Al-Nahrain Journal for Engineering Sciences **NJES** 26(1)24-30, 2023
http://doi.org/10.29194/NJES.26010024

Real-Time Objects Detection, Tracking, and Counting Using Image Processing Techniques

Dr. Mohammed Hussein Ali

Authors affiliations:
Department of Electronic and Communication Engineering, College of Engineering, Al-Nahrain University, Baghdad-Iraq.
mohammed.hussein.1@nahrainuniv.edu.iq

Paper History:

Received: 16th Nov. 2022

Revised: 24th Dec. 2022

Accepted: 7th Feb. 2023

Abstract

As a result of the tremendous development taking place in modern systems and technologies in the field of electronic monitoring. Intelligent monitoring, decision making, and automated response systems have become common subjects at this time, especially after the development of machines responsible for these processes. Traffic surveillance is a trend goal nowadays using different techniques and equipment. In this article, real-time Object detection and tracking techniques were proposed for traffic surveillance using image processing techniques. A state was specifically examined for its ability to detect and count passing motorcycles on a highway in a specific area. The results showed good reliability, with a frame processing time of approximately about (30 ms) and the achievement of real-time performance. The main contribution of this article is reaching the best result implemented by the performance the real-time process using image process technique and tracking the object by depending on the sequencing of frames and can stands with rationally not so powerful machines. Several tools have been used for different types of necessary tasks that will be part of the required application such as Python 3.7; which was used to build the basic algorithms,Visual studio code (VSC) as an Integrated Development Environment (IDE), and Anaconda navigator for downloading many useful libraries. The specifications of the used device were Intel(R) Core (TM) i7- 10750H CPU @ 2.60GHz 2.59 GHz, RAM 16.0 GB, NVIDIA GeForce GTX 1650 GPU, 64-bit operating system, x64-based processor.

Keywords: Real Time, Object Detection, Tracking and Counting, Traffic Surveillance, Image Processing Techniques.

اكتشاف الاجسام، تعقبها، و عدها في الوقت الحقيقي باستخدام تقنيات معالجة الصور

د.محمد حسين علي

الخلاصة:

أصبحت المراقبة الذكية واتخاذ القرار وأنظمة الاستجابة الآلية موضوعات شائعة في الوقت الحاضر ، خاصة بعد نمو الأداء لمجموعة واسعة من الأنظمة ، وهو نتيجة لتطور الآلات المتخصصة في هذه العمليات. تعد مراقبة حركة المرور أحد الأهداف المهمة في الوقت الحاضر باستخدام تقنيات ومعدات مختلفة. في هذا البحث ، تم اقتراح تقنيات الكشف عن الاجسام وتتبعها في الوقت الحقيقي لمراقبة حركة المرور باستخدام تقنيات معالجة الصور. تم استخدام العديد من الأدوات لأنواع مختلفة من المهام الضرورية التي ستكون جزءًا من التطبيق المطلوب مثل Python 3.7 ؛ والتي تم استخدامها لبناء الخوارزميات الأساسية ، ورمز الاستوديو المرئي (VSC(كبيئة تطوير متكاملة (IDE(، ومتصفح Anaconda Navigator لتنزيل المكتبات المتوافقة المختلفة.كانت مواصفات الجهاز المستخدم هي: معالج انتل Core (TM) i7- 10750H CPU @ 2.60 GHz 2.59 GHz ، ذاكرة الوصول العشوائي ١٦,٠ جيجابايت ، نظام تشغيل، ٦٤ بت مستند إلى معالج x64، و NVIDIA GeForce GTX 1650 GPU.

NJES is an **open access Journal** with **ISSN 2521-9154** and **eISSN 2521-9162**

1. Introduction

1.1 Object Detection

The capability of distinguishing objects by vision is naturally congenital owned by humans, but for machines, it is an issue [1,2]. Object detection is a computer vision technology that is strongly related to image processing that concerns with detecting specifically required objects (such as pedestrians, cars, etc.) in a video streaming or images [1,3,4]. Detection of objects is one of the most important challenges in applications of computer vision and even more important than pattern detection or recognition, has received great attention in recent years [5].

The continuously developing of software techniques for object detection, enhanced the performance and made it reach its peak ability using the routine programming procedures, so the need for developing the ways of designing the frameworks has been grown day after day till the time of using Artificial Intelligence (AI). The function of object detection is developing computational models and techniques that provide one of the most basic pieces of information needed by computer vision applications [2].

The researches conducted on this topic can be classified into two main areas: "General Object Detection" and "Detection Applications", where the first area focuses on finding ways to explore different objects within a common framework to simulate human vision and perception, while the second focuses on exploration in different ways, in a specific application, such as pedestrian detection, face detection: text detection, etc. [1, 2, 6, 7]. Object detection is used in a wide range of real-world applications, such as video surveillance, robotics, autonomous driving, robot vision, and other similar applications [1]. Object detection is increasingly important in practical applications that require real-time execution, **Fig. 1** shows the exponential growth in the number of research papers over the past two decades in this field. [2].

The tracking of objects in videos in real-time is crucial for both video surveillance and autonomous driving. Conventional object detectors typically focus on detecting objects in individual images, neglecting the relationships between frames in a video. Research on improving detection by taking into account the spatial and temporal connections between frames is a significant area of study [2, 7].

Many detectors employ significant algorithms and techniques, such as color thresholding (which is a technique used to segment an image based on color. It involves converting an image from the RGB color space to a binary image, where all pixels that fall within a specified range of colors are set to 1, and all other pixels are set to 0) and contour detection (which is the process of identifying the boundaries of objects in an image. The edges of an object are traced along its perimeter, creating a continuous line or contour that defines the object's shape.). In contrast, object tracking employs processes such as comparing sequential frames [2, 3].

In this work colors thresholding will help for easy identification and separation of objects or regions of interest in an image based on their color, while contours detection will be used to detect the edges of the object which can be used to detect the contours of the object.

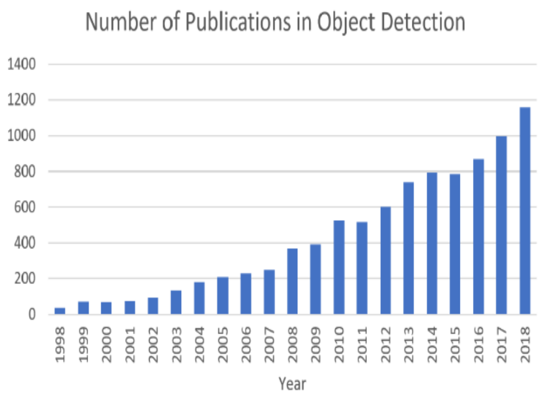

Figure (1): Publication numbers growth of "object detection" [3].

1.2 Literature Survey

Many attempts were done with different properties and performances, here are some examples:

1.2.1 Viola Jones Detectors

P. Viola and M. Jones achieved for the first time without any constraints real-time human faces detection. Due to the importance of this achievement, this algorithm was named after the authors in appreciation of the important contribution [2]. Although the algorithm has a straight forward path to follow and it seems to be very simple, but it has very complicated calculations which is far from the power of the computers in its time. The algorithm adopts sliding windows in its working method, to detect the human face in the image from the windows of variable size and position. [2, 8].

1.2.2 Histogram of Oriented Gradients (HOG) Detector.

This algorithm was first proposed by N. Dalal and B. Triggs in 2005 [9]. It can be considered as an improvement of the scale-invariant feature transform [10, 11] and shape contexts [12] of its time. The HOG descriptor is mainly designed to be computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization (on "blocks") to improve the accuracy. The main motive was to solve the problem of pedestrian detection, however, it can be used to detect a variety of object classes. The HOG detector rescales the input image multiple times with remaining the detection window size unchanged for detecting objects of different sizes. This detector formed the basic structure for different object detectors and different applications of computer vision for a long time [13, 14]

1.2.3 Deformable Part-Based Model (DPM)

It was proposed by P. Felzenszwalb [13] in 2008, and is considered to be the best conventional object detector. The DPM is the expansion of the HOG detector, after that many improvements have been updated by R. Girshick [14]. The philosophical type of the model DPM is "divide and conquer", in which the training can be defined as a proper way learning for decomposing an object. For instance, detecting a "car" can be done by detecting its window, body, and wheels [13].

2. Applications

Object detection with its basic concepts can be imagined that it has a lot of applications, and after the performance evolution and simplicity in achieving different systems, the applications became more attractive in various fields.
Here are some of the mentioned applications to get acquainted with their capabilities and possible features:

2.1 Optical Character Recognition:

It can be said it is a tool more than a simple application, it can be cleared as the process of converting the printed text or even handwritten into computer text. This application can be a tool for a wide range of larger applications with various type of inputs like vehicle license plate or paper documents into a database [15, 16].

2.2 Self Driving Cars:

One of the main requirement for self-driving cars is the object detection. Which will be the tool that helps with the decision made like accelerate, brake and turning into specific direction [17, 18, 19].

2.3 Face Detection:

Face detection is a widely used computer vision tool. It can be noticed the manner that the facebook using it with detecting our face after uploading a photo. Face detection can be regarded as a specific case of object-class detection. This is a simple application of object detection that we see in our daily life [20, 21, 22].

2.4 Activity Recognition:

This application, in general, is very useful and especially with long-time or full-time surveillance systems. It can produce a notification if any notable change happens under different conditions and circumstances. It can be applied with rationally humble device specifications [23, 24, 25, 26].

2.5 Pedestrian Detection:

Detecting and locating people is a very important tool that can be used with different applications like self-driving cars, customers counting breaching into forbidden areas [27, 28].

2.6 Ball Tracking in Sports:

After the increasing number of sports lovers like baseball, football, etc. it was essential to present and analyze more details concerned with the game with multidimensional information visualization. As these games depend mainly on the ball position accurately, object tracking is an important tool for tracking the ball and extracting information about it [29].

2.7 Object Counting:

This tool can be used with a lot of applications concerned with statistics calculations and automated systems like traffic surveillance (as it will be seen later), pedestrian counting, and automated manufacturing processes [30, 31].

2.8 Automatic Image Annotation:

Also, it can be called automatic image tagging which is automatically assigning processes by the machine in the form of captioning or keywords to a digital image. This process can be used with various types of applications like automatically tagging people on social media [32, 33, 34].

The reason for developing a traffic surveillance system is the increasing number of vehicles on the roads. As the number of vehicles continues to grow, traffic problems become more common and can have serious consequences such as delays, accidents, and pollution. A traffic surveillance system can help to monitor and manage these problems by providing real-time information on traffic flow and identifying areas where congestion is likely to occur. This information can be used by planning engineers to design and implement solutions to improve traffic flow and safety. Additionally, a traffic surveillance system can be used for a variety of applications, depending on the needs of the situation, making it a versatile tool for addressing traffic challenges.

3. Methodology

In order to have a full view of how this system works, the procedure will be divided into small parts to be easy for understanding.
Some points need to be clear before going deep in the process like dealing with the incoming video stream is actually done by processing it frame by frame with some interactions between them (for tracking purposes), the detected area of each object must be calculated according to the specific situation that the camera is mounted, etc. The video available at [35].

The **first** part is concerned with receiving the video from the camera and converting it into a stream of frames, as shown in **Fig. 2**.

Figure (2): Converting the video into a stream of frames

Third, this allocated area will be processed by passing it through a specified threshold level (In this work, the value of the threshold was chosen using the try-and-error method and it can be adjusted according to the required task in a specific situation) in order to convert it to a black and white image and prepare it to the next process which is the extraction of the contours.

After that, contours in the result are allocated and their areas are calculated for deciding whether take them into consideration or not by depending on the comparison with the area threshold. If the calculated area was in the acceptable range (which depends on the specific examined situation), then this object is added to the detection list in that frame, and it will be tracked with the next frame until it will be gone outside the region of interest. In this work the range of the acceptable area is limited between 700 and 1600 squared unit area.

Then, the parameters represented by the position, width, and height are passed into the tracking algorithm in order to track the object in the next frame and not assume is as a new object if it was not. The tracking process will count from 1 for the first detected object, then will decide for each detected object if it was new the count will be increased by one, and if it was the same in the previous frame then the count remains as the object is not new.

Lastly, this output can be used by different processes that require the number of passed vehicle through a specific area.

The whole algorithm is shown in the following Process flowchart**.**

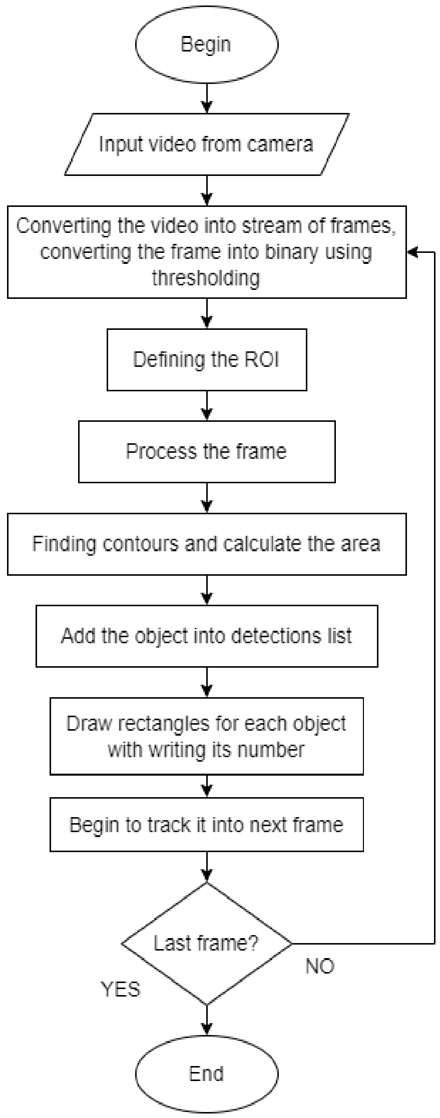

Figure (3): Process flowchart

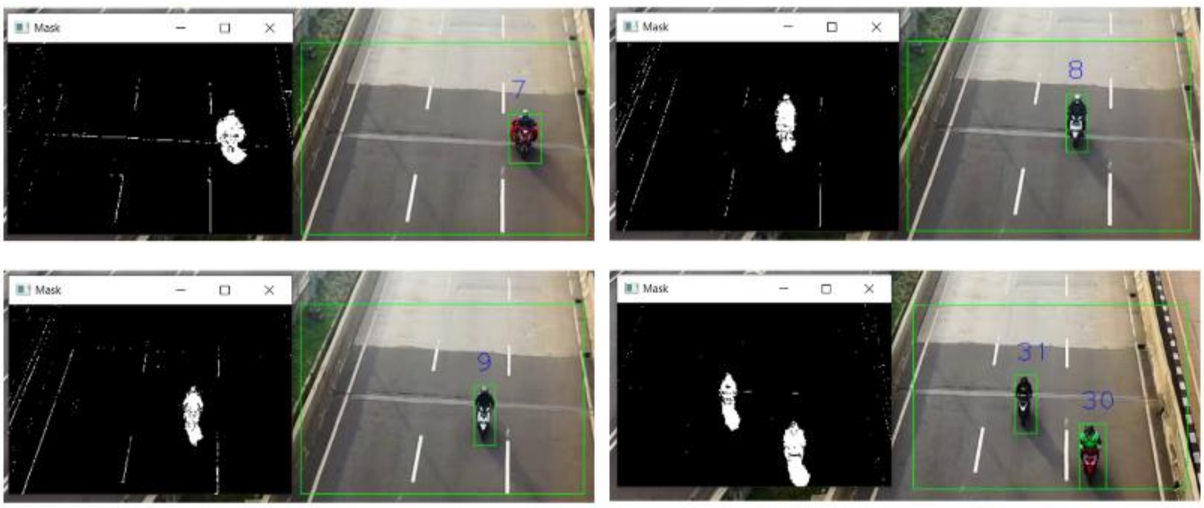

Figure (4): The result of the process

4. Experiments and Results

The process was applied for calculating the number of motorcycles as an example for measuring the performance of the process. As a specific path was the target to count the number of motorcycles occupying it, the specific area was allocated of this path with taking in count avoiding interfering with other moving parts. Then the process began and the result was as shown in **Fig. 4**.

The time consumed for processing each frame ranges between (29.025ms) and (31.502ms), while the frame per second (FPS) achieved a maximum number of (34) frames per second, which can be used for real-time applications as shown in **Fig. 5**.

```
time= 0.03150200843811035        FPS= 31.74400775000189

time= 0.029025793075561523       FPS= 34.45211634344486
```

Figure (5): Processing time and frame per second rates

5. Conclusion

Many important points can be concluded from this article and its results such as:

1. Various object detection methods were proposed with different algorithms, advantages, disadvantages, usages, and applications that can be applied with them.
2. Great evidence about not reaching the final limits of the performance that can be done using the computers, is the continuously developing of the methods and algorithms.
3. As traffic surveillance systems are essential parts of the controlling systems that can grant great benefits by applying these algorithms for it, object detection was used for surveillance in this article.
4. The results were obviously good and acceptable with different conditions that can be applied.
5. One of the main things that must be mentioned, is the system was real-time and without any lagging.
6. Another important point must be cleared is that this work did not use any AI or deep learning algorithms but mainly depended on the image processing filters.

This provided many advantages for this work like real-time processing, low CPU and GPU specifications requirements, and simplicity in implementation and development.

6. Future Work

A lot of ideas can be implemented in this area of researches and applications, like:

1. This field represented by object detection with traffic surveillance can be developed using a lot of other algorithms with taking into count the advantages that can benefit from, and the disadvantages that can avoid.
2. The process can be applied all over the roads and highways in order to have a full view of the traffic in the whole city, and this can be considered a type of smart city application.
3. Also, it can be used in small manners like automated monitoring of the garages of different institutes of various sizes.
4. The resulted data can be used after arranging them in tables in databases and used for statistical calculations for future road development by finding the essential roads that must be done in the future for solving specific traffic issues.
5. For developing the system for more accuracy and better performance with taking into count the additional hardware requirements, deep learning can be used in various methods like automatically specifying the area of the required vehicles, or even detecting specific vehicle shapes, enhancing the tracking of each vehicle, and adapt the size of the area according to the different conditions that it may face in the practical environment.

6. References

[1] Abdul Vahab, et al. "Applications of object detection system," *International Research Journal of Engineering and Technology (IRJET),* vol. 6, issue 4, pp. 4186-4192, April 2019.
https://www.irjet.net/archives/V6/i4/IRJET-V6I4920.pdf

[2] Z. Zou, Z. Shi, Y. Guo and J. Ye. (2019, May). *Object detection in 20 years: A survey.* in *arXiv:1905.05055v2* [online].
https://arxiv.org/abs/1905.05055

[3] Benenson, R., Omran, M., Hosang, J., Schiele, B., "Ten years of pedestrian detection, what have we learned?," *In: Computer Vision - ECCV 2014 Workshops*, L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part II, LNCS 8926, 2015, pp. 613–627, Springer International Publishing (2015).
https://doi.org/10.1007/978-3-319-16181-5_47

[4] Dollár, P., Wojek, C., Schiele, B., Perona, P., "Pedestrian Detection: An Evaluation of the State of the Art," *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743-761, April 2012.
https://doi.org/10.1109/TPAMI.2011.155

[5] S. Yanushkevich, P. Wang, M. Gavrilova, M. Nixon and S. Srihari, *Image Pattern Recognition: Synthesis and Analysis in Biometrics,* Singapore: World Scientific, 2007.
https://www.nlb.gov.sg/biblio/12941648

[6] Singla, N., "Motion detection based on frame difference method," *International Journal of Information Computation Technology*. Vol. 4, issue 15, pp. 1559–1565, 2014.
https://www.ripublication.com/irph/ijict_spl/ijictv4n15spl_10.pdf

[7] Nayagam, M. Gomathy, and K. Ramar. "A survey on real time object detection and tracking

algorithms," *International Journal of Applied Engineering Research*, vol.10, number 9, pp 8290-8297, 2015.
https://www.researchgate.net/publication/274569172_A_survey_on_Real_time_Object_Detection_and_Tracking_Algorithms

[8] P. Tavallali, M. Yazdi and M. R. Khosravi, "An Efficient Training Procedure for Viola-Jones Face Detector," *2017 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2017, pp. 828-831, https://doi.org/10.1109/CSCI.2017.143

[9] 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, vol. 1, pp. 886-893. https://doi.org/10.1109/CVPR.2005.177

[10] D. G. Lowe, "Object recognition from local scale-invariant features," *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Kerkyra, Greece, 1999, vol. 2, pp. 1150-1157. https://doi.org/10.1109/ICCV.1999.790410

[11] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004. https://www.cs.ubc.ca/~lowe/papers/ijcv04.pdf

[12] S. Belongie, J. Malik and J. Puzicha, "Shape matching and object recognition using shape contexts," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509-522, April 2002. https://doi.org/10.1109/34.993558

[13] P. Felzenszwalb, D. McAllester and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," *2008 IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, USA, 2008, pp. 1-8. https://doi.org/10.1109/CVPR.2008.4587597

[14] P. F. Felzenszwalb, R. B. Girshick and D. McAllester, "Cascade object detection with deformable part models," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, 2010, pp. 2241-2248. https://doi.org/10.1109/CVPR.2010.5539906

[15] Eikvil, Line. "Optical character recognition," *citeseer. ist. psu. edu/142042. html* 26 1993. https://doi.org/10.3403/00116086U

[16] Ravina Mithe, Supriya Indalkar, and Nilam Divekar. "Optical character recognition." *International journal of recent technology and engineering (IJRTE)*, vol 2, issue 1, pp. 72-75, March 2013. https://www.ijrte.org/wp-content/uploads/papers/v2i1/A0504032113.pdf

[17] Simhambhatla, Ramesh; Okiah, Kevin; Kuchkula, Shravan; and Slater, Robert (2019) "Self-Driving Cars: Evaluation of Deep Learning Techniques for Object Detection in Different Driving Conditions," *SMU Data Science Review*: Vol. 2: No. 1, Article 23. https://scholar.smu.edu/datasciencereview/vol2/iss1/23

[18] Abhishek Gupta, Alagan Anpalagan, Ling Guan, Ahmed Shaharyar Khwaja, "Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues," *Array*, vol. 10, February 2021. https://doi.org/10.1016/j.array.2021.100057.

[19] S. H. Naghavi, C. Avaznia and H. Talebi, "Integrated real-time object detection for self-driving vehicles," *2017 10th Iranian Conference on Machine Vision and Image Processing (MVIP)*, 2017, pp. 154-158. https://doi.org/10.1109/IranianMVIP.2017.8342340

[20] H. Jiang and E. Learned-Miller, "Face Detection with the Faster R-CNN," *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pp. 650-657. https://doi.org/10.1109/FG.2017.82

[21] B. Froba and A. Ernst, "Face detection with the modified census transform," *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings*, pp. 91-96. https://doi.org/10.1109/AFGR.2004.1301514

[22] C. P. Papageorgiou, M. Oren and T. Poggio, "A general framework for object detection," *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, 1998, pp. 555-562. https://doi.org/10.1109/ICCV.1998.710772

[23] Tripathi, R.K., Jalal, A.S. & Agrawal, S.C. "Suspicious human activity recognition: a review," *Artif Intell Rev* 50, pp. 283–339, 2018. https://doi.org/10.1007/s10462-017-9545-7

[24] S. W. Pienaar and R. Malekian, "Human Activity Recognition using Visual Object Detection," *2019 IEEE 2nd Wireless Africa Conference (WAC)*, 2019, pp. 1-5. https://doi.org/10.1109/AFRICA.2019.8843417

[25] Xu, Xin, Jinshan Tang, Xiaolong Zhang, Xiaoming Liu, Hong Zhang, and Yimin Qiu. "Exploring Techniques for Vision Based Human Activity Recognition: Methods, Systems, and Evaluation," *Sensors* 13, no. 2: pp. 1635-165, 2013. https://doi.org/10.3390/s130201635

[26] J.K. Aggarwal, Lu Xia, "Human activity recognition from 3D data: A review," *Pattern Recognition Letters*, vol.48, pp.70-80, 2014. https://doi.org/10.1016/j.patrec.2014.04.011.

[27] C. Papageorgiou and T. Poggio, "Trainable pedestrian detection," *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)*, 1999, vol. 4, pp. 35-39. https://doi.org/10.1109/ICIP.1999.819462

[28] S. Walk, N. Majer, K. Schindler and B. Schiele, "New features and insights for pedestrian detection," *2010 IEEE Computer Society Conference on*

Computer Vision and Pattern Recognition, 2010, pp. 1030-1037.
https://doi.org/10.1109/CVPR.2010.5540102

[29] Kamble, P.R., Keskar, A.G. & Bhurchandi, K.M. "Ball tracking in sports: a survey," *Artif Intell Rev 52,* pp. 1655–1705, 2019.
https://doi.org/10.1007/s10462-017-9582-2

[30] Arteta, C., Lempitsky, V., Noble, J.A., Zisserman, A. (2014). "Interactive Object Counting," in *Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science,* Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. *(*eds*)*. vol 8691. Springer, Cham.
https://doi.org/10.1007/978-3-319-10578-9_33.

[31] C. Pornpanomchai, F. Stheitsthienchai and S. Rattanachuen, "Object Detection and Counting System," *2008 Congress on Image and Signal Processing*, Sanya, China, 2008, pp. 61-65.
https://doi.org/10.1109/CISP.2008.108

[32] A. Hanbury, "A survey of methods for image annotation," *Journal of Visual Languages & Computing,* vol. 19, No. 5, pp.617-627, 2008.
https://doi.org/10.1016/j.jvlc.2008.01.002

[33] Dengsheng Zhang, Md. Monirul Islam, Guojun Lu, "A review on automatic image annotation techniques," *Pattern Recognition,* vol. 45, issue 1, pp. 346-362, January 2012.
https://doi.org/10.1016/j.patcog.2011.05.013

[34] Jeon, J., Manmatha, R. (2004). "Using Maximum Entropy for Automatic Image Annotation," in *Image and Video Retrieval. CIVR 2004. Lecture Notes in Computer Science,* Enser, P., Kompatsiaris, Y., O'Connor, N.E., Smeaton, A.F., Smeulders, A.W.M., EDS. vol 3115. Springer, Berlin, Heidelberg.
https://doi.org/10.1007/978-3-540-27814-6_7

[35] Tom Fisk, (2020, August, 16th). *Drone Footage Of Expressway During Daytime* [online].
https://www.pexels.com/video/drone-footage-of-expressway-during-daytime-4673661