CULTURA CIENTÍFICA Y TECNOLÓGICA • ISSN 2007-0411

Ciudad Juárez, Chihuahua, México
26 de noviembre de 2020

ADRIÁN MARISCAL TORRES
UNIVERSIDAD AUTÓNOMA DE CIUDAD JUÁREZ
PRESENTE.–

Por este conducto me permito hacer de su conocimiento que el manuscrito titulado:

INTELLIGENT SURVEILLANCE SYSTEMS: A REVIEW

en coautoría con Leticia Ortega Máynez y José Manuel Mejía Muñoz, que fue sometido para su posible publicación en la revista *Cultura Científica y Tecnológica*, ha sido **ACEPTADO** en la modalidad de **ARTÍCULO DE REVISIÓN** por el grupo editorial de la Revista y será publicado en el volumen 17, número 2, mayo-diciembre de 2020.

Una vez más, gracias por someter su manuscrito a CULCYT.

Reciba un cordial saludo.

DRA. NELLY GORDILLO CASTILLO
EDITORA EN JEFE
REVISTA CULTURA CIENTÍFICA Y TECNOLÓGICA

Artículo de revisión

CULTURA CIENTÍFICA Y TECNOLÓGICA • ISSN 2007-0411

Fecha de envío: 26 de 8 de 2020

Intelligent Surveillance Systems, a Review

Vigilancia Inteligente, una Revisión Bibliográfica

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 para 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 convertirlo en un avance significativo con respecto a los sistemas programados únicamente por algoritmos de visión tradicionales, abriendo la posibilidad de convertirse 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, además, se han utilizado únicamente artículos científicos, evitando tecnologías comerciales.

Palabras clave: seguridad inteligente; aprendizaje profundo; seguridad.

ABSTRACT

Security refers to the perceptions about an environment protection, it means without worry of suffer harm. This investigation 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, 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.

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 that a predator is on an urban property [15], or rural [16], 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, this can help the authorities to act before any attack occurs, increasing the chances that the animal will survive, reducing the human impact, 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], 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 can control access to specific areas, without the need to carry any device, 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. METODOLOGY

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.

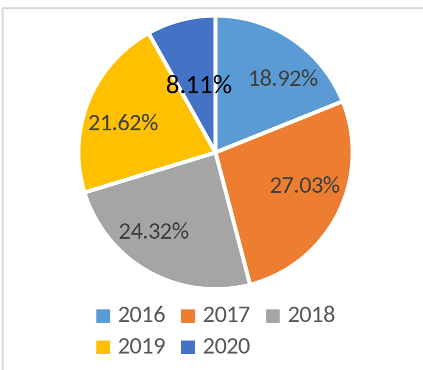

Figure 1. Distribution of the articles reviewed in the range of 2016-2020 (source: made by the authors).

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], 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 [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 [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 [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 [26] an intelligent surveillance system with Complex Event Processing technology implemented, 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 [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. In [28], they 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.

The authors of [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.

In [30] they 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 [31] and [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 [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 [39] and [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 [39] a heuristic programming method to solve this problem is proposed, 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 [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 [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.

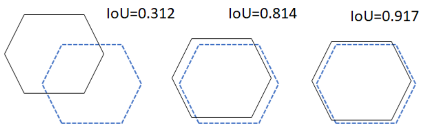

Figure 2. Intersection over union (source: made by the authors).

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.

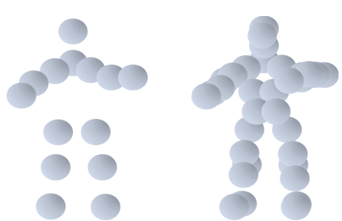

Figure 3. Pose, using 14 and 32 features (source: made by the authors).

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

TABLE 4

COMPARISON FOR FACIAL RECOGNITION SYSTEMS

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

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.

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 y M. Hoffmann, "Industry 4.0", *Business & information systems engineering*, vol. 6, no. 4, pp. 239-242, junio 19, 2014. DOI: s12599-014-0334-4

[2] S. Vaidya, P. Ambad y 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 y 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

H. Flatt, S. Schriegel, J. Jasperneite, H. Trsek y 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] Chhetri, S. R., Rashid, N., Faezi, S., & Al Faruque, M. A. (2017, November). Security trends and advances in manufacturing systems in the era of industry 4.0. In 2017 IEEE/ACM International Conference on Computer-Aided Design (ICCAD) (pp. 1039-1046). IEEE.

[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

[7] Andrejevic, Mark. Automating surveillance. Surveillance & Society, 2019, vol. 17, no 1/2, p. 7-13.

[8] González-Ramallo, Víctor J., and Antonio Segado-Soriano. "Veinticinco años de hospitalización a domicilio en España." (2006): 332-333.

[9] Llorente, T. M., Gallardo, P. S., Fernández, C. D. R., & Alba, R. M. (2016). Repercusiones en el cuidador principal del niño hospitalizado a domicilio en cuidados paliativos pediátricos. Medicina Paliativa, 23(2), 79-92.

[10] Zhang, Zizheng, et al. Danger-pose detection system using commodity Wi-Fi for bathroom monitoring. Sensors, 2019, vol. 19, no 4, p. 884.

[11] Weintraub, A., Gregory, D., Patel, A. R., Levine, D., Venesy, D., Perry, K., Konstam, M. A. (2010). 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. Journal of cardiac failure, 16(4), 285-292.

[12] Kuo, Mu-Hsing; Wang, Shu-Lin; Chen, Wei-Tu. Using information and mobile technology improved elderly home care services. Health Policy and Technology, 2016, vol. 5, no 2, p. 131-142. DOI: https://doi.org/10.1016/j.hlpt.2016.02.004

[13] Sharafaldin, Iman; Lashkari, Arash Habibi; GHORBANI, Ali A. Toward generating a new intrusion detection dataset and intrusion traffic characterization. En ICISSP. 2018. p. 108-116.

[14] Ashfaq, Rana Aamir Raza, et al. Fuzziness based semi-supervised learning approach for intrusion detection system. Information Sciences, 2017, vol. 378, p. 484-497.

[15] Radhakrishnan, Saieshwar; Ramanathan, R. A Support Vector Machine with Gabor Features for Animal Intrusion Detection in Agriculture Fields. Procedia computer science, 2018, vol. 143, p. 493-501.

[16] Nikhil, R.; Anisha, B. S.; Kumar, P. Ramakanth. Real-Time Monitoring of Agricultural Land with Crop Prediction and Animal Intrusion Prevention using Internet of Things and Machine Learning at Edge. EasyChair, 2020.

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

[18] SUN, Zehao, et al. SOS: Real-time and accurate physical assault detection using smartphone. Peer-to-Peer Networking and Applications, 2017, vol. 10, no 2, p. 395-410.

[19] HARSHITHA, Ch Gayathri; RAO, M. Kameswara; KUMAR, P. Neelesh. A Novel Mechanism for Host-Based Intrusion Detection System. En First International Conference on Sustainable Technologies for Computational Intelligence. Springer, Singapore, 2020. p. 527-536.

[20] WILKINSON, Christer J. Airport Staff Access Control: Biometrics at Last. En 2018 International Carnahan Conference on Security Technology (ICCST). IEEE, 2018. p. 1-8.

[21] BOWLING, Ben; WESTENRA, Sophie. ‘A really hostile environment’: Adiaphorization, global policing and the crimmigration control system. Theoretical Criminology, 2018, p. 1362480618774034.

[22] ANTONIOU, Antreas; ANGELOV, Plamen. A general purpose intelligent surveillance system for mobile devices using deep learning. En 2016 International Joint Conference on Neural Networks (IJCNN). IEEE, 2016. p. 2879-2886.

[23] Zhao, T., Yan, Y., Peng, J., Mi, Z., & Fu, X. (2019). Guiding intelligent surveillance system by learning-by-synthesis gaze estimation. Pattern Recognition Letters, 125, 556-562.

[24] Dautov, R., Distefano, S., Merlino, G., Bruneo, D., Longo, F., & Puliafito, A. (2017, September). Towards a global intelligent surveillance system. In Proceedings of the 11th International Conference on Distributed Smart Cameras (pp. 119-124).

[25] Kumar, Neeraj; Lee, Jong-Hyouk; Rodrigues, Joel JPC. Intelligent mobile video surveillance system as a Bayesian coalition game in vehicular sensor networks: Learning automata approach. IEEE Transactions on Intelligent Transportation Systems, 2014, vol. 16, no 3, p. 1148-1161.

[26] Shahad, R. A., Bein, L. G., Saad, M. H. M., & Hussain, A. (2016, November). Complex event detection in an intelligent surveillance system using CAISER platform. In IEEE 2016 International Conference on Advances in Electrical, Electronic and Systems Engineering (ICAEES) (pp. 129-133).

[27] Al-Nawashi, Malek, Obaida M. Al-Hazaimeh, and Mohamad Saraee. "A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments." Neural Computing and Applications 28.1 (2017): 565-572.

[28] Yang, Liang Wei, and Chung Yen Su. "Low-Cost CNN Design for Intelligent Surveillance System." 2018 International Conference on System Science and Engineering (ICSSE). IEEE, 2018.

[29] Tang, Y., Zhang, C., Gu, R., Li, P., & Yang, B. (2017). Vehicle detection and recognition for intelligent traffic surveillance system. Multimedia tools and applications, 76(4), 5817-5832.

[30] Kurniawan, W., Ibrahim, S., & Sulistyo, M. (2020, April). People detection and tracking methods

for intelligent surveillance system. In AIP Conference Proceedings (Vol. 2217, No. 1, p. 030110). AIP Publishing LLC.

[31] Jungong Han, Minwei 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.

[32] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland. Pfinder: real-time tracking of the human body. IEEETransactions on Pattern Analysis and Machine Intelligence,19(7):780–785, July 1997.

[33]. Ko, K. E., & Sim, K. B. (2018). Deep convolutional framework for abnormal behavior detection in a smart surveillance system. Engineering Applications of Artificial Intelligence, 67, 226-234. DOI: https://doi.org/10.1016/j.engappai.2017.10.001

[34] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

[35] Ren, K. He, R. Girshick, J. Sun, and R. Faster. Towards real-time object detection with region proposal networks, 2016. arXiv preprint arXiv:1506.01497.

[36] K. He, G. Gkioxari, P. Dollar, and R. Girshick. Mask r-cnn. In The IEEE International Conference on Computer vision (ICCV), Oct 2017.

[37] DALAL, Navneet; TRIGGS, Bill. Histograms of oriented gradients for human detection. En 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). IEEE, 2005. p. 886-893.

[38] GUO, Yanming, et al. A review of semantic segmentation using deep neural networks. International journal of multimedia information retrieval, 2018, vol. 7, no 2, p. 87-93.

[39] Lao, W., Han, J., & De With, P. H. (2009). Automatic video-based human motion analyzer for consumer surveillance system. IEEE Transactions on Consumer Electronics, 55(2), 591-598. DOI 10.1007/s11042-016-3342-1

[40] RAJAEI, Karim, et al. Beyond core object recognition: Recurrent processes account for object recognition under occlusion. PLoS computational biology, 2019, vol. 15, no 5, p. e1007001.

[41] S. V. P. Peursum, H. Bui and G. West. Robust recognition and segmentation of human actions using hmms with missing observations. 13:2110–2126, 2005.

for consumer surveillance system. 55(2):591–598, May 2009. DOI https://doi.org/10.1155/ASP.2005.2110

[42]. Wei, S. E., Ramakrishna, V., Kanade, T., & Sheikh, Y. (2016). Convolutional pose machines. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (pp. 4724-4732).

[43] LIU, Jingwen; GU, Yanlei; KAMIJO, Shunsuke. Customer behavior classification using surveillance camera for marketing. Multimedia Tools and Applications, 2017, vol. 76, no 5, p. 6595-6622.

[44] El-Sayed, M. A., & Hamed, K. (2015). Study of Similarity Measures with Linear Discriminant Analysis for Face Recognition. Journal of Software Engineering and Applications, 08(09), 478–488. DOI: 10.4236/jsea.2015.89046

[45] Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. 2014 IEEE Conference on Computer Vision and Pattern Recognition. DOI: 10.1109/cvpr.2014.220

[46] Common Objects in Context. (n.d.). Retrieved from http://cocodataset.org.

[47] Cakir, F., He, K., Xia, X., Kulis, B., & Sclaroff, S. (2019). Deep metric learning to rank. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1861-1870).

[48] PURI, Divyansh. COCO Dataset Stuff Segmentation Challenge. En 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA). IEEE, 2019. p. 1-5.

[49] Hui, J. (2019, April 3). mAP (mean Average Precision) for Object Detection. Retrieved from https://medium.com/@jonathan_hui/map-mean-average-precision-for-object-detection-45c121a31173

[50]Y.Yang and D. Ramansan. Articulated human detection with flexible mixtures of parts. PAMI'13

[51] The PASCAL Visual Object Classes (VOC) Challenge Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J. and Zisserman, A. International Journal of Computer Vision, 88(2), 303-338, 2010

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