Condensed Object Representation with Corner HOG Features for Object Classification in Outdoor Scenes

Tin Tin Yu
University of Computer Studies, Mandalay
Myanmar
baybayyu@gmail.com

Nu War
Computer University, Mandalay
Myanmar
nuwar81@gmail.com

***Abstract*—Nowadays, HOG (Histogram of Gradient) feature is extracted from the objects and using it in the classification tasks among the many visual application systems such as object tracking, action recognition and automated video surveillance. Most techniques of extraction HOG feature are based on cells and blocks. Although the HOG feature on cell and block are being robust for current visual systems, the alternative way to extract HOG feature that focus on corner points are presented in this paper. HOG features on corner points is extracted for multiple object detection system in which single or multiple moving objects are classified and labeled. And also comparison results on outdoor challenging sequences for HOG feature extraction on blocks and corners are provided with experimental results.**

Keywords— HOG (Histogram of Gradient), Object Tracking, Action Recognition

I. INTRODUCTION

Classification and recognizing objects is the process of tracing the location of interested objects in video sequences. The main application of object tracking comes in the field of automatic video surveillance. In recent years, almost all buildings and roads are provided with closed circuit television (CCTV) security cameras. The visuals from these CCTV cameras serve a great deal in monitoring human activities. Thus these become very much helpful in prevention of crime, theft, etc. The main problem associated with this is that a lot of human workforces are required for the monitoring of these video sequences. It is in this context that the need for an automatic surveillance system arises. Tracking an object of surveillance camera is also demanding to improve human action recognition and classification problems.

To accomplish in handling all challenging scenarios with one single approach will not be possible. Even emerging the new, robust and advanced approaches for object detection and tracking, its research trend still remains as the most interested field in visual application systems. Since changes in target appearance are the main sources of target drifting to the background, loss of target signal, handling appearance changes become critical factors, how one defines the object highly impact to the difficulty level of this challenging factors. Success in object recognition and tracking relies heavily on how robust the representation of target is against the variation challenges. If a robust target representation can be developed that can be adaptive with variations, the accuracy of classification and tracking result can improve significantly. This challenge becomes matter of primary interest for robust online tracking in the recent researches.

Most existing tracking systems are based on color-feature distribution of object in which a target object is represented by means of color. The efficiency of object representation can deteriorate when several objects move together in image sequence or when sudden illumination changes happen or when background color is similar with the target. Color histogram based approaches still challenging to deal with the complex tracking. So the new type of robust image feature needs to be utilized to find the corresponding region between the consecutive frames by means of feature point detection. HOG feature on corners will be used in this work for proposed compact representation of objects to be classified and tracked.

Related works will be described in section 2, and proposed framework on tracking will be discussed in section 3 and in which the detail progress the proposed system will be provided. Experimental result on the steps of our work will be presented in section IV. The last portion will draw conclusions.

II. RELATED WORKS

Extraction of HOG feature was proposed for human detection by Dalal and Triggs [2], and nowadays the use of HOG feature can be seen in visual applications like object detection, image registration and object tracking. Although HOG was intended for human detection, several literature surveys confirm that the popularity of HOG in visual applications. As HOG descriptor is an quantization of gradient

978-1-5090-5504-3/17/$31.00 ©2017 IEEE
SNPD 2017, June 26-28, 2017, Kanazawa, Japan

orientations that are weighted by the magnitudes of these gradient vectors, this type of feature representation is especially appropriate for some objects which have strong edges and corners such as vehicles and people.

In 2008[1], the researchers presented an approach in which contain two trackers for the purpose of tracking by detection. Using global seed features, it can support robust segmentation result via Adaboost. To improve the object segmentation mask, Adaboost was also applied on region based local seed features. For the purpose of handling the weak point for high dimension problem of pixel wise approaches, feature selection algorithm was developed in feature classification. But this approach was not able use to online tracking because it post processing was also needed especially when the occlusion occurred.

The authors [9] proposed the multiple object tracking in which object detection was done by the use of background subtraction approach. To classify and recognize object, scale invariance feature transformed (SIFT) feature and motion of object were applied. To estimate the trajectory of object, Kalman Filter was adapted. As other feature point based matching mechanism, threshold was critical factors. This tracker can track both fast or slow motion and also no manual selection need to be performed. The author [17], in 2009, proposed an on-line EM based algorithm for tracking in which objects are defined by means of SURF features as a template matching. Robust segmentation was essentially needed as a pre-processing. The performance of tracker can be degraded when object was too small or homogeneous blobs contained. The tracker robust the some challenges such as insensitive to Appearance changes, background clutter, illumination changes and occlusion. In 2012, region-based particle filter was presented for general object tracking by segmentation. In this approach, color feature was adopted to represent homogeneous regions. Due to the limitation of color based representation approach, the tracker lost the target when the illumination variation. The approach can strongly track in shape variations [15].

Hwann-Tzong Chen, et.al [6] in 2014, proposed feature point based object tracking system. That was aimed for dynamic scenes on FPGA System prototyping. The initial object segmentation and background initialization were not needed. The fixed region of interest (ROI) was used to define the specific objects to be tracked. As consequences of using fixed ROI, the involvement of background pixels in ROI problem raised. Reducing pre-processing steps such as object segmentation and background subtraction caused the degrading of computing complexity.

In order to get a compromise between the validity and efficiency, the author [5] proposed the edge-color co-occurrence centroid (ECCC) based target representation to achieve an invariant target appearance model in which the color space was non-uniformly clustered to accent gray-scale of target objects occurring in high frequency according to the color distribution of objects.

In 2015[12], adaptive covariance search window approach was proposed for tracking object. But when search window can become to full image, complexity of covariance matrix could not be largely reduced. Occlusion detection result was getting optimal and integral region computation was the one of additional complexity. During recent year 2016[4], the researcher proposed a mechanism in which the using of MC-HOG features joints with the correlation tracking framework. The target model drifting problem caused by severe occlusion could be solved by the proposed saliency map. This map was applied as prior information to condense the including background portion in the region of interest for target object. Performance of this map will decrease in multi objects as a limitation. However, the model drift problem caused by occlusion can be handled by re-ranking on saliency map.

III. Visual Feature-Based Object Classification and Tracking System with C_HoG Features

The fundamental architecture of feature-based visual system as shown in figure 1. These four steps have to be taken to accomplish the tracing of objects in video sequences.

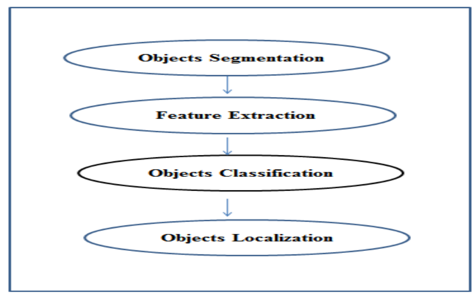

Fig. 1 Feature-based Object Classification and Localization System

The architecture of visual objects tracking via local features on corner point is presented in figure2. In the intended system, as the first step, moving objects will be detected in the form of regions from background subtracted frame with the help of Gaussian Mixture background modeling. For the portion of feature extraction, the strongest features points on corners will be extracted for each moving objects by the Features by Harris Corner Detector. Comparing with other blobs detectors such as SIFT and SURF, Harris a corner based detector with high speed with can get more feature points. One drawback of blob detector is that the object region can contain the background pixels that may reduce the discriminative power of features.

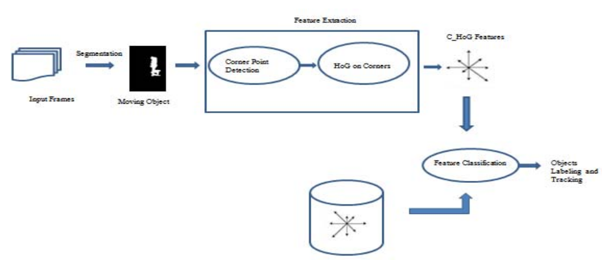

Fig. 2 Framework of Object Classification and Tracking System with C_HoG features

A. *Objects Segmentation*

Detecting of moving objects in a video sequence is a ultimate step of visual applications such as people tracking, traffic monitoring, visual surveillance. Many background models to solve with the different challenging scene. In background subtraction, the comfortable way is to model the background with only inactive pixels that are so called stationary points. Unfortunately, such background cannot available especially for clouded scenes and background usually contains dynamic changes such as objects being appeared and eliminated from sight of the sequence unexpectedly. To control these kinds of problem, several amount of background modeling techniques have been proposed in these years [20].

1) Background Subtraction:

The core and first step in tracking is also the object segmentation or object detection in which the foreground is separated from the background. All the moving parts in video sequence are regarded as the foreground parts to be tracked through the video sequence. Background subtraction method is an effective moving target detection method. Reliable background modeling and robust foreground detection are open researches for complex scenes and real time environments. The instinctive segmentation of moving targets in visual application system requires efficient algorithms that are able to robust challenging scenes. The usual method is native background subtraction because it is easy to implement. In these background subtraction techniques, each pixel will be classified whether its existence as foreground portion or background portion. Background Subtraction techniques are characterized as: 1) non-recursive and 2) recursive in [11]. An efficient BGS algorithm should be strong when it encounters with illumination changing conditions, robust when background contains the movement of tiny and slight objects and accomplish of suddenly joining new objects into the background model. All these requirements must be able to perform with acceptable computational requirements including in real time environments.

2) Gaussian Mixture Model (GMM)

This is one of the popular and widely used approaches in visual application systems including for moving object detection. In recent years, it becomes more popular because of its robustness and powerful impact of several improvement finding and enhancement on the original and typical GMM [13]. Typical model can successfully concerned with sequences including shadows, shading tree, slow motion, fasting motion and other severe problems that cause to fall down of detection rate in real world applications. In this work, moving object detection step was taken by the Adaptive GMM [10].

In adaptive GMM, background is modeled by each single particular pixel in the scene with a mixture of K Gaussian distributions. The probability that a certain pixel has a value of x_N at time N can be represented as

$$p(x_N) = \sum_{j=1}^{K} w_j \eta(x_N; \theta_j) \tag{1}$$

where,
w_k is the weight parameter of the k^{th} Gaussian component
$\eta(x; \theta_k)$ is the Normal distribution of k^{th} component represented by

$$\eta(x; \theta_k) = \eta(x; \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{\frac{D}{2}}|\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)} \tag{2}$$

where, μk is the mean and $\Sigma_k = \sigma_k^2 I$ is the covariance of the k^{th} component. The K distributions are ordered based on the fitness value w_k/σ_k and the first B distributions are used as a model of the background of the scene where B is estimated as

$$B = \arg\min_b\left(\sum_{j=1}^{b} w_j > T\right) \tag{3}$$

The threshold T is the minimum fraction of the background model. In other words, it is the minimum prior probability that the background is in the scene. Background subtraction is performed by marking a foreground pixel any pixel that is more than 2.5 standard deviations away from any of the B distributions [10].

$$\hat{w}_k^{N+1} = (1-\alpha)\hat{w}_k^N + \alpha\hat{p}(w_k|x_{N+1}) \tag{4}$$

$$\hat{\mu}_k^{N+1} = (1-\alpha)\hat{\mu}_k^N + \rho x_{N+1} \tag{5}$$

$$\hat{\Sigma}_k^{N+1} = (1-\alpha)\hat{\Sigma}_k^N + \rho(x_{N+1}-\hat{\mu}_k^{N+1})(x_{N+1}-\hat{\mu}_k^{N+1})^T \tag{6}$$

$$\rho = \alpha\eta(x_{N+1}; \hat{\mu}_k^N, \hat{\Sigma}_k^N) \tag{7}$$

$$\hat{p}(w_k|x_{N+1}) = \begin{cases} 1; & \text{if } w_k \text{ is the first match Gaussian component} \\ 0; & \text{otherwise} \end{cases} \tag{8}$$

Where w_k is the k^{th} Gaussian component
$1/\alpha$ defines the time constant which determines change.
If none of the K distributions match that pixel value, the least probable component is replaced by a distribution with the current value as its mean, an initially high variance, and a low weight parameter[10].Foreground masks for moving objects were taken by post processing such as morphological processing, removing filtering to increase the segmentation results. These segmentation results are shown in section 4A.

B. *Feature Extraction*

The key portion of feature based system is to consider what strong and robust feature will be extracted. Feature correspondence is also the challenging problem because a feature point in one region of image may have many similar points in another region of this image or another image and it tends to accuracy in feature correspondence uncertainty. Selecting the right and covered feature plays a very important role in video object tracking.

1) Block HOG *Feature Extraction (B_ HOG)*

As in Figure 3, the detector window is separated into the dense gird of cells. Image gradient vector is evaluated for all pixels within the cell. Each gradient vector has its own gradient angle and gradient magnitude. Angle of vector is voted for the corresponding orientation bin and magnitude of vector is weighted to each vote. HOG feature is compiled for each grid cell. A particular cell has the local histogram passing by orientation bins. These cells are combined into blocks and then HOG normalization is applied in order to improved accuracy. HOG normalization results tend to resist for illumination changes and shadowing [8].

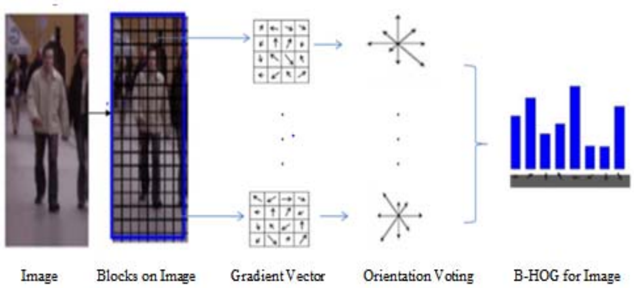

Fig 3. Block based HOG Feature Representation (one histogram per block)

2) Motivation

Edge detection purposes to identify the margin point of objects in an image. The main concept behind the edge detection is that edge is the point where sharpens intensity changes, in other words, edge points have more information rather than other points. By detection edge and processing the features on these detected edge areas tends to moderate the amount of features to be processed apparently.

Corner point is the intersection of two edge points. Commonly, the pixels occupy at the junction of two edges would be the possibility of strong and good points for feature correspondence in detection system. Corner based detectors are widely used in classification and recognition task because of their computational properties [7]. Among the corner point detectors, such as FAST, MinEigen, and Harris detector could produce the strongest corner points with the acceptable rate for tracking system. These HOG features on corner points will be used in matching objects between consecutive frames of tracking sequences.

3) Corner HOG *Feature Extraction (C_ HOG)*

In the proposed feature extraction, firstly corner point is extracted from detected ROI moving objects via Harris Corner Detector [3]. For a pixel p, use its neighbourhood (e.g. 7x7) to form the following matrix.

$$C = \begin{bmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{bmatrix} \quad (9)$$

where I_x, I_y are image gradient components [11]. If the smaller Eigen-value of this matrix is larger than a certain threshold, it is considered a corner as shown in figure4.

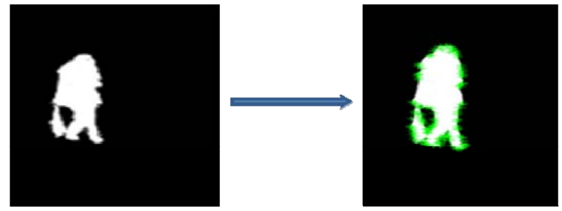

Fig 4. Visualization of Corner Point on Detected Moving Object

The strongest features points on corner are extracted for each moving objects by Harris Corner Detector and gradient of each point via HOG are visualized into bins as shown in figure 5.

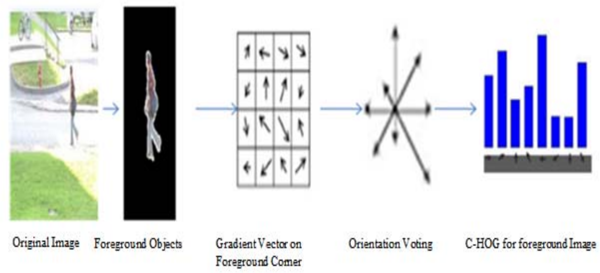

Fig 5. Proposed Corner based HOG Feature Representation (one histogram per object)

C. *Feature Classification and Object Localization*

In feature classification steps, the features on the block and corner of each moving objects are extracted, trained and tested respectively with the help of multi-support vector machine (multi-SVM). Firstly, a multiclass model is trained using SVM learners and SVM template using error-correcting output codes (ECOC) multiclass model [20]. And then the test data are predicted with the trained feature. According to the classified result on features of model, the objects are labeled as shown in figure 9.

IV. RESULTS

In these experiments, four datasets of outdoor sequences were shown with the illustration figures and tables for the result of objects segmentation step. Each dataset represent one category of challenging scenarios for baseline, bad weather, camera jitter and clouded area respectively.

A. Experimental Result for Segmentation

The experiment is performed on challenging sequences with the joint method for co-segmentation and dense correspondence estimation in two images [14]. This approach accomplished by taking the segmentation mask obtained by object segmentation step and reference mask (ground truth).

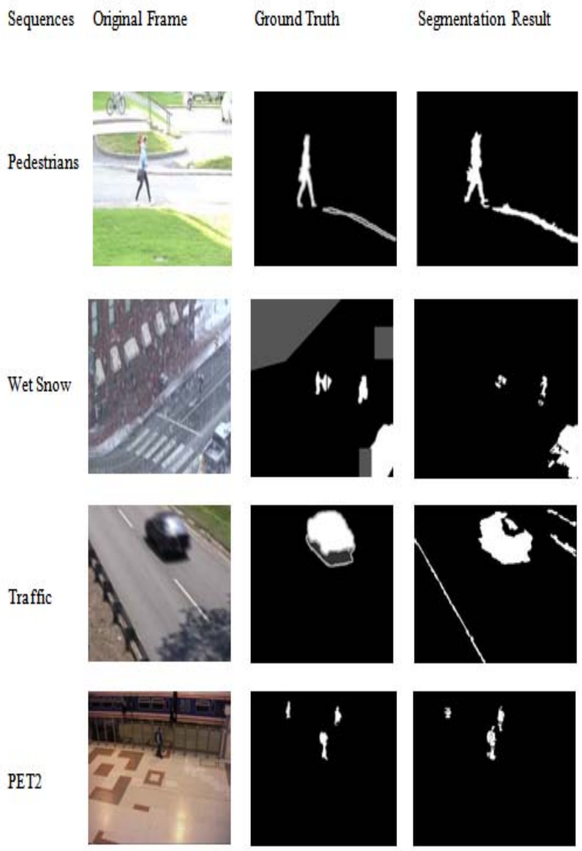

Fig.6 Segmentation Results in Objects Segmentation Step

TABLE I. PROPERTIES OF SEQUENCES

Sequences	Resolution	Challenges
Pedestrians	240*360	Base Line
Wet Snow	540*720	Bad Weather
Traffic	240*320	Camera Jitter
PET2	576*720	Clouded Area

TABLE II. SEGMENTATION ACCURACY ON EACH SEQUENCE

Sequences	Segmentation Accuracy
Pedestrians	0.864917
Wet Snow	0.654332
Traffic	0.498295
PET2	0.725397

B. Feature Classification and Object Labeling

The classification results on three dataset are shown in table III with the properties of trained feature, tested features and classification accuracy on each feature types (CHOG and BHOG) is proved with confusion matrix in table IV and V . Classification accuracy and time for each feature are also demonstrated with bar chats in figure 7 and figure 8. According to the accuracy results, CHOG features are more robust when dataset contains people while BHOG features have strong discriminative on car or vehicle. Both these two features depend on the segmentation result from the object detection step. Visualization of object localization and labeling objects with corner HOG feature can be seen in figure 9.

TABLE III. OBJECT CLASSIFICATION RESULTS

Dataset	Feature	Feature Length	Classification Accuracy	Classificati on Time
Pedestrians Trained =307 Test=1053	C_HOG	81	0.6486(65%)	14.0603s
	B_HOG	45936	0.6087(61%)	461.4877s
Highway Trained =333 Test=285	C_HOG	81	0. 8147(81%)	3.8055s
	B_HOG	45936	0. 8307 (84%)	124.9041s
PET 2006 Trained =500 Test=700	C_HOG	81	0.7948(79%)	9.3468s
	B_HOG	45936	0.7501(75%)	306.7819s

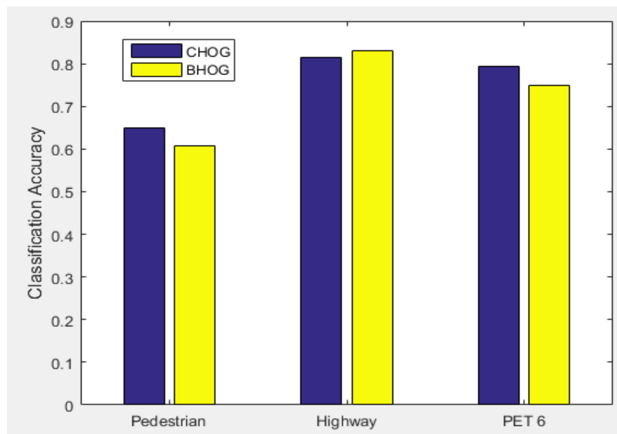

Fig .7 Demonstration of Classification Accuracy for C_HOG and B_HOG

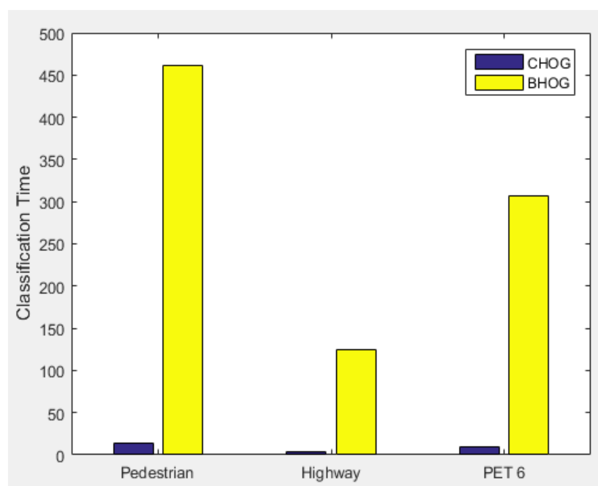

Fig 8. Demostration of Classification Time for C_HOG and B_HOG

TABLE IV. CONFUSION MATRIX FOR BLOCK BASED HOG (B_HOG) FOR PEDESTRIANS DATASET

Output Class (columns) / Target Class (rows)

Objects	1	2	3	4	5	6
1	333	0	0	0	0	0
2	0	32	61	0	0	31
3	0	0	117	0	17	60
4	0	0	53	0	0	0
5	0	0	83	0	17	21
6	0	0	83	0	3	142

TABLE V. CONFUSION MATRIX FOR CORNER BASED HOG (C_HOG) PEDESTRIANS DATASET

Output Class (columns) / Target Class (rows)

Objects	1	2	3	4	5	6
1	333	0	0	0	0	0
2	0	35	34	26	0	29
3	0	0	142	0	0	52
4	0	2	8	3	0	40
5	0	0	51	30	0	40
6	0	3	22	33	0	170

Fig. 9 Visualization of Detection and Objects Labeling

V. CONCLUSIONS

This proposed work is intended to propose an extraction method of discriminative features on corners and compact object representation to handle the large amount of local features in feature correspondence. The alternative way of how to getting robust HOG features on corner points (C_ HOG) as a compact target representation scheme is presented as the main theme of this paper. Since accessing HOG feature on corner points has a moderate amount of feature length, computation time is desirably achieved over block based HOG on classification time. Using these features in labeling objects between consecutive frames of challenging sequences will be one of the contribution points of the intended research work. Exploring in better segmentation methods especially for shadow detection and removal will be included in future work in order to achieve the more accurate results.

REFERENCES

[1] Fan-Tung Wei,et.al," A region-based object tracking scheme using Adaboost-based feature selection" IEEE International Symposium on Circuits and Systems, Ipp. 2753 – 2756, 2008.

[2] Dalal, N., and Triggs, B. "Histograms of oriented gradients for human detection", In CVPR, volume 1, 886–893. IEEE. 2005.

[3] C.Harris and M. Stephens (1988). "A combined corner and edge detector " (PDF). Proceedings of the 4th Alley Vision Conference. pp. 147–151.

[4] Guido Zhu, et.al. "MC-HOG Correlation Tracking with Saliency Proposal" , 2016, Association for the Advancement of Artificial Intelligence (AAAI).

[5] Hong Lu, "Edge and color contexts based object representation and tracking," , ELSEVIER Optik (August 2014).

[6] Hwann-Tzong Chen, et.al ,"Feature Points Based Video Object Tracking for Dynamic Scenes and Its FPGA System Prototyping" : Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2014 Tenth International Conferences

[7] J. Shi and C. Tomasi (June 1994). ", 9th IEEE Conference on Computer Vision and Pattern Recognition. Springer.
Mahmud Farshbaf Doustar ,: "MATLAB implementation of HOG (Histogram of Oriented Gradients) " Saturday July 26, 2014 by at 11:5611:56

[8] Md. Saidur Rahman,"Multi object tracking in video sequences based on background subtraction and SIFT feature Matching", IEEE Transactions on Image Processing,2009.

[9] P. Kaewtrakulpong, R. Bowden, "An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection" . In Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems, AVBS01, And VIDEO BASED SURVEILLANCE SYSTEMS: Computer Vision and Distributed Processing (September 2001).

[10] Robert Collins: Lecture 06: Harris Corner Detector, CSE486, Penn State. Reading: T&V Section 4.3

[11] Ruthann He · ,et.al, :Integral region-based covariance tracking with Occlusion detection: Springer Science+Business Media New York 2015.

[12] Stauffer, C. and Grimson, W.E.L, " Adaptive Background Mixture Models for Real- Time Tracking". Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, Vol. 2 (06 August 1999), pp. 2246-252 Vol. 2.

[13] Tatsunori Taniai, Sudipta N. Sinha, Yoichi Sato "Joint Recovery of Dense Correspondence and Co- segmentation in Two Images", CVPR 2016.

[14] Varas, et.al. "Region-based particle filter for generic object tracking and segmentation", 19th IEEE International Conference on Image Processing (ICIP), pp. 1333 – 1336, 2012.

[15] Wei He,et.al, :"SURF Tracking" :2009 IEEE 12th International Conference on Computer Vision (ICCV)

[16] https://uk.mathworks.com/help/vision/examples/detecting-cars-using-gaussian mixture-models.html.

[17] https://hal.archives-ouvertes.fr/hal-338206/document.

[18] https:// sites.google.com/site/tracker benchmark/benchmarks/v10

[20] https://www.mathworks.com/help/stats/fitcecoc.html