Detection of theft activities using AI based surveillance system

Shambhuraje Mohite[1], Pratik Patil[2], Tushar Patil[3], Prasad Patil[4], Swati Chandurakar[5]

[1,2,3,4,5]Department of Computer Engineering,
Pimpri Chinchwad College of Engineering, Pune, India

***Abstract*: Theft is a standard criminal activity that's prevailing over the years and is increasing day by day. To tackle this downside several investigation systems are introduced within the market. Some are merely videos monitored by a security whereas some are AI-based system capable of detecting suspicious activity that can raise an alarm. However, none of them are intelligent enough to identify what suspicious activity is being taking place and what protecting measures need to be taken in time.**

***Keywords*: Anti-Theft System, Surveillance, AI, Detection.**

Introduction:-

Theft is a relatively prevalent crime that occurs all around the world. According to the National Crime Records Bureau (NCRB), stealing accounts for 80% of all criminal offences. People suffer financially and emotionally as a result of rising theft rates.. Therefore, there's a requirement to develop a system, which is convenient to use, free from false alarms, minimize human interference, and cost-effective toward surveillance. Machine Learning (ML) techniques are useful in developing such efficient systems. The major help is related to:-Detection of motion within the still place. Recognizing facial expressions and also detecting people who wear mash using the model. To detect suspicious activity in the immediate vicinity, if any weapon is present, and to notify the appropriate authorities.

Literature Review:-

The authors of this paper[5] suggest using ml models employed in surveillance to identify a handgun weapon in real time. To detect a handgun, they employed a sliding window and a region-based technique. With 84.21 percent precision, 100 percent recall, and higher true negatives, the Faster Region based Convolutional Neural Network (Faster R-CNN) offers faster, more precise, and exact results. They employed Alert Activation Time per Interval (AATpI), which validates the k number of the following frame and then makes the decision, to create the most accurate decision on raising security alarm when detecting the firearm. Some of their project's big positives include detecting weapons in real time, testing on low-quality YouTube footage, and receiving predicted findings. However, the project's disadvantages include the inability to detect the handgun weapon in the background and faster moving objects, as well as the ability to detect only the handgun.

The authors of this paper[6] proposed a method for detecting the cold-blooded weapon knife in real time. They used an R-CNN based machine learning model with datasets that included images of various types of knives, objects that often appear in the background with knives, and related images that can be used as knives. They present DaCoLT, a brightness-guided preprocessing strategy that increases the model's resistance to brightness changes during both the learning and testing stages. Images with higher brightness are preprocessed in the DaCOLT process at the learning stage by multiplying with a determined darkening factor, and then the contrast of the image is increased with the aid of the CLAHE algorithm to boost the quality. After five consecutive true positives, it successfully triggers the alarm in 19 situations in an average time of 0.41 seconds. This cold steel safety system has a range of applications, including real-time detection of cold steel weapons in video surveillance and parental monitoring of violent videos or videos. Some of the system's drawbacks include detecting weapons in outdoor locations where moving objects may be present in the background and unfavorable weather conditions.

The authors of this paper[7] suggested the concept of weapon detection using a binocular image fusion method in the ml model. They recorded the video using two cameras separated by 9cm in this process. Their goal is to reduce the number of false positives that occur in normal detection as a result of the creation of a weapon-like figure in the background. They first acquired frames from two cameras, then calculated disparity maps using Block Matching (BM) and Semi Global Block Matching (SGBM) to detect objects in three dimensions, then eliminated the context by pre-selecting an area of interest, and finally detected the weapon. In comparison to not using the binary fusion method, accuracy has increased by 13.47 percent, precision by 16 percent, F1 by 10.89 percent, and the number of false positives has decreased by 49.47 percent.

The author in this paper [9], proposed Human pose estimation from 3D image even though it is noisy upto lower threshold. After taking that framed 3D pose while converting them into 2D, there are lots of similar and overlapping configurations of body pose. So they used off-the-shelf detector algorithm to estimate the positions(including ambiguity) of 2D pose structure. For sharpness heat map techniq and Gaussian distributions are used. Finally we use kinematic constraints as well as geometric constraints for

separating Disambiguated Poses.The process mainly consist of three parts first is detection of 2D parts then stochastic exploration of ambiguous hypotheses and finally disambiguation technique. Advantage of this solution is using kinematic constraints ambiguity try to be nullified.

The author in this paper [10], deals with the issue of human posture estimation which is frequently characterized since the computer vision strategies that are under the circumstance of changed human keypoints. Stacked Hourglass Network can build a software to track criminal's action, or make a body language classifier based on a person's pose. Human pose data has lots of variances, which makes it hard to converge so the idea behind stacking multiple HG (Hourglass) modules network is that each HG module will produce a full heat-map for joint prediction. This preserves the location information, and then just need to find the peak of the heat-map and use that as the joint location.

The author in this paper [11], provide human pose estimation using Deep Neural Networks (DNNs).As extreme differentiability in articulations there are a lot of hidden (overlapped) body parts as per perspective of 2D image . Such a problem requires holistic reasoning which can be naturally provided by Deep Neural Networks. This solution gives powerful formulation for estimation of holistic human pose which is simpler than graphical model. Generic convolutional DNN is trained for part detection, feature representations, model topology and joint interactions. For more precision cascading-DNN is provided while dealing with joint-localization.

In this study [12], the author discusses vision-based activity recognition and prediction from videos that usually feature one or more persons. This research has proposed a practical way to detect pose by vision based recognition on the characterstic of each object in an image. First detect the regions of interest (ROIs) in given frame. Then pattern of human motion activity is recognized. Then from the video frame sequences there is extraction of visual information. After the activity .After activity detection, Template matching methods and state space methods are user for activity recognition. Surveillance systems is also implemented.

The author in this paper[13],proposed a system with 2 phases which can detect irregular motion in traffic for crime detection.In its first phase,the modeling and motion patterns learning is done using optical flow for computing vectors of motion,then clustering itusing DBSCAN which represents existing models motion patterns.By similarity and criterions of entropy claasses resulting matched to motion pattern models.This helps in detecting the motions with some irregularities in traffic.

The author in this paper[14],proposes a process in which trajectory of the object is processed for detection of a likely suspicious behaviour.It relies on vector of displacement of object to adjust rectangle to encompass accuarately the tracked object.It compares 2 set of images groups before and after change of object trajectory and detects a remarkable change if theta is high.Once a person is find suspicious the behavious detection by predicting the intention of person takes place.

The author in this paper[15] suggests a process in which first step is to detect if there is a motion. Once motion is detected then frames of images are created and converted to its binary image. Then background substraction before thresholding of the image takes place.Then the movements like boxing,kicking, threatening etc are detected using various algorithms like MIBC, FKNN, MD, LBG, QBG and accuracy of them are compared.

The author in this paper[16] entire training process for a surveillance system may be broken down into three phases: data preparation, model training, and inference. Two neural networks, CNN and RNN(Recurrent Neural), make up the framework .The CNN algorithm is used to extract high-level features from images in order to minimise the input's complexity. For categorization, RNN is utilised, which is ideally suited for video stream processing.The suggested system makes use of a VGG-16 (Visual Geometry Group) pre-trained model that was developed on the ImageNet dataset.Currently, the model is being trained to predict behaviour based on the film. In the footage utilised to enhance the monitoring process, the model can predict suspicious or typical human behaviour.

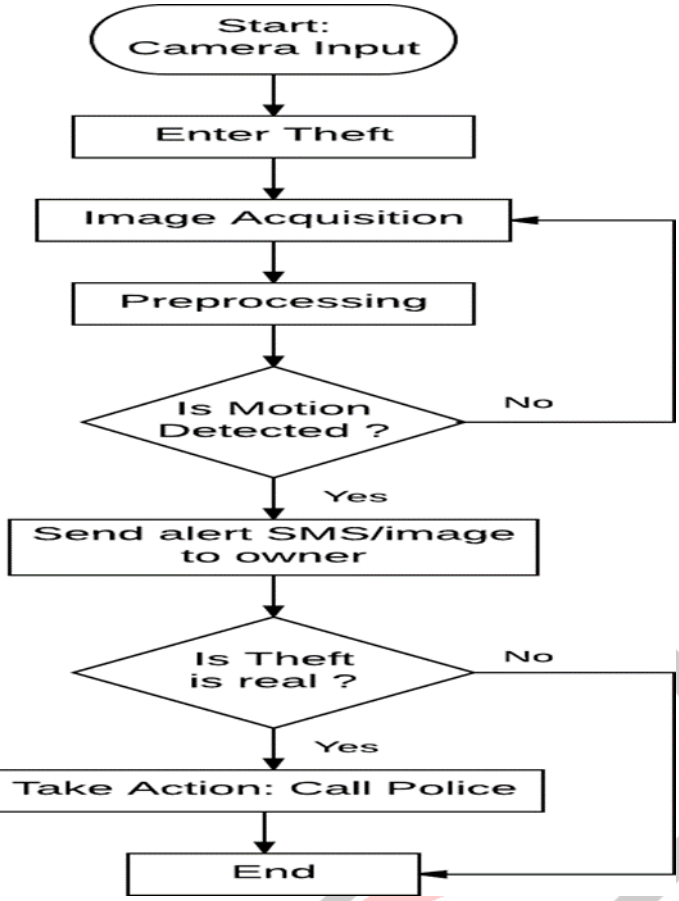
Reference	Consensus used	Contributions	Advantages	Disadvantages
[1] Jae Kyu, Suhr Sungmin, KumHo Gi ,Jung Jihye Kim May 2012	To create facial features, Gabor filters are used.	Automatic facial recognition technologies are used in mask detection technologies.	Provides early notice, allowing guards to be saved.	The effectiveness of facial recognition is limited by poor image quality.
[2] Gahyun Kim,Jae Kyu Suhr,Ho Gi Jung,Jaihie Kim 2010	techniques based on grayscale images	Check to see whether faces are substantially obscured.	The system architecture achieves great performance while still being cost-effective.	The camera should be of decent quality.
[3] Rui Min,Angela D'Angelo,Jean-Luc Dugelay 2010	algorithm for detecting scarves.	detection of partial blockage of the face	An effective tool for improving the security system's performance.	Facial Recognition Is More Difficult With Smaller Image Sizes
[4] Gahyun Kim,Jae Kyu Suhr, Ho Gi Jung, Jaihie Kim 2010	B-spline active contour to motion edges (B-spline active contour to motion edges).	B-spline Active Contour and Skin Color Information for Face Occlusion Detection	This will attempt to eliminate the possibility of ATM theft-related fraud.	Facial recognition technology may be limited by data processing and storage.
Olmos, R., Tabik, S. and Herrera, F., 2018.	R-CNN	It contributes to surveillance .	Detection of a handgun in real time.	It only detects handgun as weapon
Castillo, A., Tabik, S., Pérez, F., Olmos, R. and Herrera, F., 2019.	Faster R-CNN	It contributes in video surveillance and parental monitoring of violent videos.	It shows good precision in robust brightness due to brightness guided technique	Detection of moving objects and adverse weather conditions are some of the system's disadvantages.
Olmos, R., Tabik, S., Lamas, A., Perez-Hernandez, F. and Herrera, F., 2019	CNN	It improves the reliability of the detection in the security field.	It detects the 3d information with dual cameras to focus on the main region in detection by eliminating background.	It requires more hardware and self captured data for training
Pérez-Hernández, F., Tabik, S., Lamas, A., Olmos, R., Fujita, H. and Herrera, F., 2020.	Object Detection with Binary Classifiers based on deep learning (ODeBiC) methodology .	Application in video surveillance.	This methodology decreases the false positives by differentiating similar objects	It can only filter some instances that can cause confusion in the model.
Simo-Serra, E., Ramisa, A., Alenyà, G., Torras, C. and Moreno-Noguer, F., 2012, June.	Hypotheses Clustering and one-class Support Vector Machine (OCSVM).	Detecting disambiguation , pose estimation and configuration even for noisy frames in 3D.	using kinematic constraints ambiguity try to be nullified	in this approach up to 30 pixels only it can bare errors for the pose's 2D localization.

Munea, T.L., Jembre, Y.Z., Weldegebriel, H.T., Chen, L., Huang, C. and Yang, C., 2020.	Stacked Hourglass module	DNN for pose regressor and refine	HG (Hourglass) modules instead of forming a giant encoder and decoder network use HG modules which produce a full heat-map for joint prediction.	peoples interactions that causes complex spatial interference due to occlusion of individual parts by clothes, contact, and limb articulations which makes it difficult for association of parts.
Toshev, A. and Szegedy, C., 2014.	DNN CNN	Joint coordinates have DNN-based regression for formulation of the problem.	DNN has the ability to capture the entire context of all body joints.	Less precision in case of rough image.
Xu, X., Tang, J., Zhang, X., Liu, X., Zhang, H. and Qiu, Y., 2013.	vision surveillance, performance evaluation	quantitative evaluation for performance of recognition in human activity.	proposed solution take the advantages of both CNNs and vision based recognition	Temporal interval is very short so long-term model having temporal correlation is not that much effective.
El Maadi, A. and Djouadi, M.S., 2013, October.	DBSCAN	Focused on the detection of irregular or abnormal motion in traffic.	Works well for detecting abnormalities in crowded traffic.	Errors were caused due to noise which caused discarding of short time detections.
Airfares, W., Kobbane, A. and Krioula, A., 2016, September.	vector of displacement of objects	It focuses on detecting a person's suspicious motions by his sudden and fast movements.	The value of theta calculated mathematically helps a lot in detecting suspicious and sudden movements.	It works good if suspicious movements displace from the original position but doesn't work to that accuracy if suspicious movements are slow.
Yasin, H. and Khan, S.A., 2008, April..	MIBC,Background Subtraction	It helped to compare various algorithms for threat and crime related movements.It gave us the most accurate of the various algorithms compared.	The MIBC is the most efficient of the algorithms.	If we want to increase accuracy we need to train more datasets which causes a big amount of time increase which is a disadvantage of this algorithm.
Amrutha, C.V., Jyotsna, C. and Amudha, J., 2020	CNN,RNN	This model is helpful in any scenario where model needs to be trained accordingly with the suspicious activity suiting for that scenario.	The model can be improved and used in various different scenarios.	The model needs a large dataset to train and reach the accuracy of its potential.

Outline : The components listed below comprise the working model of our system are:

1. Collecting footage and converting it into useful data
2. Data acquisition and analysis
3. Decision making to check criminal activity
4. Trigger the alert system

Capturing video and turning it into usable information:
An IP camera is used to capture video, which is connected to the remote organizer. After the video has been shot, it is processed and broken down into image outlines. These image outlines are then fed to machine learning models. During data cleaning, a lexicon of one-of-a-kind words that appear in all of the dataset picture captions is generated and saved to disc. Informing. And then the corresponding data is fed for next process.

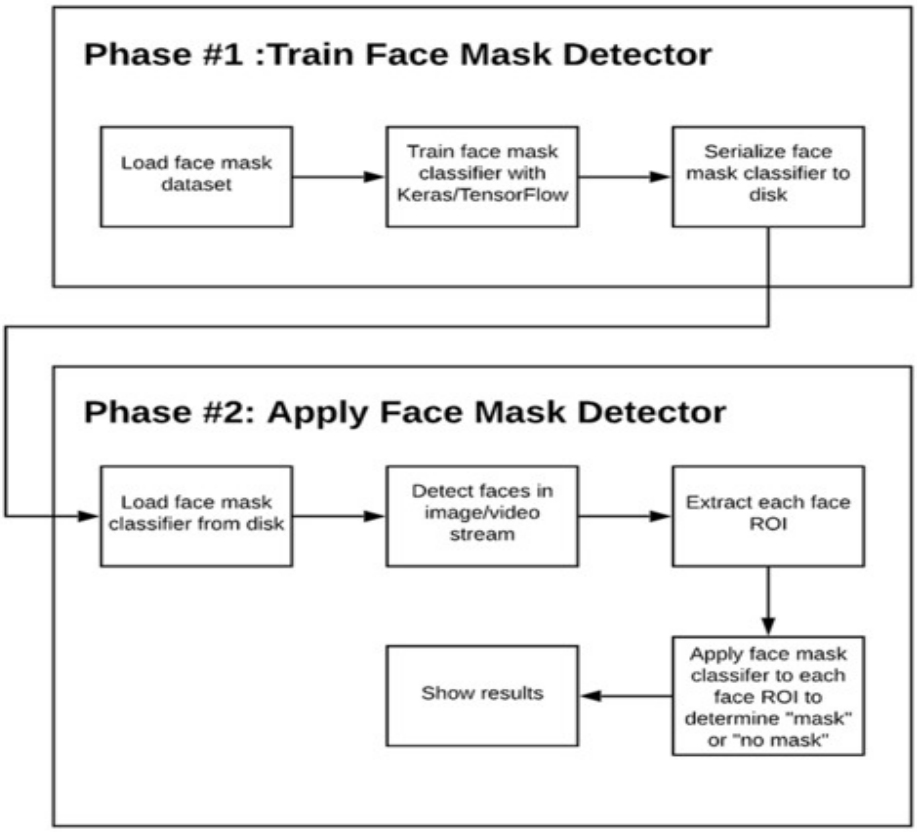


The above activity diagram i.e figure 4.3.2 gives the complete flow of working of the system from the capturing of an image to sending the alert message to the owner.
Start: Camera module inside the room or shop, the Anti-theft device will require a 5-12V power supply. The whole system will start working on providing the continuous power supply to the device.
Enter Theft Once the system is installed and sufficient power supply is provided to it, it will start capturing the image. After closing the shop or room, if any motion happens in front of the camera it will detect it as a Theft entry and start following the next steps of its algorithm.
Decision Making-I The proposed system has a very simple decision-making algorithm, based on feature matching. The first captured image will be stored in the database as the reference image. Objects detected as a result of the new current image are matched with the database values using the technique of feature matching. If there is a match between the current image frame and the reference image frame stored in the dataset, the proposed system will send an alert message as well as the captured image as compared to the large original video to the user/owner. If there will be no match between the two frames, then it continues the process from the image acquisition steps.
Decision Making–II The decision-making algorithm depends on the user/owner of the shop, after getting the alert and capturing the image. The user will see the image and then decide whether the theft is real or not and whether the action has to be taken or not.
Take Action After watching the image, if the user finds it to be a real theft then he can call the police using the “Call Police” option provided in the application of the user or neglect it by clicking the “Neglect” option

Data acquisition and analysis:
Image frames are processed and fed into various machine learning models. Each model completes a task in a specific order in order to examine various evaluation parameters. In order to analyse the results, we used the following ml models.

I) Mask Detection:

Mask detection is done in two phases. In first phase initially we load the face mask dataset. Then train face mask classifier with Keras / tensorflow. Further serialize face mask classifier to disk.
In the second phase load face mask classifier from disk which will result in detection of faces captured in image frames(video stream). Then figure out region of interest for each face. And finally apply the face mask classifier to each face region of interest to determine mask or no mask.
As per this it will show the result.



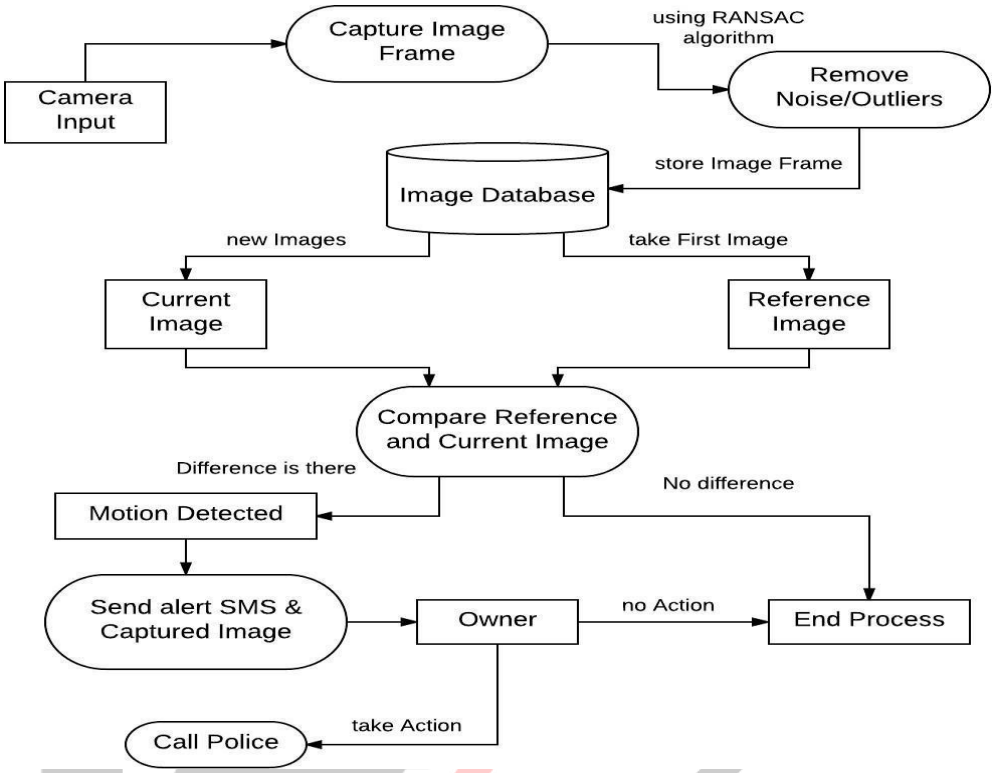
2) Weapon Detection:
While detecting whether an activity is harmful or not, weapon detection will play a measured role. Because in some frames of images if a weapon got detected then the probability of crime occuring will increase. Here for classification KNN algorithm can classify as per trained classes like no weapon, knife, handgun etc.
K-Nearest Neighbour is a Machine Learning technique for Regression and Classification that is based on the Supervised Learning approach.

- The K-NN approach considers the new case/data to be comparable to previous cases, and it assigns the new case to the category that is closest to the existing categories.
- Use Knn to train the model to detect the items you want to categorise (in our example, 0 = No Weapon, 1 = Handgun, and 2 = Knife).
- **Step-1:** Select the number K of the neighbors
- **Step-2:** Calculate K number of neighbours' Euclidean distance.
- **Step-3:**Using the estimated Euclidean distance, find the K closest neighbours.
- **Step-4:**Count the number of data points in each category among these k neighbours.
- **Step-5:** Assign the new data points to the category with the greatest number of neighbours.
-

3) Motion Detection:-
In this we use Change of trajectory by theta angle method proposed by W.Kobanne which is used for detecting suspicious motion. It consists of the following steps:-.
1. Initialization: An initial rectangle encompassing the object (blue curve.We Manually initialize a rectangle surrounding the object in the first frame
2. Then the extracting of interesting points of the object (within the rectangle) in the frame
3.Where *x* is the distance from the origin in the horizontal axis,In the vertical axis, y is the distance from the origin.
4. After every ten frames, a mean displacement vector is determined, as well as an angle THETA between two successive mean displacement vectors. |a|. |b|. cos = a.b
5.Compute the THETA angle between Vu. i+k) and V (i+k,i+k+JO)

6. If theta 1 > theta 2 then Suspicious behavior detection by Object tracking. **Decision making to check criminal activity:**
The system comprises of many levels of surveillance. At each level, the action in each frame of the video will be meticulously monitored using machine learning models that have been specifically taught to do their job**.** In this situation, the model will use numerous parameters to determine whether or not the conduct is criminal. The output of various sub- models, such as mask detection, weapon detection, pose detection, and motion detection, is taken into account as a parameter according to their priority settings**.**

Trigger the alert system:
Finally, the output of these modules is combined to produce an input for another machine learning model that selects to whom the alert message should be delivered, and we employed multiple classification approaches to develop such a model, including logistic regression, SVM, and others. However, decision trees produce the best results. Actions such as raising alarms, sending alert messages to the owner only, sending Alert messages to cops only, and sending an alert message to the owner and cops both will be taken based on the results acquired from the ML models.



Conclusion:-
Though a decent amount of research in past has been done to solve such security problems but still it remains challenging due to increase in the complexity and various thefts taking place daily. The system capture images only when there is any motion in the frame and motions exceed a certain criteria and then further detection takes place. We also got to know the various challenges including accuracy and image quality in environment. Thus the various Algorithms would be helpful in developing the complete system.

References:-
1. Wen, C.Y., Chiu, S.H., Tseng, Y.R. and Lu, C.P., 2005. The mask detection technology for occluded face analysis in the surveillance system. Journal of Forensic Science, 50(3), pp.1-9.
2. Suhr, J.K., Eum, S., Jung, H.G., Li, G., Kim, G. and Kim, J., 2012. Recognizability assessment of facial images for automated teller machine applications. Pattern Recognition, 45(5), pp.1899-1914.
3. Min, R., d'Angelo, A. and Dugelay, J.L., 2010, August. Efficient scarf detection prior to face recognition. In 2010 18th European Signal Processing Conference (pp. 259-263). IEEE.
4. Kim, G., Suhr, J.K., Jung, H.G. and Kim, J., 2010, December. Face occlusion detection by using B-spline active contour and skin color information. In 2010 11th International Conference on Control Automation Robotics & Vision (pp. 627-632). IEEE.
5. Olmos, R., Tabik, S. and Herrera, F., 2018. Automatic handgun detection alarm in videos using deep learning. Neurocomputing, 275, pp.66-72
6. Castillo, A., Tabik, S., Pérez, F., Olmos, R. and Herrera, F., 2019. Brightness guided preprocessing for automatic cold steel weapon detection in surveillance videos with deep learning. Neurocomputing, 330, pp.151-161.
7. Olmos, R., Tabik, S., Lamas, A., Perez-Hernandez, F. and Herrera, F., 2019. A binocular image fusion approach for minimizing false positives in handgun detection with deep learning. Information Fusion, 49, pp.271-280.

8. Pérez-Hernández, F., Tabik, S., Lamas, A., Olmos, R., Fujita, H. and Herrera, F., 2020. Object detection binary classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillance. Knowledge-Based Systems, 194, p.105590.
9. Simo-Serra, E., Ramisa, A., Alenyà, G., Torras, C. and Moreno-Noguer, F., 2012, June. Single image 3D human pose estimation from noisy observations. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2673-2680). IEEE.
10. Munea, T.L., Jembre, Y.Z., Weldegebriel, H.T., Chen, L., Huang, C. and Yang, C., 2020. The progress of human pose estimation: a survey and taxonomy of models applied in 2D human pose estimation. *IEEE Access*, *8*, pp.133330-133348.
11. Toshev, A. and Szegedy, C., 2014. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1653-1660).
12. Xu, X., Tang, J., Zhang, X., Liu, X., Zhang, H. and Qiu, Y., 2013. Exploring techniques for vision based human activity recognition: Methods, systems, and evaluation. *sensors*, *13*(2), pp.1635-1650.
13. El Maadi, A. and Djouadi, M.S., 2013, October. Suspicious motion patterns detection and tracking in crowded scenes. In 2013 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR) (pp. 1-6). IEEE.
14. Airfares, W., Kobbane, A. and Krioula, A., 2016, September. Suspicious behavior detection of people by monitoring cameras. In 2016 5th International Conference on Multimedia Computing and Systems (ICMCS) (pp. 113-117). IEEE.
15. Yasin, H. and Khan, S.A., 2008, April. Moment invariants are based on human mistrustful and suspicious motion detection, recognition and classification. In Tenth International Conference on Computer Modeling and Simulation (uksim 2008) (pp. 734-739). IEEE.
16. Amrutha, C.V., Jyotsna, C. and Amudha, J., 2020, March. Deep learning approach for suspicious activity detection from surveillance video. In 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) (pp. 335-339). IEEE.