Volume: 17 · Issue: 2
May-December 2020
Pages 1-9

CULCYT
CULTURA CIENTÍFICA Y TECNOLÓGICA
ISSN (electrónico) 2007-0411

REVIEW ARTICLE

DOI: 10.20983/culcyt.2020.2.3.1

Intelligent Surveillance Systems: A Review

Sistemas de vigilancia Inteligente: una Revisión Bibliográfica

A. Mariscal-Torres[1], *L. Ortega-Máynez*[1], *Jose Mejia*[1]
[1] Universidad Autónoma de Ciudad Juárez

ABSTRACT

Security refers to the perceptions about an environment protection, it means without worry of suffer harm. This research offers a literature review about security subject, focused on autonomous surveillance, gathering in a single document the technical novelties about surveillance systems, their applications, and central components. During this research , we observe that deep learning its being applied for surveillance purpose, opening new research horizons, in an area which does not have been significant changes during about ten years, and we also found that new vast datasets are being produced to solve issues regarding security. We have also seen that, in terms of security, deep learning is highly viable to solve problems that have been implicit in security systems for a long time, this being able to turn deep learning into a new breakthrough with respect to systems programmed only by traditional vision algorithms, opening the possibility of becoming a mandatory accessory for security of systems. This research has been limited only on civil area surveillance systems, also we only use scientific articles for this, avoiding commercial technologies.

KEYWORDS: smart surveillance, deep learning, security.

RESUMEN

La seguridad es la percepción que se tiene en algún entorno, de estar protegido, sin miedo a sufrir algún daño. Este documento ofrece una revisión de literatura sobre el tema de seguridad inteligente, basada en técnicas de inteligencia artificial enfocada a vigilancia autónoma, y reúne avances técnicos de sistemas de supervisión, sus aplicaciones y componentes centrales. En el transcurso de esta investigación, se observó que el aprendizaje profundo está siendo aplicado para propósitos de vigilancia, abriendo nuevos horizontes de investigación en un área que no había tenido cambios significativos durante aproximadamente diez años. Se encontró también que se han estado produciendo bases de datos, de tamaño vasto, para resolver problemas referentes a la seguridad y probar los algoritmos. Hemos visto, además, que en cuestión de seguridad el aprendizaje profundo es altamente viable para solventar problemas que durante mucho tiempo han estado implícitos en los sistemas de seguridad, pudiendo esto convertirse en un avance significativo con respecto a los sistemas programados únicamente por algoritmos de visión tradicionales y abrirse la posibilidad de que se convierta en un accesorio obligatorio para esta clase de sistemas. Cabe mencionar que, debido a lo amplio del tema, se ha limitado esta investigación únicamente a la búsqueda de sistemas de vigilancia de ámbito civil y, además, se han utilizado solo artículos científicos, evitando tecnologías comerciales.

PALABRAS CLAVE: seguridad inteligente; aprendizaje profundo; seguridad.

Correspondencia:

DESTINATARIO: José Manuel Mejía Muñoz
INSTITUCIÓN: Universidad Autónoma de Ciudad Juárez, Instituto de Ingeniería y Tecnología
DIRECCIÓN: Av. del Charro no. 450 norte, col. Partido Romero, Ciudad Juárez, Chihuahua, México, C. P. 32310
CORREO ELECTRÓNICO: jose.mejia@uacj.mx

Fecha de recepción: 26 de agosto de 2020. Fecha de aceptación: 26 de noviembre de 2020. Fecha de publicación: 2 de diciembre de 2020.

acceso abierto
UACJ

REVISIÓN POR PARES

I. INTRODUCTION

The surveillance and intrusion prevention area has always been an active area. Since ancient times, security measures have always been important for the protection and survival of the mankind. During the passage of time, man has always been inventing ways to detect risk situations more effectively, using equipment such as: audible systems, like bells; visual elements such as movement ropes, torches, or signals emitted by people; and lately vision equipment, which allows it to detect intrusion of wildlife or strange people.

Currently, with the so-called fourth industrial revolution [1], [2], [3], which promotes the integration of massive amounts of information with existing equipment, in addition to its analysis through the use of artificial intelligence technologies, there is the premise that complete areas of the industry will be improved in a way that has not been seen before. This also includes the security aspect [4], where intelligent surveillance applications begin to be implemented in an increasingly accelerated manner [5], through applications focused not only on the industrial but also on the civil sphere. Some applications that are in development or have been implemented, are aimed, for example, to know where the industrial workers really are or the surveillance of restricted areas.

There are several works that describe modern aspects of surveillance, for example, in [6], [7] several of the important characteristics of surveillance systems are described, which can be used to evaluate current systems. Here we list a set of common characteristics for a surveillance system and some of the applications of a surveillance system.

- **Home health care/ home monitoring**: such as geriatric care and home hospitalization [8], [9]. These techniques could use surveillance systems to be aware of any abnormal behavior of patients, in addition to a more precise follow-up for people with diseases that require more attention such as Alzheimer's. This would help reduce important costs, such as having a permanent nurse at home [10], a camera could be able to learn actions such as falls, extreme anxiety, vomiting and could even detect serious injuries [11], [12].

- **Intrusion detection**: Detecting a person trespassing the property limits would help to take appropriate actions such as calling the 911 emergency system, locking doors, or activating an audible alarm [13]-[14].

- **Animal Intrusion**: Detect when a predator is on an urban property [15] or rural [16], and determine if that specific animal needs a call to the police. The systems could have a list of local animals that require to be reported to police departments, especially if they are predators, and this can help the authorities to act before any attack occurs, increasing the chances that the animal will survive, reducing the human impact, and saving people from an attack [17].

- **Home assault**: Detecting a person or group of people with the intention of assaulting a resident of a house [18]-[19] and can help the authorities to have more precise alarms, sending information on how many people there are if they are armed and their hostile behavior.

- **Restrict access**: An intelligent surveillance system that can control access to specific areas without the need to carry any device and it could even trigger an alarm if a specific person is detected outside its permitted limits [20], [21].

Furthermore, according to the literature review, it is possible to obtain a general organization of intelligent surveillance systems into three general groups according to the type of algorithm used:

- Systems based on classic algorithms, which are based on a well-known and specific formula, such as hidden Markov chain algorithms, Gaussian mixing;

- Systems based on machine learning algorithms, such as those based on support vector machines and deep learning, and

- Mix of the two types or mixed.

Because of the rise of systems based on machine learning and to limit the number of works reviewed, in this study we will focus on those of the second group: systems based on machine learning algorithms.

II. METHODOLOGY

In this section, we describe the methodology to search, analyze, and select information from the literature. The present work is a narrative review, incorporating features of a systematic review such as some analysis of the literature reviewed. The methodology is described in the following subsections.

Identification of literature sources

For this work the following sources where considered: EBSCO, IEEE, Science Direct, and Google Academic databases.

Search Strategies

To limit the range of years of the reviewed works, a filter between 2016 and 2020 was employed, however, papers of different years were included when they contain relevant concepts or examples. In figure 1, we show a distribution of the range of revision. For the initial collecting, a random search method with the keywords surveillance system was performed. Afterwards, each article was carefully reviewed to localize the required information and to find relevant aspects or attributes of surveillance systems for study selection and data abstraction.

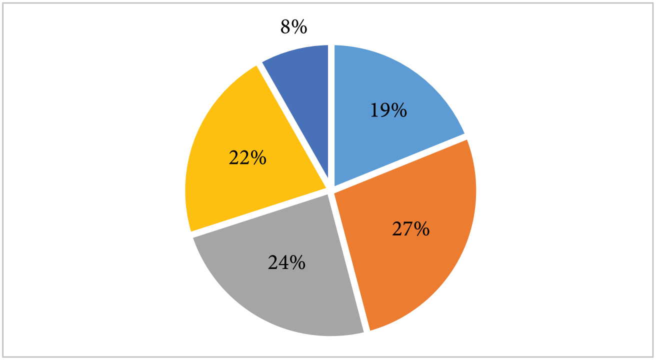

Figure 1. Distribution of the articles reviewed in the range of 2016-2020. Source: made by the authors.

III. RESULTS

There are works in the area of systems based on machine learning algorithms and here we review some of the most representative: A work by Antreas and Plamen [22], in which they propose a design of an intelligent surveillance system based on smartphones. Such system has two modules, one of detection, based on the method of background subtraction with optical flow, and the second of classification based on deep learning.

On the other hand, in Zhao *et al.* [23], a module is proposed to estimate the gaze direction of a surveillance system. The novelty of the module is that it can be trained by means of synthetic images using a generating artificial neural network of the type GAN (Generative adversarial network). Also, in the work of Dautov *et al.* [24] they combine technologies from the Internet of things, the cloud, Edge Computing and Big Data in an intelligent surveillance system and focus on the analysis of the data generated in the system's sensors. In Kumar *et al.* [25] they direct their work to an intelligent mobile video surveillance system that uses the concepts of Bayesian coalition game theory and learning automata algorithms. Their system is mainly focused on reducing the delay that occurs during the transmission of video to the closest access points and suggest how to select the best route. In Shahad *et al.* [26] an intelligent surveillance system, with Complex Event Processing technology implemented, is proposed, which is used for the detection of intrusions through data correlation. In addition, four classifiers are used in the engine to predict the occurrence of events from the recognition of patterns in the data sequence acquired from door sensors and surveillance cameras.

In Al-Nawashi *et al.* [27], it is proposed an automatic surveillance system for academic environments based on video and capable of monitoring a scene semantically and detecting anomalies. Such system consists of three modules: pre-processing, detection of abnormal human activity and content-based image retrieval phase, with support vector machine type classifiers. Wei Yang and Yen Su [28] use a convolutional neural network for image recognition, and they also want the system to be low-cost, so they use a hardware accelerator called Neural Compute Stick instead of a GPU (graphic processing unit) which is normally used in deep learning, but has a high cost.

Tang *et al.* [29] analyze the automatic detection and recognition of vehicles applied to a traffic surveillance system. To this end, they use AdaBoost algorithms to extract characteristics and build classifiers to detect the vehicle on the input image. Even more characteristics are extracted, but at different scale by means of the Gabor transform. Finally, a nearest neighbor classifier is used for the final classification.

Kurniawan *et al.* [30] propose an intelligent surveillance system to automatically analyze data from surveillance cameras using foreground-background segmentation and people detection using a mixture of Gaussian and oriented-gradient histogram and a multi-object tracking method to tracking people.

Within the algorithms examined within this group, the following main characteristics were identified: people-oriented detection, identification of hidden objects, pose estimation, behavior classification/ action recognition, and facial recognition and identification. These topics are addressed in the following sections.

People-oriented detection

In Han *et al.* [31] and Wren *et al.* [32] the problem of face-to-face detection is analyzed and they agree that one of the main problems is detecting and segmenting a human figure in the environment. Various general-purpose algorithms have been proposed for this purpose: YOLO [33], R-CNN [34], Faster-R-CNN [35] and the most recently, Mask-RCN [36]. These algorithms solve the problem using deep learning, far surpassing all conventional algorithms [37], differing among them due to speed and precision [38]. The fastest algorithm is YOLO, coming to be considered as in real time, that is, it runs in a range above 45 frames per second, one of its variants reaches up to 155 frames per second, with an average precision of 63.5 focused on people. However, the most accurate algorithm is Mask-RCN with 78.9% accuracy but running at a speed of 5 frames per second.

Identification of hidden objects

A different problem is to find a completely exposed object in contrast to find a partial hidden object. In Lao *et al.* [39] this task is resolved. Generally, this is addressed, first by the object segmentation from the background and finally distinguishing if an occlusion is present [40]. All this is resolved using histograms, non-lineal regressions, random Markov fields, and then limit detection using neural algorithms. However, recent algorithms as YOLO and Mask-RCN are capable of finding persons through statistical analysis within the neural network. However, there is no precise data on how much concealment percentage is allowed, so it is difficult to specify which algorithm is better.

Pose estimation

This feature enables the security system to catch what is registered in the image, in order to classify the type of behavior, which in Lao *et al.* [39] and Peursum *et al.* [41] is achieved using hidden Markov models. It should be considered that [39] classifies the pose, but it does not generate a skeleton in the image. Open Posse [42] uses a deep learning model, convolutional by stages, which first finds each of the identified key points to form a pose. These points consist of eye (left and right), nose, mouth, neck, right shoulder, left shoulder, right elbow, left elbow, right hand and left hand, center of trunk, right knee, left knee, right foot and left foot are also detected. These points are then joined by an algorithm called the affinity of parts, which restricts them to a certain angle and distance, which is able to successfully determine pose.

Behavior classification/ action recognition

The analysis of behavior is one of the most important topics for surveillance camera system [43]. In the work of Lao *et al.* [39] a heuristic programming method to solve this problem is proposed and, additionally, their algorithm is immune to certain types of occlusions in the scene. This is done by eliminating information about textures and only paying attention to shapes, from where it is obtained information about trajectories that subsequently helps to classify different types of posture to estimate behavior.

Facial recognition and identification

This feature is extremely useful for any intelligent surveillance system, because it improves the detection of threats, generating actions depending on each situation. In El-Sayed and Hamed [44], several ways in which this image differentiation can be carried out are listed, such as Euclidean distance, quadratic distance, distance between two blocks, Minkowski distance, among others. In Taigman *et al.* [45], a neural model is proposed, which processes the image of the face seeking to align it in 3D and making a process called “frontalization” to reduce the loss of important data for identification.

Metrics for machine learning-based algorithms

In this section, we review the metrics to quantify the performance of the various algorithms specialized in people-oriented detection and identification of hidden objects, pose estimation, behavior classification, and facial recognition and identification. In general, surveillance systems based on machine learning algorithms evaluate their performance based on the capacity of the algorithms used, so the evaluation consists of metrics that quantify their precision to classify, detect, and response time to the events to which you want the system to respond.

In [46], methods for comparing performance of different object classifiers are described. Intersection over Union (IoU) refers to the intersection area of the frame of the detected object, against the ground-truth, which consists of a region selected by a human. Figure 2 exemplifies this concept.

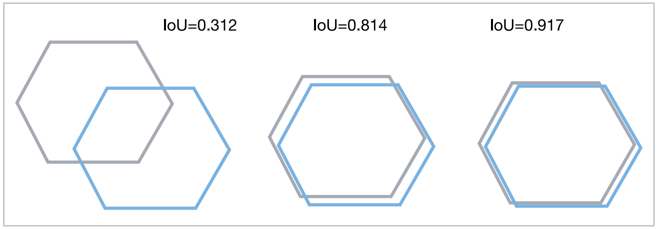

Figure 2. Intersection over union. Source: made by the authors.

AP_{IoU50}=0.50 refers to the average precision of a classifier that has managed to frame an object in an image, with an intersection above 50 percent of the ground-truth, while a more precise method is the one that obtains an AP_{IoU75}=0.75.

Mean Accuracy (MA) [47] is a metric used in object detection as a method to characterize the performance of an object detector. It corresponds to the average of the multiple intersection junctions founded. In the case of the competitions carried out using the COCO database [48], it was proposed an average of 10 AP values obtained for each object.

Another metric, Accuracy, quantifies the results of a classification or prediction operation [49], showing how good they are. The Accuracy has the following formula,

$$Accuracy = Positive\ /\ (Positive + False\ Positive)$$

It is measured from a specific region, setting how many values were correctly classified and the number of false positives obtained are punished. This metric is used in both object detection and facial recognition.

The pose estimation problem merits another metric, the one used by OpenPose being the PCK, described in [50], which is a modification of an older metric, called the percentage of correctly located parts or PCP. In this case, the overlaps obtained from detecting the parts of people's joints (elbows, shoulders, knees, etc.) are evaluated against human handmade ones. Figure 3 shows these regions for a pose identifying 14 and 32 feature parts respectively.

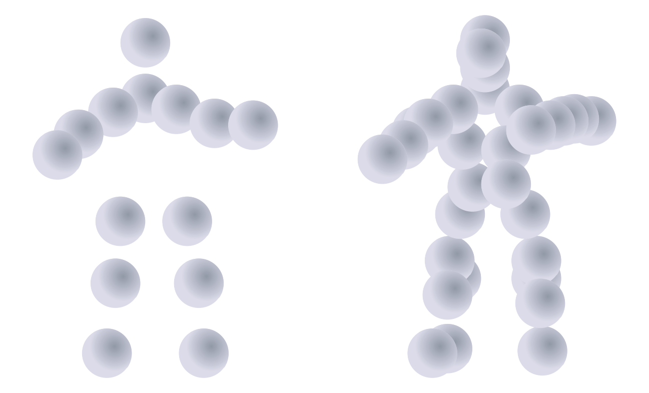

Figure 3. Pose, using 14 and 32 features. Source: made by the authors.

Comparison of algorithms used in surveillance systems

In this section we compare some of the algorithms reviewed. For this task we employ the metrics used in the original research papers, and it is possible that the metrics may not coincide between different investigations. Even so, we believe that it is relevant to make the comparison.

Results for in human detection

Table 1 shows a comparison of the detection percentage of each studied system, using the COCO and VOC databases [51]. As can be seen in Table 1, the work of [31] has better performance than the more popular architectures based on RCNN. However, RCNN-based algorithms have been tested in databases with a greater number of images, which is not the case of [31], thus, results on architectures based on RCNN are more stable while those reported in [31] could have more variation.

TABLE 1
Comparison of the Detection Percentage

Algorithm	Performance	Databases
[31] Real Time V	93.2%	Small database
[34] RCNN	58.7%	VOC 2007
[35] Fast-RCN	52.3%	VOC 2007*
[36] Mask-RCN	53.6%	COCO test

*General

Results for pose estimation

Table 2 shows a comparison of the systems focused on determining the pose of people.

TABLE 2
Comparison for Pose Estimation

Algorithm	Performance	Metric
[36] Mask-RCN	87.3%	COCO test, AP_{50}
[39] Weilun Lao	77.4%	Mean
[42] OpenPose	87.95%	COCO, LSP, PCKh-0.5

As can be seen in Table 2, none of the three systems studied have an equivalent metric, so their comparison is difficult. However, the results reported in OpenPose are tested on more databases than those reported in [39], [42]. However, OpenPose requires much more computational power to operate in real-time than [42].

Results for action recognition

Table 3 lists the results of the systems that are focused on recognizing actions.

TABLE 3
Action Recognition Algorithms

Algorithm	Performance	Metric
[33] Smart Surv.	83.33%	Mean Accuracy
[41] Peursum *et al.*	98.81%	Mean Accuracy
[52] TORNADO	96.79%	UCF-Sports>0.50

It was observed that in a certain way there are no general criteria for recognizing actions, none of the three proposed systems has the same actions to detect nor is it compatible with the same type of database.

Results for facial recognition algorithms

In this section we compare two facial recognition systems: Simil [44] and DeepFace [45]. In both systems different databases were used. In [44] it is used the ORL database, which consists of 40 people, 10 images per person, while in [45] are used two large databases, the LFW database with 5749 and the YTF database with 3425 videos, 1595 different people. The results are shown in Table 4.

TABLE 4
Comparison for Facial Recognition Systems

Algorithm	Performance	Metric
[44] Simil.	90.0%	Mean Accuracy
[45] DeepFace	97.5%	Mean Accuracy

IV. CONCLUSIONS

From all the articles reviewed, we conclude with several remarks. The first of these is that the surveillance area continues to be an active topic to investigate; there still exist an abundance of problems raised which needs to be solved or improved. Second, it is observed a sustained improvement on the applications and systems. This is mainly due to deep learning techniques, such as the use of convolutional neural networks, of two and three dimensions, applied to surveillance systems. An important point is the available databases, both in quantity and quality, which make developing systems and algorithms sometimes difficult to evaluate. Finally, it would be interesting to see if the identification of people, as part of a surveillance system, improves or not the level of protection of an entity, which would imply the design of a series of tests to quantify this item.

REFERENCES

[1] H. Lasi, P. Fettke, H. G. Kemper. T. Feld and M. Hoffmann, "Industry 4.0," *Bus. Inf. Syst. Eng.*, vol. 6, no. 4, pp. 239-242, Jun. 19, 2014. DOI: s12599-014-0334-4.

[2] S. Vaidya, P. Ambad and S. Bhosle, "Industry 4.0 – A Glimpse," *Procedia Manufacturing*, vol. 20, pp. 233-238, 2018. DOI: j.promfg.2018.02.034.

[3] T. Pereira, L. Barreto and A. Amaral, "Network and information security challenges within Industry 4.0 paradigm," *Procedia manufacturing*, vol. 13, pp. 1253-1260, 2017. DOI: j.promfg.2017.09.047.

[4] H. Flatt, S. Schriegel, J. Jasperneite, H. Trsek and H. Adamczyk, "Analysis of the Cyber-Security of industry 4.0 technologies based on RAMI 4.0 and identification of requirements," *2016 IEEE 21st International*

Conference on Emerging Technologies and Factory Automation (ETFA), Berlin, 2016, pp. 1-4, DOI: 10.1109/ETFA.2016.7733634.

[5] S. R. Chhetri, N. Rashid, S. Faezi and M. A. Al Faruque, "Security trends and advances in manufacturing systems in the era of industry 4.0," *2017 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, Irvine, CA, 2017, pp. 1039-1046, DOI: 10.1109/ICCAD.2017.8203896.

[6] Weiming Hu, Tieniu Tan, Liang Wang and S. Maybank, "A survey on visual surveillance of object motion and behaviors," in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 34, no. 3, pp. 334-352, Aug. 2004, DOI: 10.1109/TSMCC.2004.829274.

[7] M. Andrejevic, "Automating surveillance," *Surveillance & Society*, vol. 17, no 1/2, pp. 7-13, 2019, DOI: 10.24908/ss.v17i1/2.12930.

[8] V. J. González-Ramallo and A. Segado-Soriano, "Veinticinco años de hospitalización a domicilio en España," *Med. Clín.*, vol. 126, no. 9, pp. 332-333, 2006, DOI: 10.1157/13085746.

[9] T. M. Llorente, P. S. Gallardo, C. D. R. Fernández and R. M. Alba, "Repercusiones en el cuidador principal del niño hospitalizado a domicilio en cuidados paliativos pediátricos," *Medicina Paliativa*, vol. 23, no. 2, pp. 79-92, 2016, DOI: 10.1016/j.medipa.2013.12.004.

[10] Z. Zhang, S. Ishida, S. Tagashira and A. Fukuda, "Danger-pose detection system using commodity Wi-Fi for bathroom monitoring," *Sensors*, vol. 19, no 4, pp. 884, 2019, DOI: 10.3390/s19040884.

[11] A. Weintraub, D. Gregory, A. R. Patel, D. Levine, D. Venesy, K. Perry and M. A. Konstam, "A multicenter randomized controlled evaluation of automated home monitoring and telephonic disease management in patients recently hospitalized for congestive heart failure: the SPAN-CHF II trial," *J. Card. Fail.*, vol. 16, no. 4, pp. 285-292, 2010. DOI: 10.1016/j.cardfail.2009.12.012.

[12] M. H. Kuo, S- L. Wang and W. T. Chen, "Using information and mobile technology improved elderly home care services," *Health Policy and Technology*, vol. 5, no 2, p. 131-142, 2016, DOI: 10.1016/j.hlpt.2016.02.004.

[13] I. Sharafaldin, A. H. Lashkari and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," *4th International Conference on Information Systems Security and Privacy - Volume I: ICISSP*, Funchal, Madeira, Portugal, 2018, pp. 108-116, DOI: 10.5220/0006639801080116.

[14] R. A. Raza Ashfaq, X-Z. Wang, J. Zhexue Huang, H. Abbas and Y-L. He, "Fuzziness based semi-supervised learning approach for intrusion detection system," *Information Sciences*, vol. 378, pp. 484-497, February 2017, DOI: 10.1016/j.ins.2016.04.019.

[15] S. Radhakrishnan and R. A. Ramanathan, "A Support Vector Machine with Gabor Features for Animal Intrusion Detection in Agriculture Fields," *Procedia computer science*, vol. 143, p. 493-501, 2018, DOI: 10.1016/j.procs.2018.10.422.

[16] R. Nikhil, B. S. Anisha and R. Kumar P., "Real-Time Monitoring of Agricultural Land with Crop Prediction and Animal Intrusion Prevention using Internet of Things and Machine Learning at Edge," *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, Bangalore, India, 2020, pp. 1-6, DOI: 10.1109/CONECCT50063.2020.9198508.

[17] N. Manohar, Y. H. Sharath Kumar and G. Hemantha Kumar, "An Approach for the Development of Animal Tracking System," *International Journal of Computer Vision and Image Processing (IJCVIP)*, vol. 8, no 1, p. 15-31, 2018, DOI: 10.4018/IJCVIP.2018010102.

[18] Z. Sun, S. Tang, H. Huang, Z. Zhu, H. Guo, Y. Sun and L. Huang, "SOS: Real-time and accurate physical assault detection using smartphone," *Peer-to-Peer Netw. Appl.*, vol. 10, no 2, pp. 395-410, 2017. DOI: 10.1007/s12083-016-0428-5.

[19] Ch. Gayathri Harshitha, M. Kameswara Rao and P. Neelesh Kumar, "A Novel Mechanism for Host-Based Intrusion Detection System," *First International Conference on Sustainable Technologies for Computational Intelligence (ICTSCI 2019)*, Jaipur, Rajasthan, India, 2019, pp. 527-536, DOI: 10.1007/978-981-15-0029-9_42.

[20] C. J. Wilkinson, "Airport Staff Access Control: Biometrics at Last?," *2018 International Carnahan Conference on*

Security Technology (ICCST), Montreal, QC, 2018, pp. 1-8, DOI: 10.1109/CCST.2018.8585592.

[21] B. Bowling and S. Westenra, "'A really hostile environment': Adiaphorization, global policing and the crimmigration control system," *Theoretical Criminology*, vol. 24, no. 2, pp. 163-183, 2018, DOI: 10.1177/1362480618774034.

[22] A. Antoniou and P. Angelov, "A general purpose intelligent surveillance system for mobile devices using Deep Learning," *2016 International Joint Conference on Neural Networks (IJCNN)*, Vancouver, BC, 2016, pp. 2879-2886, DOI: 10.1109/IJCNN.2016.7727563.

[23] T. Zhao, Y. Yan, J. Peng, Z. Mi and X. Fu, "Guiding intelligent surveillance system by learning-by-synthesis gaze estimation," *Pattern Recognit. Lett.*, 125, pp. 556-562, 2018.

[24] R. Dautov, S. Distefano, G. Merlino, D. Bruneo, F. Longo and A. Puliafito, "Towards a global intelligent surveillance system," *2017: Proceedings of the 11th International Conference on Distributed Smart Cameras (ICDSC)*, Stanford CA USA, September 2017, pp. 119-124, DOI: 10.1145/3131885.3131918.

[25] N. Kumar, J. Lee and J. J. P. C. Rodrigues, "Intelligent Mobile Video Surveillance System as a Bayesian Coalition Game in Vehicular Sensor Networks: Learning Automata Approach," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1148-1161, June 2015, DOI: 10.1109/TITS.2014.2354372.

[26] R. A. Shahad, L. G. Bein, M. H. M. Saad and A. Hussain, "Complex event detection in an intelligent surveillance system using CAISER platform," *2016 International Conference on Advances in Electrical, Electronic and Systems Engineering (ICAEES)*, Putrajaya, 2016, pp. 129-133, DOI: 10.1109/ICAEES.2016.7888023.

[27] M. Al-Nawashi, O. M. Al-Hazaimeh and M. Saraee, "A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments," *Neural Comput. & Applic.*, 28, pp. 565-572, 2017, DOI: 10.1007/s00521-016-2363-z

[28] L. Wei Yang and C. Yen Su, "Low-Cost CNN Design for Intelligent Surveillance System," *2018 International Conference on System Science and Engineering (ICSSE)*, New Taipei, 2018, pp. 1-4, DOI: 10.1109/ICSSE.2018.8520133

[29] Y. Tang, C. Zhang, R. Gu, P. Li and B. Yang, "Vehicle detection and recognition for intelligent traffic surveillance system," *Multimedia tools and applications*, vol. 76, no. 4, pp. 5817-5832, 2017, DOI: 10.1007/s11042-015-2520-x.

[30] W. Kurniawan, S. Ibrahim and M. Sulistyo, "People detection and tracking methods for intelligent surveillance system," *AIP Conference Proceedings*, vol. 2217, no. 1, p. 030110, Apr. 2020, DOI: 10.1063/5.0001022.

[31] J. Han, M. Feng and P. H. N. de With, "A real-time video surveillance system with human occlusion handling using nonlinear regression," *2008 IEEE International Conference on Multimedia and Expo*, Hannover, 2008, pp. 305-308, DOI: 10.1109/ICME.2008.4607432.

[32] C. R. Wren, A. Azarbayejani, T. Darrell and A. P. Pentland, "Pfinder: real-time tracking of the human body," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780-785, July 1997, DOI: 10.1109/34.598236.

[33] K. E. Ko and K. B. Sim, "Deep convolutional framework for abnormal behavior detection in a smart surveillance system," *Engineering Applications of Artificial Intelligence*, vol. 67, pp. 226-234, 2018, DOI: 10.1016/j.engappai.2017.10.001.

[34] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, 2014, pp. 580-587, DOI: 10.1109/CVPR.2014.81.

[35] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", *arXiv:1506.01497*, Jan. 2016. Available: https://aps.arxiv.org/abs/1506.01497v2.

[36] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386-397, 1 Feb. 2020, DOI: 10.1109/TPAMI.2018.2844175.

[37] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *2005 IEEE Computer Society*

Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 2005, pp. 886-893 vol. 1, DOI: 10.1109/CVPR.2005.177.

[38] Y. Guo, Y. Liu, T. Georgiou and M. S. Lew, "A review of semantic segmentation using deep neural networks," *Int. J. Multimed. Info. Retr.*, 2018, vol. 7, no 2, p. 87-93, DOI: 10.1007/s13735-017-0141-z.

[39] W. Lao, J. Han and P. H. n. De With, "Automatic video-based human motion analyzer for consumer surveillance system," in *IEEE Transactions on Consumer Electronics*, vol. 55, no. 2, pp. 591-598, May 2009, DOI: 10.1109/TCE.2009.5174427.

[40] K. Rajaei, Y. Mohsenzadeh, R. Ebrahimpour, S. M. Khaligh-Razavi, "Beyond core object recognition: Recurrent processes account for object recognition under occlusion," *PLoS Comput. Biol.*, vol. 15, no 5, p. e1007001, 2019, DOI: 10.1371/journal.pcbi.1007001.

[41] P. Peursum, H. H. Bui, S. Venkatesh and Geoff West, "Robust recognition and segmentation of human actions using HMMs with missing observations," *EURASIP J. Adv. Signal Process*, no. 870416, 2005, DOI: 10.1155/ASP.2005.2110.

[42] S. Wei, V. Ramakrishna, T. Kanade and Y. Sheikh, "Convolutional Pose Machines," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 2016, pp. 4724-4732, DOI: 10.1109/CVPR.2016.511.

[43] J. Liu, Y. Gu and S. Kamijo, "Customer behavior classification using surveillance camera for marketing," *Multimed. Tools Appl.*, vol. 76, no 5, p. 6595-6622, 2017, DOI: 10.1007/s11042-016-3342-1.

[44] M. A. El-Sayed and K. Hamed, "Study of Similarity Measures with Linear Discriminant Analysis for Face Recognition," *Journal of Software Engineering and Applications*, vol. 8, no. 9, pp. 478–488, 2015, DOI: 10.4236/jsea.2015.89046.

[45] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, 2014, pp. 1701-1708, DOI: 10.1109/CVPR.2014.220.

[46] "*What is COCO?*," 2020. Accessed on: 2020. [Online]. Available: http://cocodataset.org.

[47] F. Cakir, K. He, X. Xia, B. Kulis and S. Sclaroff, "Deep Metric Learning to Rank," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 1861-1870, DOI: 10.1109/CVPR.2019.00196.

[48] D. Puri, "COCO Dataset Stuff Segmentation Challenge," *2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*, Pune, India, 2019, pp. 1-5, DOI: 10.1109/ICCUBEA47591.2019.9129255.

[49] J. Hui, *mAP (mean Average Precision) for Object Detection*, Johathan Hui web page, Mar. 6, 2018. Accessed on: April 3, 2019. [Online]. Available: https://medium.com/@jonathan_hui/map-mean-average-precision-for-object-detection-45c121a31173.

[50] Y. Yang and D. Ramanan, "Articulated Human Detection with Flexible Mixtures of Parts," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2878-2890, Dec. 2013, DOI: 10.1109/TPAMI.2012.261.

[51] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn and Andrew Zisserman, "The PASCAL Visual Object Classes (VOC) Challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303-338, 2010, DOI: 10.1007/s11263-009-0275-4.

[52] H. Zhu, R. Vial and S. Lu, "TORNADO: A Spatio-Temporal Convolutional Regression Network for Video Action Proposal," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, 2017, pp. 5814-5822, DOI: 10.1109/ICCV.2017.619.