (ORB) descriptor were used to track people crossing, adding unique features of identifying people and for the occlusion using a single camera. However, the occlusion of multiple people using multi-cameras encountered problems, like increasing the hardware cost and limiting the camera installation. A deep 3D convolutional network based on 3DCNN was proposed in [53] to find efficient spatial-temporal features of infrared videos. The difference between infrared and visible video data was the spatial domains, while behavior recognition was concerned with information in the time domains. There was little infrared data available for training. The proposed approach was compared with others, but it did not care about the preprocessing costs; besides, it was a hard-coded convolution to extract greyscale, gradient, and optical flow from the video. In [54], a method based on preference distribution and KL distance similarity was introduced to classify and label videos in an intelligent video surveillance system. Intelligent video surveillance could identify tracks, analyze video images, and automatic monitoring [55]. The preference distribution method included multi-tag classification and multi-tag sorting. However, this method was subjective and did not improve the optical flow, which implied that motion in a video is achieved by matching points on objects over frames. A ghost elimination method was proposed in [56] to extract moving objects and automatically recognize human behavior for foreground detection based on a vibe algorithm. SVM and the star model were used to select the training set to allow the classification model to detect human behavior and extract features, respectively. Still, the measured average classification rate was not comparable to that of others. Thus, the results were not verified. The multi-type feature fusion and irrelevant feature reduction (MtRF) approach was presented in [57] for human behavior and action recognition. Features like shape and color were extracted based on serial methods to reduce the irrelevant and redundant features. One-Against-All (OAA) multi-class SVM classifiers selected and recognized the reduced features. In [58], an ATM real-time video surveillance system was presented to alert security officers for emergencies. The system was applied in a cloud environment using multi-camera, but human resources were required for real-time observed cameras. In [59], a cognitive computing model based on context-aware data flow was proposed for large-scale data analysis and processing, especially in human behavior analysis. The CART algorithm models the pattern in the data, and the K-means clustering method is optimized. It is through simulation experiments and the real data application that the performance of the algorithm is evaluated. Some of the identified gaps include the broader exploration of the analysis of human behavior within an Internet application environment and the extraction of meaningful information from large datasets. It also emphasizes the importance of cognitive computing in understanding human behavior and potential personalized services because of the development of Internet behaviors and effective methods for processing data in computing user cognition and providing personalized services. also points to the challenges arising from extracting meaningful data from large-scale datasets, which require proper processing. It then

suggests that current methods of analyzing and processing such data are inadequate. TABLE 2 summarizes the studies that consider video analysis using surveillance cameras for human behavior analysis.

TABLE 2 Summarization of the state-of-the-art techniques for human behavior analysis.

Ref #	Video Analysis Purpose	Applied Techniques	Context	Dataset	Evaluation Criteria	No. of Cameras	Analysis Type
[47]	Automated detection of risk situations	Expert system	Mall	CAVIAR dataset	Detection Rate (DR), Specs, False Alarm Rate (FAR)	Multi-camera	Real-time
[48]	Complex nonlinear correlations and interactions across Modalities.	Multi-model human behavior based on (KCCA) and (MV-HCRF)	N/A	Canal 9 dataset	agreement and disagreement recognition	Single	Offline
[49]	Non-parametric model dependent on HDP	(IHCRF) based on (HDP) HCRF (C++) IHCRF (MATLAB)	N/A	Canal9 Dataset, UNBC dataset	Performance, F1 measure	Single	Offline
[52]	Analyzing human behavior in a scene	Intelligence surveillance System (embedded system)	Indoor environments	fifteen videos in different scenes.	Accuracy Performance	Single	Offline
[54]	Security in intelligent video surveillance	Method based on preference distribution and uses KL distance similarity	N/A	Panic, gang Fight, Stampede, and Fall	Optical flow	Single	Offline
[56]	Automatic recognition human behavior	Ghost elimination method, Vibe Algorithm	N/A	Actions	Average success rate	Single	Offline
[53]	Behavior recognition in infrared video	infrared-3D network	N/A	UCF101, infrared datasets	Accuracy	Single	Offline
[57]	HBR or HAR	MtFR	N/A	Muhavi, WVU and YouTube	Recognition Rate (RR)	Single	Offline
[58]	Alert security officers	ATM video surveillance system	Cloud environment			Multi-camera	Real-time

2) Crowd behavior analysis.

The analytical process of crowd behavior analysis (crowd scene analysis or crowd abnormality analysis)[61] includes stages such as data collection, preprocessing, feature extraction, behavior modeling, and anomaly detection, which enable one to precisely model and predict the dynamics of the crowd. Steps involved in crowd behavior analysis (macroscopic) include the following: detection, tracking, feature extraction, and behavior classification, as shown in Fig.2. Real-time video surveillance, particularly in crowded areas, is challenging but very important for public security; it involves detecting and analyzing abnormal activities in real time[62], [63],[64]. Research involves intelligent methods such as deep learning and machine learning algorithms that automatically discriminate against normal and abnormal activities. Common approaches include pre-trained CNNs, novel optical flow-based features for abnormal behavior detection, spatial and temporal feature combinations, object tracking, and motion analysis.

Benchmark datasets, including CUHK Avenue and UCSD Ped1/Ped2[62], are very common for assessment; however, quantitative performance comparison across methods is relatively rare. Datasets in [64] include UMN and PETS 2009. Posited gaps include improving denoising techniques, looking at unsupervised approaches, raising the computational cost, and dealing with real-world challenges while integrating multimodal data to improve anomaly detection. In [65], a comprehensive review is provided of techniques for crowd behavior analysis, including crowd counting, crowd tracking, and crowd anomaly detection. Advantages and disadvantages are discussed by comparing traditional approaches using handcrafted features like HOG, SIFT, etc., with deep learning approaches that learn features automatically. It provides performance comparisons of the proposed method with existing state-of-the-art approaches using several benchmark datasets, including UCSD, UMN, Avenue, and others. The tracking performance is evaluated using various metrics such as MAE (Mean Absolute Error) and MSE (Mean Squared Error). In crowd tracking, the MOT Accuracy (MOTA) metric is used to evaluate the performance of the proposed method on datasets like TownCenter and PETS2009, among others. For crowd anomaly detection, the evaluation metrics are Accuracy, Equal Error Rate (EER), and Area Under ROC Curve (AUC). The deep learning techniques are shown, in general, to outperform traditional approaches for crowd behavior analysis tasks. The paper concludes that deep learning methods exhibit better results than traditional crowd behavior analysis approaches across all three categories on most benchmark datasets.

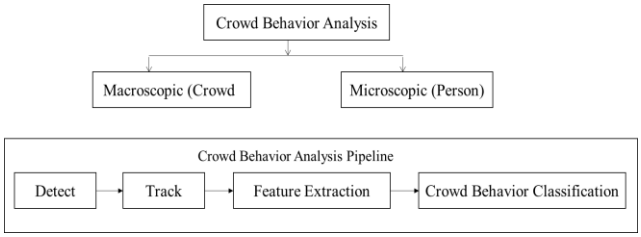

Fig.2. Crowd Behavior Analysis Pipeline.

Abnormal human behavior detection and classification in crowded scenes was discussed in [66] using image processing techniques. Abnormal behavior in crowds is detecting issues due to occlusions, varying densities of crowds, and complexity in tracking individual trajectories. Optical flow features were extracted from the video frames to capture crowd motion patterns. Based on optical flow, the three features computed are average kinetic energy, movement direction entropy, and crowd distance potential energy. The proposed method used the extracted optical flow features and SVM classifiers to classify normal versus abnormal crowd behavior. With optical flow features and an SVM classifier, the proposed method performed 96.75% on the testing accuracy of classifying abnormal behavior, outperforming logistic regression (96.45%) and KNN (96.68%). The primary evaluation metric used is classification accuracy on a testing dataset with videos labeled as normal or abnormal crowd behavior. The gaps are: No information is given on the dataset used for training or testing, and nothing is revealed on the types of abnormal behavior being considered. The comparative analysis is conducted on well-known crowd datasets like UCSD, UMN, PETS, Avenue, etc. However, it is unclear if the proposed method in this paper used any of these datasets.

Computer vision and machine learning techniques such as convolutional neural networks (CNNs) are used to detect anomalies in crowd behavior from video data [67]. Several public datasets, such as UMN, UCSD, and ShanghaiTech, are available to evaluate models for crowd anomaly detection. Examples of needed gaps are developing generalized models from scene to scene, alleviating the computational costs, and deploying application-specific models. Future work is expected to provide generative models, graph-based methods, online learning, ensemble techniques, and meta-learning for improving performance and robustness. This paper presents a survey of the state of the art in crowd anomaly detection, taxonomies, methods, datasets, and evaluation metrics, overviewing current progress. Deep learning and other methods confront challenges in real-time anomaly detection, especially in crowded scenarios. These include high computational costs, the need for large annotated datasets, and the privacy issue. At the same time, limits are placed on what can be detected when considering the detection of new types of anomalies.

Additionally, there are questions of scalability, dealing with multiple anomaly types simultaneously, dealing with environmental changes such as occlusion or changes in lighting, and explaining the detected anomalies at their root causes. Then, it is observed that current models lack generalization across different scenes or datasets, which calls for scenario-specific model designs to improve their effectiveness in any surveillance context. Abnormal behavior detection in crowded videos, targeting detecting various types of motion in a complex environment, was discussed in [68]. This work identifies the correct detection and tracking of anomalies based on low-level features—global, local, and feature features—to identify anomalies in the motions of the objects. This is tested using a challenging dataset to improve the diagnosis of anomalies with the help of video tracking and feature extraction. The work further attempts to improve the accuracy of anomaly detection and identification in video content by putting forward options and comparing them to perform better in real-time and outdoor scenarios. The process of video modeling is also

pointed out with improved accuracy and less processing time to achieve perfection in real-time and outdoor scenarios.

Anomalous crowd behavior detection in video surveillance data was discussed in[69]. Anomalies could include some people moving abnormally compared to regular pedestrian traffic or others still in certain areas. KLT feature tracking has been used to extract motion cues from video frames, with dynamic graph sequences representing crowd movement based on the proximity and velocity of tracked objects. A max-flow/min-cut algorithm can identify graph components or communities far away from the rest, showing abnormal behavior. The model gets trained offline, and then anomaly detection is done online in a computationally efficient manner suitable for real-time applications. Datasets to be used: UCSD Pedestrian 1 and 2 datasets. Private MCG (Melbourne Cricket Ground) dataset with six camera views. On UCSD Ped1, 91.6% AUC and 17.5% EER were achieved for frame-level anomaly detection, 63.5% AUC, and 39.5% EER for pixel level. On UCSD Ped2, 88.3% AUC and 17.5% EER were achieved for frame-level detection, outperforming most existing methods. On Avenue achieved 85.6% AUC and 23.15% EER for frame level, the second-best in AUC and best in EER compared to the rest of the prior work. However, results suggest this method is not necessarily the best on all datasets/metrics compared with some deep learning approaches. Computational time was not extensively analyzed; however, it is stated to be competitive for real-time use. Parameters like time window size were selected empirically on one dataset and may need tuning for others. Table 3 summarizes the studies that consider video analysis using surveillance cameras for crowd behavior analysis.

TABLE 3 Summarization of the state-of-the-art techniques for crowd behavior analysis.

Ref #	**Video Analysis Purpose**	**Applied Techniques**	**Context**	**Dataset**	**Evaluation Criteria**	**No. of Cameras**	**Analysis Type**
[61]	detecting and analyzing abnormal activities	CNNs (GoogLeNet, VGGNet, AlexNet, etc.), spatial CNN features with LSTM, SVMs, autoencoders	crowded public scenes	CUHK Avenue and UCSD Ped1/Ped2	ROC curves, accuracy	N/A	Real-time
[62], [68]	Detect and identify anomalous	A novel optical flow based features for abnormal crowd behavior detection	crowded public scenes	UMN and PETS 2009	Accuracy, analyze the impact of window size and feature vector dimensions	N/A	Real-time
[63]	Review on crowd behavior analysis techniques	Traditional techniques like HOG, SIFT, and deep learning	Crowd	UCSD, UMN, Avenue	MAE (Mean Absolute Error) and MSE (Mean Squared Error)	N/A	N/A
[64]	detect and classify abnormal human behavior	Optical flow features and SVM Classifier	Crowd Scene	Don't mention	Accuracy	N/A	N/A
[65]	detect anomalies	convolutional neural networks (CNNs)	crowd behavior	UMN, UCSD, and ShanghaiTech	Accuracy	N/A	N/A
[67]	Detect anomalous	Graph-Based	Crowd behavior	UCSD PED1, PED2, MCG, Avenue	Accuracy	N/A	Real-Time

III. DATASETS

Over recent years, crowd behavior and anomaly detection datasets have grown. These datasets may be used to analyze, compare, and better the performance of crowd behavior and anomaly detection systems. Signed up for the crowds-related application includes counting, density estimation, categorization, activity recognition, and anomaly detection and recognition. Most visual real-world crowd datasets have focused on detection tasks, including UMN, UCSD, CUHK Avenue, UCF-crime, Hajj, Shanghai Tech, etc. The details of these datasets and many others available for crowd anomaly detection and recognition are discussed in TABLE 4. The name of the dataset, the size of the dataset, a small description, ground truth, and problems are mentioned in the table. Also, TABLE 5. Show performance comparison on UMN, UCSD-Ped1/Ped2, and CUHK Avenue datasets with the names of the methods used.

TABLE 4 Summarization of Most Common Datasets in Crowd Analysis.

Name	Abbreviation	Size	Description	Scenario	Ground-truth
UMN [69], [62]	University of Minnesota	Small	It contains 11 videos of 3 different indoor and outdoor scenes.	Unusual crowd activity (Anomaly Detection)	yes
UCSD [70]	University of California, San Diego.	Ped1, small Ped2, large (Pedestrian)	A stationary camera mounted at an elevation level gained the dataset, overlooking the pedestrian walkways. It consists of Ped1 and Ped2. Positively, it contains Ped1 groups of people facing towards and away from the camera, with 34 training samples and 36 testing video samples. Ped2: scenes that have pedestrian movement parallel to the camera plane. It contains 16 training video samples and 12 testing video samples. For each clip, the ground truth annotation includes a binary flag per frame, indicating whether an anomaly is present at that frame.	Anomaly Detection	yes
CUHK Avenue [71]	Chinese University of Hong Kong	Small	It comprises 16 videos for training and 21 for testing. Captured in the Chinese University of Hong Kong (CUHK) campus avenue, each video has 3,065 frames in total (15,328 training, 15,324 testing). The training videos contain normal situations. The testing videos include both normal and abnormal events.	Abnormal Event Detection	yes
UCF-Crime[72]	University of Central Florida	large-scale dataset of 128 hours of videos	It consists of 1900 long and untrimmed real-world surveillance videos, with 14 realistic anomalies including Abuse, Arrest, Arson, Assault, Road Accident, Burglary, Explosion, Fighting, Robbery, Shooting, Stealing, Shoplifting, normal Videos and Vandalism.	anomaly detection	NO

HAJJ V2 [73], [75]	Hajj	Large-scale Crowds	The Hajjv2 dataset has 18 manually captured videos from the annual event. All the videos are stored in an mp4 extension. The videos depict abnormal behaviors of individuals in the massive crowd. The videos were taken at different scenes and places in the wild during the Hajj event. Five videos are filmed in the "Massaa" scene, while the remaining videos are filmed in "Jamarat," "Arafat," and "Tawaf.". In these videos, individual abnormal behaviors involve standing, sitting, sleeping, running, moving in different or opposite crowd directions, and non-pedestrian entities, such as cars and wheelchairs. These behaviors are potentially dangerous for crowd flow on a large scale.	Abnormal Behaviors Detection	NO
Shanghai Tech [74]	Research university located in Shanghai, China.	large-scale crowd	It consists of 1,198 annotated crowd images. The dataset is subdivided into Part-A, containing 482 images and Part-B containing 716 images. The two sets of images are used for Part-A and Part-B only to create train/test splits. Part-A contains 300 images in the train set and 182 images in the test set. Part-B consists of 400 and 316 images for training and testing, respectively. In each crowd image, each person is annotated by a single point close to the center of the head. This dataset is summarized by 330,165 annotated people. The images in Part-A were procured from the Internet, while images from Part-B were collected on the busy streets of Shanghai.	crowd counting	No

TABLE 5 Performance comparison on UMN and UCSD, and Avenue CHUK datasets

Dataset	**UMN**	**UCSD-Ped1**	**UCSD- Ped2**	**Avenue CUHK**
Method				
3DCNN- LSTM	99.5%			
CNN- Residual LSTM	98.20%			
Optical flow	98.03%			
CNN+RF	99.77%			
max-flow/min-cut pedestrian flow optimization scheme (MFMCPOS).		91.6%- frame level, and 63.6%-pixel level	88.3%- Frame Level	85.6%
IVADC-FDRL (Intelligent Video Anomaly Detection and Classification using Faster RNN with Deep learning reinforcement Learning)		98.5%- Test004 and 94.80%- Test007		

Interactive Trajectory-Level Behavior Computation	85%
Conv-LSTM	83.00%
Conv LSTM + Conv Encoder	95.20%
Vgg-16 LSTM	95%
Cascaded Attention CNN	97.40%

IV. Discussion and research gaps

Activity Recognition Integration. Few studies combine abnormal behavior detection with activity recognition, which could provide a valuable context for interpreting crowd behavior [28],[35]. There are many applications in Human behavior analysis in areas as diverse as security, education, and healthcare. Some examples are detecting suspicious behavior in malls [47] or, more recently, recognizing agreements/disagreements in debates [48]. Data analyzed with computational help can be taken from surveillance cameras, audio recordings, or sensor readings. Some studies achieve real-time processing; others, however, lack focus on optimizing processing speed [53]. Deriving meaningful insights from large datasets requires advanced processing techniques beyond current approaches [59]. Although multi-camera systems have advantages, hardware costs can be raised, and installation complexity can increase [52]. Exploiting meaningful features from infrared videos is challenging due to the scarcity of available data [53]. Feature extraction needs to be developed into robust methods, and the irrelevant information needs to be minimized. A reliable crowd anomaly detection algorithm evaluates local and global density and reliably anticipates crowd behavior. There are some classifications of crowd methods, such as traditional or deep learning methods. Another classification was object-based approaches and holistic approaches [60]. In the object-based approach, an individual is traced in the crowd. It can be performed by detecting an individual via segmentation to analyze group behavior. Many tools and methods, such as Laplacian and graph, can be used to trace the individuals in the group. However, this approach works only for sparse or moderate crowds. For the dense crowds, it is less appropriate due to occlusion. Occlusion has been a major challenge in video analysis and is even difficult for crowd anomaly detection.

Occlusion due to overlapping objects remains a significant challenge for accurate behavior detection [42]. Occlusion can limit the effectiveness of crowd anomaly detection systems as it makes identifying individuals in a crowd demanding. Whereas, in holistic approaches, the crowd is considered a global identity that focuses on global features like fully convolutional neural networks (FCNN) to learn appearance features and motion features, spatiotemporal CNN, trajectories path, optical flow, etc. learning methods have become the new research trend to analyze the behavior of the crowd. Deep learning approaches using Convolutional Neural Networks (CNNs) are increasingly adopted for crowd behavior analysis tasks and outperformed traditional approaches that rely on handcrafted features [63]. Other approaches have successfully demonstrated the effectiveness of deep learning methods, such as Convolutional Neural Networks and Generative Adversarial Networks, in detecting abnormal crowd behavior [26,40,41]. The methods can learn complex features from video data, which results in better detection accuracy. However, despite such advancements, the development of robust abnormal behavior detection systems is still hindered by several challenges. Many methods cannot provide real-time processing speeds, which hinders their practical use in surveillance settings [30]. Systems usually cannot adapt to conditions such as background clutter, crowd density, and illumination changes [27,32]. The lack of large and diverse datasets for training and evaluation hinders generalization models.

V. Future Directions and Conclusion

Development of more robust but complex feature extraction methods capable of working with complex crowd scenarios. Building more extensive and diverse datasets that include scenarios with different densities of people, environments, and abnormal behaviors is also necessary. Improved data pre-processing techniques to reduce noise and remove blurs may also be helpful. Research should focus on methods that optimize the algorithms and use hardware acceleration for real-time processing in practical applications. Contextual information such as scene type and patterns of activities can also boost the accuracy and interpretability of abnormal behavior detection. Research can also investigate the presentation of these deep learning methods and other techniques, such as optical flow analysis and graph-based representations, to garner more robust solutions.

We consider real-time processing and data privacy as the two most urgent issues to be solved. Indeed, real-time analysis is at the core of many applications related to surveillance, such as crowd control or emergency response. To reach real-time performance, efficient development of algorithms, along with hardware acceleration, will be one of the keys. Besides, because

of increasing usages, the privacy of data has become the prime concern of research on the surveillance system. Research related to privacy-preserving techniques such as federated learning and differential privacy should be a priority.

3D data extraction and automation of surveillance tasks can be facilitated with collaborative and stereo cameras assisted by machine learning [47]. Variational Inference for IHCRF: There is a need to examine variational inference approaches for the performance enhancement of Infinite Hidden Conditional Random Fields (IHCRFs) [49]. Better understanding of human behavior and enabling personalized services can be improved by integrating cognitive computing approaches. Human behavior analysis in internet application environments will be a promising avenue for future studies [59]. Practices for processing noise in video data need to improve [62]. Unsupervised learning is yet another area that needs to be explored to reduce dependence on labeled data [62]. Incorporating data from multiple sources, such as audio and sensors, is likely to help improve anomaly detection. One main issue is developing models that fare well over various crowd scenarios [65]. Computation is also an essential consideration as an application that is deployable in a real-time setting remains to be studied. Another highly relevant issue is developing models that can explain how anomalies are detected. Models for crowd behavior need to be scalable to accommodate large crowds and adaptable to environmental changes [65]. However, there is a need to balance security objectives with privacy when collecting and analyzing data for crowds [65]. In this respect, generative models, graph-based methods, and online learning potentially improve performance and robustness [65]. This enables modeling and predicting crowd dynamics. Due to occlusion and computational expense, real-time analysis of crowded scenes is yet to be accomplished for dynamic scenes [61,62].

Behavior analysis is a powerful tool for understanding human actions and interactions in many contexts. This kind of analysis is widely applicable to education and healthcare, as well as to security and public safety. Nevertheless, the review focused on two major areas within behavior analysis: human behavior analysis and crowd behavior analysis. Both areas have seen groundbreaking work, primarily because of the increased use of deep learning. These techniques allow researchers to interpret deep, informative features from data sources such as video recordings and sensor readings. Because of these, behavior recognition and anomaly detection have been vastly improved.

However, there are still many challenges that need to be solved. Researchers are still not adequately addressing real-time processing data privacy concerns and developing generalizable models capable of performing equally well in different scenarios. These focus on data from various sources and developing explainable AI models whose outputs reveal the reasoning of detected anomalies.

As research advances, behavior analysis can become a multifunctional tool for all these applications. This can take the form of more sophisticated ways to ensure public safety, help educational and health professionals in their daily tasks, and carry out a much deeper analysis of patterns of human behavior in all kinds of scenarios. By solving hidden problems and understanding novel research directions, behavior analysis may play a significant role in building a safer, more effective, and more insightful future.

REFERENCES

[1] C. S. Regazzoni, A. Cavallaro, Y. Wu, J. Konrad, and Arun Hampapur, "Video Analytics for Surveillance: Theory and Practice," *IEEE Signal Process. Mag.*, no. September, pp. 16–17, 2010, doi: 10.1109/MSP.2010.937451.

[2] C. C. Loy, T. Xiang, and S. Gong, "Time-delayed correlation analysis for multi-camera activity understanding," *Int. J. Comput. Vis.*, vol. 90, no. 1, pp. 106–129, 2010, doi: 10.1007/s11263-010-0347-5.

[3] T. Huang, "Surveillance video: The biggest big data," *Comput. Now*, vol. 7, 2014, [Online]. Available: https://computingnow.computer.org/web/computingnow/archive/february2014

[4] A. Oussous, F. Z. Benjelloun, A. Ait Lahcen, and S. Belfkih, "Big Data technologies: A survey," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 30, no. 4, pp. 431–448, 2018, doi: 10.1016/j.jksuci.2017.06.001.

[5] I. S. Kim, H. S. Choi, K. M. Yi, J. Y. Choi, and S. G. Kong, "Intelligent visual surveillance - A survey," *Int. J. Control. Autom. Syst.*, vol. 8, no. 5, pp. 926–939, 2010, doi: 10.1007/s12555-010-0501-4.

[6] S. Vishwakarma and A. Agrawal, "A survey on activity recognition and behavior understanding in video surveillance," *Vis. Comput.*, vol. 29, no. 10, pp. 983–1009, 2013, doi: 10.1007/s00371-012-0752-6.

[7] F. C. Cheng, S. C. Huang, and S. J. Ruan, "Scene analysis for object detection in advanced surveillance systems using laplacian distribution model," *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 41, no. 5, pp. 589–598, 2011, doi: 10.1109/TSMCC.2010.2092425.

[8] D. Gowsikhaa, S. Abirami, and R. Baskaran, "Automated human behavior analysis from surveillance videos: a survey," *Artif. Intell. Rev.*, vol. 42, no. 4, pp. 747–765, 2014, doi: 10.1007/s10462-012-9341-3.

[9] P. Guler, D. Emeksiz, A. Temizel, M. Teke, and T. T. Temizel, "Real-time multi-camera video analytics system on GPU," *J. Real-Time Image Process.*, vol. 11, no. 3, pp. 457–472, 2016, doi: 10.1007/s11554-013-0337-2.

[10] A. Dore, M. Soto, and C. S. Regazzoni, "Bayesian tracking for video analytics," *IEEE Signal Process. Mag.*, vol. 27, no. 5, pp. 46–55, 2010, doi: 10.1109/MSP.2010.937395.

[11] N. Li, X. Wu, D. Xu, H. Guo, and W. Feng, "Spatio-temporal context analysis within video volumes for anomalous-event detection and localization," *Neurocomputing*, vol. 155, pp. 309–319, 2015, doi: 10.1016/j.neucom.2014.12.064.

[12] LHARRITY, "WHAT IS VIDEO ANALYTICS?," 2015, [Online]. Available: https://www.worldeyecam.com/blog/general/what-is-video-analytics.html

[13] P. Kirve, "Video analytics: Why organizations should crave for it?," 2019, [Online]. Available: https://bigdata-madesimple.com/video-analytics-why-organizations-should-crave-for-it/

[14] Y. Ye, S. Ci, A. K. Katsaggelos, Y. Liu, and Y. Qian, "Wireless video surveillance: A survey," *IEEE Access*, vol. 1, pp. 646–660, 2013, doi: 10.1109/ACCESS.2013.2282613.

[15] G. Ananthanarayanan *et al.*, "Real-Time Video Analytics: The Killer App for Edge Computing," *Computer (Long. Beach. Calif).*, vol. 50, no. 10, pp. 58–67, 2017, doi: 10.1109/MC.2017.3641638.

[16] T. Yu, B. Zhou, Q. Li, R. Liu, W. Wang, and C. Chang, "The design of distributed real-time video analytic system," *Proc. first Int. Work. Cloud data Manag.*, pp. 49–52, 2009, doi: 10.1145/1651263.1651273.

[17] J. Wulff, L. Sevilla-Lara, and M. J. Black, "Optical flow in mostly rigid scenes," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 6911–6920, 2017, doi: 10.1109/CVPR.2017.731.

[18] R. V. Babu, M. Tom, and P. Wadekar, "A survey on compressed domain video analysis techniques," *Multimed Tools Appl.*, vol. 75, pp. 1043–1078, 2016, doi: 10.1007/s11042-014-2345-z.

[19] L. Di Stefano, G. Neri, and E. Viarani, "Analysis of Pixel-Level Algorithms for Video Surveillance Applications," *Int. Conf. Image Anal. Process.*, pp. 541–546, 2001, doi: 10.1109/ICIAP.2001.957066.

[20] W. Shi *et al.*, "Real-Time Single Image and VideoSsuper-Resolution Using An Efficient Sub-Pixel Convolutional Neural Network," *IEEE Conf. Comput. Vis. pattern Recognit.*, pp. 1874–1883, 2016, doi: 10.1109/cvpr.2016.207.

[21] X. Cao, F. Wang, B. Zhang, H. Fu, and C. Li, "unsupervisd pixel-level video foreground object segmentation via shortest path algorithm," *Neurocomputing*, 2015, doi: 10.1016/j.neucom.2014.12.105.

[22] A. E. Rolls *et al.*, "The use of video motion analysis to determine the impact of anatomic complexity on endovascular performance in carotid artery stenting," *J. Vasc. Surg.*, vol. 69, no. 5, pp. 1482–1489, 2019, doi: 10.1016/j.jvs.2018.07.063.

[23] P. V. K. Borges, N. Conci, and A. Cavallaro, "Video-based human behavior understanding: A survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 11, pp. 1993–2008, 2013, doi: 10.1109/TCSVT.2013.2270402.

[24] O. Patsadu, C. Nukoolkit, and B. Watanapa, "Human gesture recognition using Kinect camera," *JCSSE 2012 - 9th Int. Jt. Conf. Comput. Sci. Softw. Eng.*, pp. 28–32, 2012, doi: 10.1109/JCSSE.2012.6261920.

[25] Y. Xu, J. Dong, B. Zhang, and D. Xu, "Background modeling methods in video analysis: A review and comparative evaluation," *CAAI Trans. Intell. Technol.*, vol. 1, no. 1, pp. 43–60, 2016, doi: 10.1016/j.trit.2016.03.005.

[26] N. C. Tay, T. Connie, T. S. Ong, K. O. M. Goh, and P. S. Teh, "A robust abnormal behavior detection method using convolutional neural network," *Lect. Notes Electr. Eng.*, vol. 481, pp. 37–47, 2019, doi: 10.1007/978-981-13-2622-6_4.

[27] M. Razavi, H. Sadoghi Yazdi, and A. H. Taherinia, "Crowd analysis using Bayesian Risk Kernel Density Estimation," *Eng. Appl. Artif. Intell.*, vol. 82, no. April, pp. 282–293, 2019, doi: 10.1016/j.engappai.2019.04.011.

[28] H. Bouma, J. Baan, S. Landsmeer, C. Kruszynski, G. van Antwerpen, and J. Dijk, "Real-time tracking and fast retrieval of persons in multiple surveillance cameras of a shopping mall," *Multisensor, Multisource Inf. Fusion Archit. Algorithms, Appl. 2013*, vol. 8756, no. June 2014, p. 87560A, 2013, doi: 10.1117/12.2016090.

[29] K. V. Sriram and R. H. Havaldar, "Human detection and tracking in video surveillance system," *2016 IEEE Int. Conf. Comput. Intell. Comput. Res. ICCIC 2016*, no. 1, pp. 1–3, 2016, doi: 10.1109/ICCIC.2016.7919692.

[30] S. Xie, X. Zhang, and J. Cai, "Video crowd detection and abnormal behavior model detection based on machine learning method," *Neural Comput. Appl.*, vol. 31, no. s1, pp. 175–184, 2019, doi: 10.1007/s00521-018-3692-x.

[31] M. Kesraoui, D. Acheli, and H. Chebi, "Crowd events recognition in a video without threshold value setting," *Int. J. Appl. Pattern Recognit.*, vol. 5, no. 2, p. 101, 2018, doi: 10.1504/ijapr.2018.10013764.

[32] W. C. Hu, C. H. Chen, T. Y. Chen, D. Y. Huang, and Z. C. Wu, "Moving object detection and tracking from video captured by moving camera," *J. Vis. Commun. Image Represent.*, vol. 30, pp. 164–180, 2015, doi: 10.1016/j.jvcir.2015.03.003.

[33] S. Kamal, A. Jalal, and D. Kim, "Depth Images-based Human Detection , Tracking and Activity," *J Electr Eng Technol*, vol. 11, pp. 1921–1926, 2016, doi: 10.5370/JEET.2016.11.6.1857.

[34] V. Gajjar, Y. Khandhediya, and A. Gurnani, "Human detection and tracking for video surveillance: A cognitive science approach," *Proc. - 2017 IEEE Int. Conf. Comput. Vis. Work. ICCVW 2017*, vol. 2018-Janua, pp. 2805–2809, 2018, doi: 10.1109/ICCVW.2017.330.

[35] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, A. Bouridane, and A. Beghdadi, "A combined multiple action recognition and summarization for surveillance video sequences," *Appl. Intell.*, vol. 51, no. 2, pp. 690–712, 2021, doi: 10.1007/s10489-020-01823-z.

[36] Y. Kong and Y. Fu, *Human Action Recognition and Prediction: A Survey*, vol. 130, no. 5. Springer US, 2022. doi: 10.1007/s11263-022-01594-9.

[37] P. Pareek and A. Thakkar, *A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications*, vol. 54, no. 3. Springer Netherlands, 2021. doi: 10.1007/s10462-020-09904-8.

[38] M. H. Arshad, M. Bilal, and A. Gani, "Human Activity Recognition: Review, Taxonomy and Open Challenges," *Sensors*, vol. 22, no. 17, pp. 1–33, 2022, doi: 10.3390/s22176463.

[39] M. G. Pallewar, V. R. Pawar, and A. N. Gaikwad, "Human Anomalous Activity detection with CNN-LSTM approach," *J. Integr. Sci. Technol.*, vol. 12, no. 1, pp. 1–7, 2024.

[40] L. Luo, Y. Li, H. Yin, S. Xie, R. Hu, and W. Cai, "Crowd-Level Abnormal Behavior Detection via Multi-Scale Motion Consistency Learning," *Proc. 37th AAAI Conf. Artif. Intell. AAAI 2023*, vol. 37, pp. 8984–8992, 2023, doi: 10.1609/aaai.v37i7.26079.

[41] T. Alafif, B. Alzahrani, Y. Cao, R. Alotaibi, A. Barnawi, and M. Chen, "Generative adversarial network based abnormal behavior detection in massive crowd videos: a Hajj case study," *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 8, pp. 4077–4088, 2022, doi: 10.1007/s12652-021-03323-5.

[42] A. A. Khan *et al.*, "Crowd Anomaly Detection in Video Frames Using Fine-Tuned AlexNet Model," *Electron.*, vol. 11, no. 19, 2022, doi: 10.3390/electronics11193105.

[43] R. Kaur and S. Singh, "Background modelling, detection and tracking of human in video surveillance system," *Proc. Int. Conf. Innov. Appl. Comput. Intell. Power, Energy Control. with Their Impact Humanit. CIPECH 2014*, no. November, pp. 54–58, 2014, doi: 10.1109/CIPECH.2014.7019097.

[44] S. R. Goniwada, *Introduction to Datafication*. Apress, Berkeley, CA, 2023. doi: 10.1007/978-1-4842-9496-3_7.

[45] Jennifer L. Austin, "25 Essential Skills for the Successful Behavior Analyst," in *School-Based Behavior Analysis*, 2nd Editio., Taylor & Francis Online, 2023, p. 13.

[46] C.-H. S. C.-J. L. T.-S. W. P.-T. L. C.-Y. Shih, "Behavior Analysis based on Local Object Tracking and its Bed-exit Application," in *IEEE 4th International Conference on Knowledge Innovation and Invention (ICKII)*, 2021, pp. 101–104. doi: 10.1109/ICKII51822.2021.9574741.

[47] R. Arroyo, J. J. Yebes, L. M. Bergasa, I. G. Daza, and J. Almazán, "Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls," *Expert Syst. Appl.*, vol. 42, no. 21, pp. 7991–8005, 2015, doi: 10.1016/j.eswa.2015.06.016.

[48] Y. Song, L. L. Morency, L. Angeles, R. Davis, I. P. Recognition, and A. Signal, "Multimodal Human Behavior Analysis : Learning Correlation and Interaction Across Modalities," *Proc. 14th Int. Conf. Multimodal Interact.*, pp. 27–30, 2012, doi: 10.1145/2388676.2388684.

[49] K. Bousmalis, S. Zafeiriou, L. P. Morency, and M. Pantic, "Infinite hidden conditional random fields for human behavior analysis," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 24, no. 1, pp. 170–177, 2013, doi: 10.1109/TNNLS.2012.2224882.

[50] A. Quattoni, S. Wang, L.-P. Morency, M. Collins, and T. Darrell, "Hidden-state conditional random fields.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1848–1853, 2007, doi: 10.1109/TPAMI.2007.1124.

[51] A. Gunawardana, M. Mahajan, A. Acero, and J. C. Platt, "Hidden Conditional Random Fields for Phone Classification," *Ninth Eur. Conf. Speech Commun. Technol.*, pp. 1117–1120, 2005, [Online]. Available: /home/kermorvant/refbase_files/2005/hiddenconditionalrandomfieldsforphoneclassification/1308_hiddenconditionalrandomfieldsforphoneclassification2005.pdf

[52] C. J. Yang, T. Chou, F. A. Chang, C. Ssu-Yuan, and J. I. Guo, "A smart surveillance system with multiple people detection, tracking, and behavior analysis," *2016 Int. Symp. VLSI Des. Autom. Test, VLSI-DAT 2016*, 2016, doi: 10.1109/VLSI-DAT.2016.7482569.

[53] R. Yang, M. Ding, X. Zhang, and X. Jiang, "Research on Behavior Recognition in Infrared video," *Adv. Comput. Sci. Res.*, vol. 232–236, no. Cnci, pp. 2276–2280, 2019, doi: 10.4028/www.scientific.net/amr.255-260.2276.

[54] B. Zhai, "Identification of Abnormal Human Behavior in Intelligent Video Surveillance System," vol. 173, no. Wartia, pp. 318–324, 2018, doi: 10.2991/wartia-18.2018.58.

[55] S. E. R. Egnor and K. Branson, "Computational Analysis of Behavior," *Annu. Rev. Neurosci.*, vol. 39, no. 1, pp. 217–236, 2016, doi: 10.1146/annurev-neuro-070815-013845.

[56] D. Li, L. Qian, L. Peng, Z. Zhu, P. Bai, and X. Wang, "Detection and Analysis of Human Behavior in Video Monitoring," *Int. Conf. Wirel. Commun. Netw. Multimed. Eng.*, vol. 89, no. Wcnme, pp. 213–215, 2019, doi: 10.2991/wcnme-19.2019.51.

[57] K. Aurangzeb *et al.*, "Human Behavior Analysis Based on Multi-Types Features Fusion and Von Nauman Entropy Based Features Reduction," *J. Med. Imaging Heal. Informatics*, vol. 9, pp. 662–669, 2019, doi: 10.1166/jmihi.2019.2611.

[58] M. Sivabalakrishnan, R. Menaka, and S. Jeeva, "Smart Video Surveillance Systems and Identification of Human Behavior Analysis," *Countering Cyber Attacks Preserv. Integr. Availab. Crit. Syst.*, pp. 64–97, 2019, doi: 10.4018/978-1-5225-8241-0.ch004.

[59] Z. Lv, L. Qiao, and A. K. Singh, "Advanced Machine Learning on Cognitive Computing for Human Behavior Analysis," *IEEE Trans. Comput. Soc. Syst.*, vol. 8, no. 5, pp. 1194–1202, 2021, doi: 10.1109/TCSS.2020.3011158.

[60] B. Tyagi, S. Nigam, and R. Singh, "A Review of Deep Learning Techniques for Crowd Behavior Analysis," *Arch. Comput. Methods Eng.*, vol. 29, no. 7, pp. 5427–5455, 2022, doi: 10.1007/s11831-022-09772-1.

[61] P. G. I. M. Chandrasekara, L. L. G. Chathuranga, K. A. A. Chathurangi, D. M. K. N. Seneviratna, and R. M. K. T. Rathnayaka, "Intelligent Video Surveillance Mechanisms for Abnormal Activity Recognition in Real-Time: A Systematic Literature Review," *KDU J. Multidiscip. Stud.*, vol. 5, no. 1, pp. 26–40, 2023, doi: 10.4038/kjms.v5i1.60.

[62] E. Hatirnaz, M. Sah, and C. Direkoglu, "A novel framework and concept-based semantic search Interface for abnormal crowd behaviour analysis in surveillance videos," *Multimed. Tools Appl.*, vol. 79, no. 25–26, pp. 17579–17617, 2020, doi: 10.1007/s11042-020-08659-2.

[63] S. Vahora, K. Galiya, H. Sapariya, and S. Varshney, "Comprehensive Analysis Of Crowd Behavior Techniques: A Thorough Exploration," *Int. J. Comput. Digit. Syst.*, vol. 11, no. 1, pp. 991–1007, 2022, doi: 10.12785/ijcds/110181.

[64] M. Chhirolya and N. Dubey, "Abnormal Human Behavior Detection and Classification In Crowd Using Image Processing," in *International Journal of Trend in Research and Development (IJTRD)*, 2021, no. May, pp. 11–12.

[65] S. Altowairqi, S. Luo, and P. Greer, "A Review of the Recent Progress on Crowd Anomaly Detection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 4, pp. 659–669, 2023, doi: 10.14569/IJACSA.2023.0140472.

[66] A. A. H. Altalbi, S. H. Shaker, and A. E. Ali, "Anomaly Detection from Crowded Video by Convolutional Neural Network and Descriptors Algorithm: Survey," *Int. J. online Biomed. Eng.*, vol. 19, no. 7, pp. 4–25, 2023, doi: 10.3991/ijoe.v19i07.38871.

[67] M. S. Q. & W. M. P. van der A. Alessandro Berti, Johannes Herforth, "Evolving graph-based video crowd anomaly detection," *Vis. Comput. Int. J. Comput. Graph.*, vol. Volume 40, no. January 2024, p. pages 303–318, 2024, doi: 10.1007/s00371-023-02783-4.

[68] C. Direkoglu, M. Sah, and N. E. O'Connor, "Abnormal crowd behavior detection using novel optical flow-based features," *2017 14th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2017*, 2017, doi: 10.1109/AVSS.2017.8078503.

[69] "Detection of Unusual Crowd Activity." https://mha.cs.umn.edu/proj_events.shtml#crowd

[70] "UCSD Anomaly Detection Dataset." http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm

[71] "Avenue Dataset for Abnormal Event Detection." http://www.cse.cuhk.edu.hk/leojia/projects/detectabnormal/dataset.html

[72] "UCF Crime Dataset." https://www.kaggle.com/datasets/odins0n/ucf-crime-dataset

[73] A. B. & M. C. Tarik Alafif, Bander Alzahrani, Yong Cao, Reem Alotaibi, "Generative adversarial network based abnormal behavior detection in massive crowd videos: a Hajj case study," *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 2022, pp. 4077–4088, 2021, doi: 10.1007/s12652-021-03323-5.

[74] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 589–597, 2016, doi: 10.1109/CVPR.2016.70.

[75] A. A. Shah, "A Machine Learning Model for Crowd Density Classification in Hajj Video Frames," International Journal of Advanced Computer Science and Applications, vol. 15, no. 12, 2024.