Daffodil International University

Data mining (CSE 450)

Supervisor:

Mayen Uddin Mojumder

Group Number: 4

Group Name: HELL PRIESTS

Group Members:

Members Name	Id
Ahmed Nur-A-Jalal	171-15-1324
Jahidur Rahman Fahim	171-15-1414
Fatema Tuz Johra Rupa	171-15-1392
Nusrat Jahan	171-15-1405
Md Hamidul Hasan	171-15-1394

Human Detection and Counting Using Yolov3 Model

ABSTRACT:

Real-time tracking is a major, challenging, and important field of research. Various human monitoring systems, computer communication (HCI), video surveillance etc. are being implemented on human detection in recent times. Research into human biometric authentication has come a long way but real-time human tracking has not been of much importance. Personal tracking can be used as a preliminary step in biometric face recognition. Keeping a continuous track of a person will allow you to identify the person at any time. The program consists of three parts first person detection, second tracking and three counting the amount of person present in the frame. The model is based on YOLO v3 and the system is trained with real time collected data and manually annotating them.

This monitoring system with real-time video surveillance could be deployed for places with a huge human gathering.

The face is the most important part of a person representing the most important details about that person. The eye is an important biometric element used in human identification. Facial detection is performed using skin-based methods. The YC b C r color model is used to find skin regions as it represents the intensity and color information differently. For eye-guessing function and pixel calculation methods are used. Finally the mean shift algorithm is used for tracking.

This model was able to detect and recognize multiple people or other regular person activity and multiple human activities tracing support at once. After completing our project, this model manages an average accuracy of 94.6667% while recognizing image and from video file.

INTRODUCTION:

Human detection plays an important role in this modern technology era. It's a large field for research. Nowadays it has become a rising topic in the human interaction area. In past decades many researchers are working on it. Computers don't have its own brain to detect anything. They can't read the humans mind which is very important. If the computer can understand the activity of a humans, it can bring a lot of positive changes in the field of IoT. Nowadays Human detection is creating a big chaos in the technology field. As many researcher and scientists are working on it from past decades.

As we know that today's world is directly or indirectly depends on computer. Though it's very challenging task, but for making our life easier, researchers are constantly working on it. Using artificial intelligence and IoT, Human detection is developing day by day. In this sector computer monitors the actions of humans walking, seating, running, sleeping, and speaking and other body part movement. Collecting those data's through many sensors, computer starts works on it. It can sense any ages of person whether he/she is child or older. It can recognize the unusual activities of any person. Following those data, it can help any person who needs help; it can detect any person doing unethical things. It creates an environment where they use data and compare them with current situations and come to conclusion about the activity of a person. It uses visual and non-visual data to analyze the data. It provides the identity or mental states of a person which can help to analyze peoples.

But there is no that level of satisfaction output found till now. If we can successfully develop it, it can bring a big revolution. From this paper, In the end, we'll also highlight some survey of our work according to this paper. This Human detection technology can help many sectors like medical, mental health, automation sector, detecting criminal, education. It is also beneficial for ensuring privacy.

Advantage of detecting human:

By using this Human detection technology, we can get a lot of benefits which can make our life easier. It can bring a big change in medical science technology. Through sensors it can sense the activity of a person and work on the current situation of their health. By using this technology, we can reduce the crime of an area. The unusual movement of an area can give information to the control room through this technology. Police can also trace the location of a criminal through the detection of the criminal. It can also help the elder persons and children who needs help anytime. It will make the security section stronger as people will be caught red handed doing any wrong activities. As it is being done with the help of computer, the chances of any mistakes are quite low.

For smart home it is being used widely for the security purpose. For any particular areas, it can make a big change as this technology can control unethical issued by recognizing the activities of any person. As it works automatically, no need to work on it again and again. Mainly this technology is focusing on security assurance. Like in this corona situation less people are working directly for security purpose. So, using it we can assure the high security for any particular area. It can help use solving many problems.

As it can detect the activity of people, it can control the person limit in any particular area. This technology can count people from any area. So, it'll help to control the gather of any place like marriage ceremony, birthday celebration etc. where a limited person is allowed. In future, it'll work for large area for a large amount of people recognition. This will help to detect the unauthorized people who are not allowed to enter any area. This technology will alert the owner of that something fishy is happening there.

As it can detect the human, it can take decision whether it is seeing any other object or human. So, using this technology can give the correct decision. So, we can say with a high hope that this Human detection technology can bring a great revolution in case of security.

DESCRIPTION OF YOLO:

The full form of YOLO is You Only Look Once. It detects objects in different ways. YOLO didn't focus on the whole picture. It focuses on the main portion of a picture. YOLO were developed detection and classification. Surrounding box and class portend are made after one evaluation of the input image. YOLO is the best algorithm to detect object it compared with R-CNN algorithm. R-CNN algorithm also detects an object. Both are the same category to detect objects but YOLO uses a totally different approach. The old method R-CNN algorithm focuses on several portions of an image and very difficult to detect an object. These algorithms focus on the full image, and then divide the image into some portions and portend surrounding boxes and probabilities for each portion. YOLO is clever for doing object detection in real-time. YOLO algorithm is best because it is working on real-time data. If we would know the YOLO algorithm first of all we have to know what actually predicts it. To predict an object it divides an object into 2 bounding boxes. YOLO algorithms divide an image into S*S grid. When the center of objects falls into the grid then the algorithm detects the object. To portend an object each surrounding box can be described using four descriptors:

a. Center of a surrounding box.
b. Width.
c. Height.
d. Value is corresponding to a class of an object.

ADVANTAGES OF YOLO V3:

The main advantages of YOLO are it is superfast. Its speed has 45 per second which is too much better than real-time. Its accuracy is better than R-CNN. This algorithm is the best algorithm to detect objects. YOLO can Detect 49 objects at a time. Its limitation to detect an object is 7*7 grids.

YOLO algorithms increase the number of boxes and decrease the orbit for IoU for better results. It is faster to detect objects of another technique. If we use resize Image it gives us better results. YOLO algorithm Understand generalized object representation. YOLO cannot detect small objects because it focuses on-grid and portends one object.

MOTIVATION

The objective of human detection is to inspect people or human from video successions or still pictures. Large gathering, it is highly crucial for any authority or security team to continuously monitor the human activity. That's why the human monitoring system in the field of human activity detection has always been occupying a very principal position in the field of security system. Not only human detection, but this method can come in handy for many other purposes. The continuous improvement of artificial intelligence and deep learning algorithm not only helping us to transmit and get vital physiological signs to the human but also simplifies the quantification and as a result, rises the efficiency of the monitoring system. The active or smart system can use HAR technology to monitor its residential area for better security purposes. The aim of our research can also be to offer medical support, well-being services and health benefit to older adults and other security purposes for important infrastructure.

As we were studying Computer Science and Engineering in our final year, we already had some skills and knowledge to build a solution that can help us improve from a predicament. We decided to solve some of our social problems and we found out how to. We gather a group of students with the same peaceful purposes, "to monitor human movement" which can be revolutionary in the medical and other sector and the same project can also be used for security also after some modification.

After studying about the topic, a large amount of work has already been done in the video-based dataset, but a large amount of image that we have mentioned in the introduction has not been used rather used in the significantly lower number of datasets. This fact inspired us the most to work with a large number of images. This was very interesting for us because we were about to create an intelligent system that will detect human and monitor that activity intelligently. That's why we decided to took the challenge.

METHODOLOGY:

The 'You Only Look Once' v3 (YOLOv3) method is among the most widely used methods of obtaining learning-based material. We have divided our whole work into three major steps.

1. Data Collection and Analysis:
   - Data smoothing
   - Data labeling
2. Data Training
3. Result Analysis

Data Collection and Analysis:

At first we use our custom data sets and manually annotate them. We collected 1000 real life images and 150 test images.

Class	Data
1. person	1000

Then we annotated or labelled persons (humans) in them and we uploaded the dataset and into Google Colab using python.

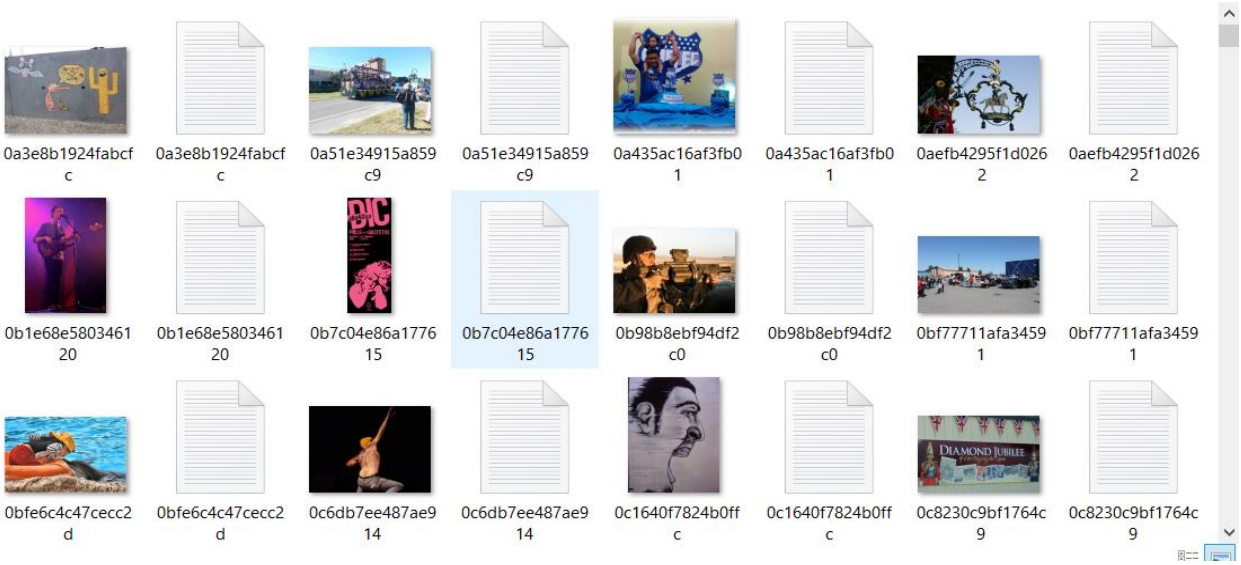


Fig: sample labelled data for train

Data Train:

Then we trained the custom object/ human/person “dataset” in the Colab using python, Darknet, pre trained weights from yolov3, OpenCv, NumPy, It uses the k-means collection method to measure the initial width and height of the predicted binding boxes. In this way, the estimated width and length are sensitive to the first sets of collections, and the processing of large databases takes time. To address these issues, a new collection method for measuring the initial width and height of the predicted binding boxes has been developed.

First, it randomly selects the values of width and height as the center of a collection that differs in the width and height of true earth boxes. Second, it builds Markov's chains according to the first selected collection and uses the final points of the entire Markov series like other original centers.

In the construction of Markov chains, the intersection-over-union method is used to calculate the distance between selected groups and each entry point, instead of the square root method. Finally, this method can be used to continuously update the collection center with new collections of width and height, which is part of the data selected from the database only. And to count persons/human or anything we let the index value of person/human is 0 and then run a loop in the frame for the “numbers of detected object as human”. The person index is 0 so we need to check if the predicted class is zero and increase the count accordingly.

Then we collected the weight as our trained model weights for testing.

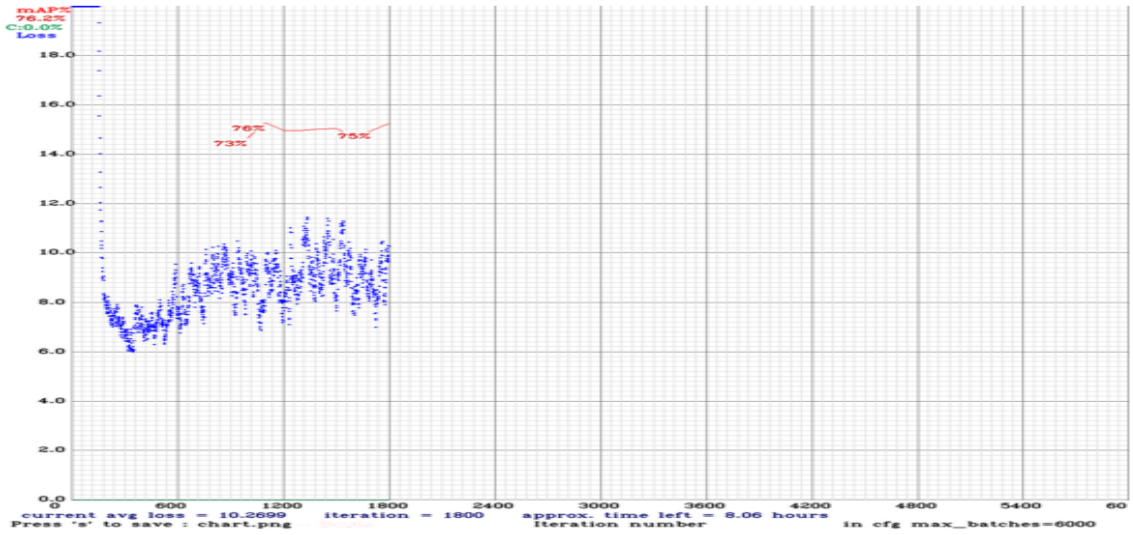


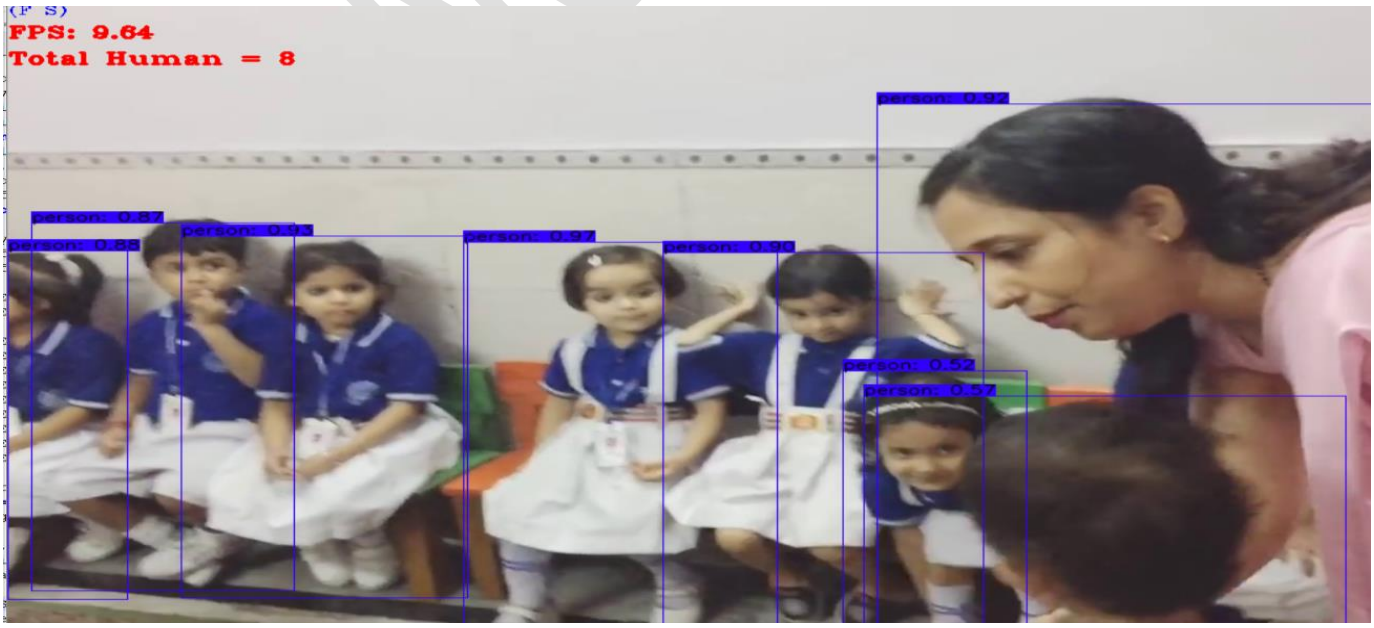
Fig: chart of loss reduction after 1800 iteration’s with 95% accuracy and loss less than 5.

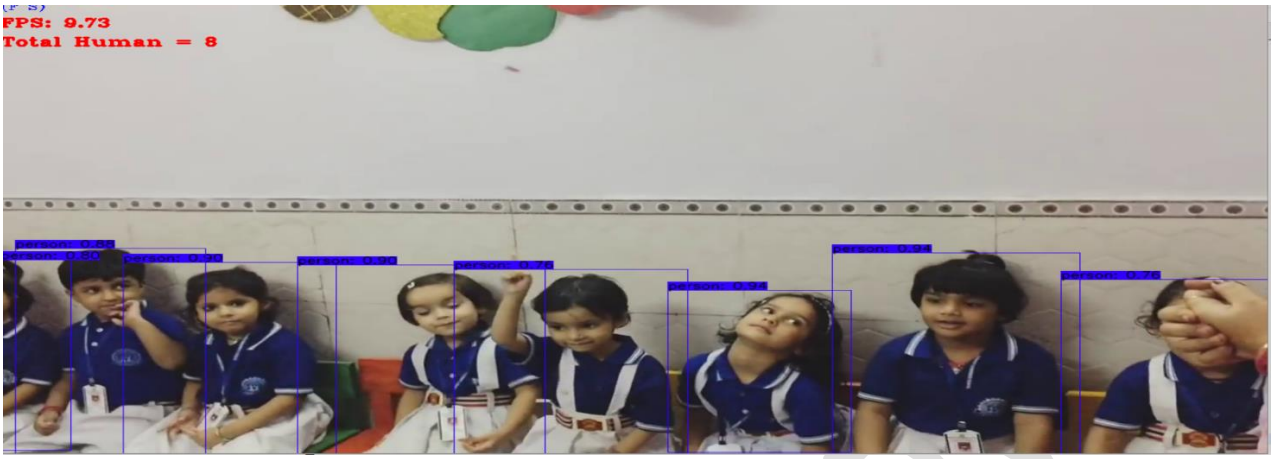
```
 calculation mAP (mean average precision)...
 Detection layer: 139 - type = 28
 Detection layer: 150 - type = 28
 Detection layer: 161 - type = 28
156

 for conf_thresh = 0.25, precision = 0.32, recall = 0.90, F1-score = 0.47
 for conf_thresh = 0.25, TP = 141, FP = 299, FN = 15, average IoU = 20.94 %

 IoU threshold = 50 %, used Area-Under-Curve for each unique Recall
 mean average precision (mAP@0.50) = 0.234849, or 23.48 %
Total Detection Time: 16 Seconds

Set -points flag:
 `-points 101` for MS COCO
 `-points 11` for PascalVOC 2007 (uncomment `difficult` in voc.data)
 `-points 0` (AUC) for ImageNet, PascalVOC 2010-2012, your custom dataset
```

Fig: calculation map: (mean average precision)

Output Result Calculation:

Now we can test, detect and count persons only in local videos and not from webcams. We have to use this model on our local webcams by implementing this model into our local computer using python language.





Our simulation results show that the new method has a faster mixing speed to start the width and height of the bound box boxes and that it can select the length of the long representation and the height of the bound boxes. Our proposed method achieves better performance than the YOLOv3 method in terms of memory, mean standard accuracy, and F1 score. .

After completing our project, this model manages an average accuracy of 94.6667% while recognizing image and from video file. We found the accuracy by calculating the average of 100 sample output manually.

CONCLUSION

The techniques approved with our difficult dataset where many jumble and uproarious information for checking more exactness are. It can recognize more than one individual's human utilizing additional jumping encloses a solitary picture. Different detection methods and a few research points that are connected to activity investigation in still pictures have been talked about in our paper. In future we are planning to add more features in this project that would make this more usable and would revolutionize human activity monitoring system. Such as human activity recognition and prediction of future activity based on present activity can be applied in this project to make it more usable as it can be used for security purposes.

RECOMMENDATION

It will be smarter to development the measure of preparing information. The dataset needs information from more various points and different wellspring of lights. Attempt other diverse profound learning calculations or deep learning algorithm to look at which is better for "Human Act". This model can be additionally utilized for different activity discovery issues like home, industry activity location issue, any sensitive area. A huge dataset request to work to perform different activity recognition task.

Reference:

[1] Ahmed Taha, Hala Zayed and El-Sayed M. El-Horbarty “Human Activity Recognition for Surveillance Applications” May 2015

[2] Chen Change Loy, "Activity Understanding and Unusual Event Detection in Surveillance Videos," PhD dissertation, Queen Mary University of London, 2010.

[3] Mao Ye, Qing Zhang, Liang Wang, Jiejie Zhu, Ruigang Yang, Juergen Gall, "A Survey on Human Motion Analysis from Depth Data," Lecture Notes in Computer Science, Springer Berlin Heidelberg, Volume 8200, pp 149-187, 2013.

[4] Lulu Chen, Hong Wei, James Ferryman, "A survey of human motion analysis using depth imagery," In Pattern Recognition Letters, Elsevier Science Inc., Volume 34, Issue 15, pp. 1995-2006, November 2013

[5] Ronald Poppe, "A survey on vision-based human action recognition," In the International Journal of Image and Vision Computing, Volume 28, Number 6, pp.976-990, June 2010

[6] Maaike Johanna, "Recognizing activities with the Kinect," Master thesis, Radboud University Nijmegen, Nijmegen, Netherlands, July 2013.

[7] Ferda Ofli, Rizwan Chaudhry, Gregorij Kurillo, René Vidal, and Ruzena Bajcsy, "Sequence of the Most Informative Joints (SMIJ): A New Representation for Human Skeletal Action Recognition," In proceedings of the IEEE Computer Vision and Pattern Recognition Workshops (CVPRW), Providence, Rhode Island, USA, PP. 8-13, June 2012.

[8] Ahmed Taha, Hala H. Zayed, M. E. Khalifa and El-Sayed M. El-Horbaty, " Human Action Recognition based on MSVM and Depth Images," In The International Journal of Computer Science Issues (IJCSI), Volume 11, Issue 4, Number 2, pp. 42-51, July 2014.

[9] Ahmed Taha, Hala H. Zayed, M. E. Khalifa and El-Sayed M. El-Horbaty, "Exploring Behavior Analysis in Video Surveillance Applications," In The International Journal of Computer Applications (IJCA), Foundation of Computer Science, New York, USA, Volume 93, Number 14, pp. 22-32. May 2014.

[10] Raviteja Vemulapalli, Felipe Arrate and Rama Chellappa, "Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group," In Proceedings of the International IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, Ohio, USA, pp.588-595, June 2014.

[11] Ahmad Jalal, Shaharyar Kamal and Daijin Kim, "A Depth Video Sensor-Based Life-Logging Human Activity Recognition System for Elderly Care in Smart Indoor

Environments," In the International Journal of Sensors, Volume 14, Number 7, pp. 11735-11759, July 2014.
[12] Samuele Gasparrini, Enea Cippitelli, Susanna Spinsante and Ennio Gambi, "A Depth-Based Fall Detection System Using a Kinect Sensor," In the International Journal of Sensors, Volume 14, Issue 2, pp. 2756-2775, February 2014.

[13] Salah Althloothia, Mohammad H. Mahoora, Xiao Zhanga, Richard M. Voylesb, "Human Activity Recognition Using Multi-Features and Multiple Kernel Learning," In Pattern Recognition Journal, Volume 47, Issue 5, pp. 1800–1812, May 2014.

[14] Jiang Wang, Zicheng Liu, Ying Wu, Junsong Yuan, "Mining Actionlet Ensemble for Action Recognition with Depth Cameras," In Proceedings of the International IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, Rhode Island, USA, pp. 1290-1297, June 2012.

[15] Xiaoxiao Dai, "Vision-based 3D Human Motion Analysis for Fall Detection and Bed-exiting," Master thesis, Faculty of the Daniel Felix Ritchie School of Engineering and Computer Science, University of Denver, USA, August 2013.

[16] Manjuatha M B, Pradeep kumar B.P., Santhosh.S.Y, "Survey on Skeleton Gesture Recognition Provided by Kinect," In the International Journal of Advanced Research in Electrical Electronics and Instrumentation Engineering (IJAREEIE), Volume 3, Issue 4, April 2014.

[17] Jungong Han, Ling Shao, Dong Xu, and Jamie Shotton, "Enhanced Computer Vision with Microsoft Kinect Sensor: A Review," In IEEE Transactions on Cybernetics, Volume 43, Number 5, pp. 1318 - 1334, October 2013.

[18] Vennila Megavannan, Bhuvnesh Agarwal, and R. Venkatesh Babu, "Human Action Recognition using Depth Maps," In proceedings of the International Conference on Signal Processing and Communications (SPCOM), Bangalore, India, pp. 1-5, July 2012.

[19] Alexandros Andre Chaaraouia, José Ramón Padilla-Lópeza, Pau Climent-Pérezb, and Francisco Flórez-Revuelta, "Evolutionary Joint Selection to Improve Human Action Recognition with RGB-D Devices," In the International Journal of Expert Systems with Applications, Volume 41, Issue 3, pp. 786–794, February 2014.

[20] Xiaodong Yang, Chenyang Zhang, and YingLi Tian, "Recognizing Actions Using Depth Motion Maps-Based Histograms of Oriented Gradients," In Proceedings of the 20th ACM International Conference on Multimedia (MM '12), New York, USA, pp. 1057-1060, November 2012.