International Journal of Recent Engineering Science (IJRES)
ISSN: 2349 – 7157 / doi: 10.1445/23497157/IJRES-V5I1P103
Volume 5 Issue 1, 11-15, Jan-Feb 2018
© 2018 Seventh Sense Research Group®

Adaptive Target Tracking based on Kernel SVM and HOG Features

[1]Yarui Wang, [2]Zuolei Sun, [3]Xiaoyu Wang
[1]*Author is with the Machine Perception and Interaction Group, MPIG, Shanghai Maritime University, China.*
[2]*Author is with the Machine Perception and Interaction Group, MPIG, professor, Shanghai Maritime University, China.*

Abstract
The traditional target tracking algorithm is popular with its real-time. Based on its discriminative model, this paper extracts HOG features into the Kernel SVM classifier to distinguish the foreground and background, which also uses a sliding window with an adaptive scale to divide the image into different scale image blocks. Meanwhile, aiming to improve the speed of calculation, we perform Fast Fourier Transform on the input image to process it in the frequency domain and then obtain rotating invariant gradient information as its feature map to distinguish the target and background the kernel SVM classifier. We performed our experiments on six challenging videos from the visual tracker benchmark and compared them with two well-known KCF and CT tracking algorithms. The experiment results show that our approach can track the targets accurately in a fast way, even though there are challenges such as motion blur and occlusion in the testing videos.

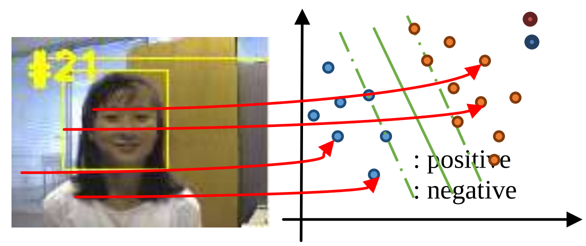
Index Terms - *Object tracking, HOG feature, kernel SVM*

I. INTRODUCTION

TARGET tracking is a blend of machine learning, motion analysis, pattern recognition, artificial intelligence, and other advanced technology. With the rapid development of machine learning, numerous relevant algorithms have been introduced in kinds of literature, and the application based on video frame is more and more popular with scientific researchers due to real-time and pragmatic [1], [2], [3], [4]. The traditional target tracking algorithm is divided into a generative algorithm and discriminative algorithms. The former first establishes the appearance of the target online or offline, then uses the sliding window to search for the candidate target to minimize the reconstruction error to determine the location of the target; The latter takes tracking problem as a classification problem between goals and background, and it aims at separating the target from the background, and then applies the various methods of machine learning to representation and updating of the target model, for the purpose that accommodating the intrinsic change of target or the external changes in the environment.

SVM can represent the target tracking problem as a binary classification problem for the foreground and background. The feasibility based on SVM for tracking lies in Fig. 1; the eigenvector is extracted from each region of the target and the surroundings in each video frame, mapping to the feature space, training respectively positive and negative samples into the classifier and then using the SVM classifier to classify the foreground and background pixels of the target area in each video frame, the target tracking on video sequence can be achieved.

Fig. 1. The illustration of target tracking for Support Vector Machine



Compared with other feature description methods, Here, we select the Histograms of Oriented Gradients features to describe our image. Since 2005, the usage of HOG features and SVM became the best choice for real-time target detection, considering that they are sensitive to geometric and optical deformation of the image; HOG features combing with SVM was first advocated to apply to human detection [5]. Though deep learning is hot now, traditional algorithms are still favoured due to their real-time. Literature [6] rotated the UAV image to align the roads of an image with the vertical and horizontal direction and further developed HOG and SVM methods to improve tracking efficiency. Literature [7] uses the Gaussian mixture model to track a specific person; simultaneously, the HOG features in the face region is provided to SVM classifier to identify human in surveillance videos. Besides, HOG and SVM also made use of road mark recognition and automated segmentation of iris images and so on [8], [9].

We proposed HOG features combing with the kernel SVM classifier to track the target of the image. Furthermore, we perform Fast Fourier Transform on the gradient and use the transformed gradient histogram to obtain rotating invariant histogram information as its feature map to distinguish the target and background by kernel SVM classifier [10], [11]. The experimental results show that the proposed tracking algorithm combing the HOG features, and kernel SVM classifier can guarantee the rotation invariance, ensuring the accuracy of in-Plane rotation and illumination variation and occlusion.

This paper is organized as follows. In section II, we describe our proposed method in detail. In section III, we revealed our experimental results and analysis. Finally, the conclusion made.

II. THE PROPOSED METHOD

The general tracking problem is divided into three steps [12], [13], [5]. Firstly, we sample near a position P_t and train a regression in the current frameI_t, where the regression can calculate the response of small sample window; Secondly, we sample at position P_t of the previous frame, then determine the response of each sample with the regression mentioned above; Finally, we can get the most robust response of sampling as the position of the next frame P_{t+1}.

The proposed paper is based on literature **Error! Bookmark not defined.**] and uses the Gamma correction method to standardize the colour space of the input image, then calculates the gradient of each pixel in each sliding scanning window, and then links all the HOG descriptors in the image to get the HOG feature descriptor. After that, we construct the training samples of the kernel SVM classifier through the cyclic matrix offset to make the data matrix become a circular matrix [15], next we transform the solution of the characteristic problem based on the cyclic matrix into the Fourier Transform domain, thus avoiding the process of matrix inversion, which dramatically reduces the complexity of the algorithm. Fig. 2 shows our target tracking framework based on kernel SVM and HOG descriptors.

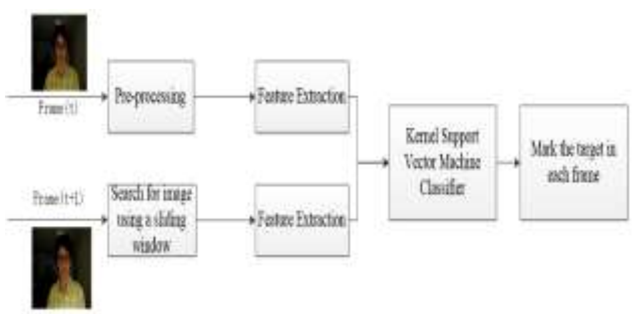


Fig. 2. The target tracking framework based on kernel SVM and HOG descriptors

A. *Feature Extraction based on HOG Features*

The core idea of the histogram of oriented gradients is that the local area of information can be well described by the direction density distribution of gradient or edge [16]. The image is susceptible to light, viewing angle, azimuth, noise and so on. To ensure illumination invariance and weaken noise, the algorithm first normalizes these local histograms in a broader range of images. And then divides the images into small connected regions, called a cell, that is, image cell units, next counts the gradient histogram of each image cell, the histograms are finally combined to save the spatial information.

As Fig. 3 shows, the size of the image is 160×40; we set each cell to 6×6 pixels and use 9 bins to compute the gradient information of 6×6 pixels, where 2×2 cells are a block. Finally, we can get the 6201×36-dimensional eigenvector.

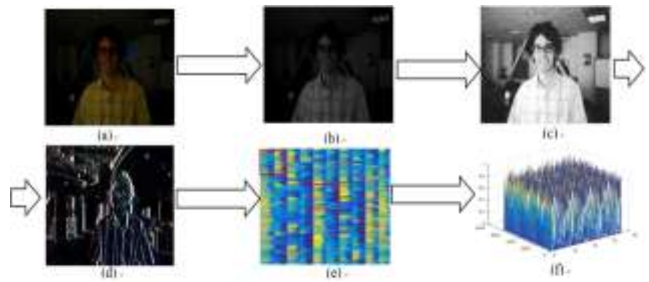


Fig. 3.The general processing of extracting HOG features. Firstly, we can normalize gamma&colour. Secondly, we computed gradients to get image (d). Thirdly, as can be seen in image (e), we will accumulate weighted votes for gradient orientation over spatial cells and then normalize contrast within an overlapping block of cells. Finally, we can collect HOGs for all blocks over sliding detection window.

B. *Classification and Discrimination based on Kernel*

SVM proposed by Vapnik is based on the principle of structural risk minimization in linear function hypothesis in high dimensional feature space [30]. It integrates several techniques such as maximum interval hyperplane, mercer kernel, convex quadratic programming, sparse solution, and relaxation variables. In several challenging and real-time applications, it has achieved the best performance so far.

The realization of the kernel SVM classifier is divided into two stages: training stage and classification stage. This paper employs the kernel SVM library that OPENCV has trained. In the classification stage, the 6201×36-dimensional eigenvector is sent to the kernel SVM classifier. Then the SVM classification function f（x） is used to determine the category with the highest confidence as the output results.

Speaking of linear classification problems, for the current frame, for the target we input, we aim to find a function $y_i = \mathrm{f}(x) = < w \cdot \mathrm{x} > = w^T x$ to train and classify target and background over samples x_i and their targety_i. To maximize margin width $\mathrm{M} = \frac{2}{||w||}$. Through minimizing the squared error, we can get equation (1), where n means the numbers of samples.

$$min \frac{1}{2} ||w||^2$$
$$subject\ to\ \ y_i[(w^T x_i) + b] - 1 \geq 0$$
$$(i = 1,2, \ldots \ldots, n) \qquad (1)$$

Then we apply Karuch-Kuhn-Tucker conditions and Lagrange dual into our equation (1) to get,

$$L(w, b, \alpha) = \frac{1}{2} ||w||^2 + \sum_{i=1}^{n} \alpha_i (y_i (w^T x_i + b) - 1) \quad (2)$$

Next,let $\theta(w) = max_{\alpha_i \ge 0} L(w,b,\alpha)$, $\min_{w,b} \theta(w) = \min_{w,b} \max_{\alpha_i \ge 0} L(w,b,\alpha) = p^*$, $\max_{\alpha_i \ge 0} \min_{w,b} L(w,b,\alpha) = d^*$, here $d^* \le p^*$, then we ask for partial guidance of $L(w,b,\alpha)$ and we can get $w=\sum_{i=1}^{n} \alpha_i y_i x_i$, $\sum_{i=1}^{n} \alpha_i y_i = 0$. As early mentioned, the optimal solution equation (3) is given by [18], [19].

$$L(w,b,\alpha)=\frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j - b\sum_{i=1}^{n} \alpha_i y_i + \sum_{i=1}^{n} \alpha_i = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (3)$$

When it comes to non-linear classification problems, sample x can be mapped to feature space using linear learners in high-dimensional feature space. Therefore, considering the hypothesis set is this type of function equation (4),

$$f(x) = \sum_{i=1}^{n} w_i \phi_i(x) + b \quad (4)$$

In the formula, ϕ: $X \rightarrow F$ is a certain mapping from the input space to a feature space, as Fig. 4 shows. That is to say, resuming the non-linear classifier needs two steps; we can employ a non-linear mapping function to transform the data into a feature space, as PCA shows [30], and then use the linear classifier in this feature space.

An important property of linear classifier can be expressed in a dual form, which means that the assumption can be expressed as a linear combination of training points, so the decision rule f(x) (classification function) can be expressed as the inner product of the test point and training point,

$$f(x)=\sum_{i=1}^{n} \alpha_i y_i < \phi(x_i)\phi(x) > +b \quad (5)$$

In this formula, α_i is a positive derivative, obtained through learning. If there is a way to calculate the inner product <(directly$(x_i), \phi(x) >$ in feature space, as in the function of the original input point, then it is possible to combine the steps to create a non-linear classifier. In this way, it only needs to carry out the inner product operation in the high-dimensional, and this inner product operation can use the original space function to achieve, we do not even need to know the form of transformation, this direct calculation method called kernel function.

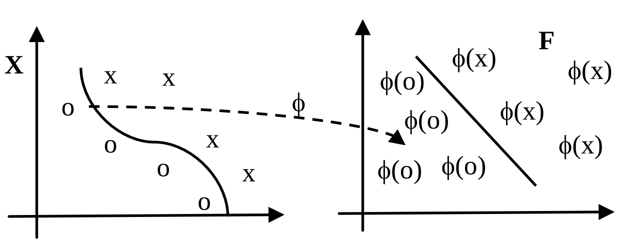


Fig. 4. Feature mapping of simple classification tasks

The commonly used kernel function consists of linear Kernel, polynomial Kernel, Gaussian Kernel sigmoid Kernel. According to the theory of generalization function, as long as one kind of kernel function satisfies the mercer condition, it corresponds to the inner product in a particular space. In this paper, the kernel function we adopted is equation (6), Gaussian Kernel,

$$K(x_1, x_2) = e^{-\frac{||x_1 - x_2||}{2\sigma^2}} \quad (6)$$

C. *The adaptive Search method of Sliding Window*

For the target tracking, there are similarity measures between the target and the model, the target, and the matching image. A good search strategy can not only eliminate false targets away from the target tracking and reduce the time complexity but also improve the robustness and real-time of the tracking algorithm. Common target search methods are: pyramid search, mean drift search [21], particle filter [22], Kalman filter [23], [24] and sliding window search [25]. The so-called sliding window search we used in this paper adopts multiple scales of the window to slide on the image and then generate multiple image blocks. It is practical in target tracking, target detection, and target classification.

A sliding window obtains Fig.4 (a) scanned the input image. The size of the window is different. In our experiment, the scale steps for multi-scale estimation adopt 0.8847, 0.9216, 0.96, 1, 1.05, 1.1025, 1.21275, 1.2734, etc. And then, we employed the window threshold; less than the threshold of the window will not be processed and saved. The size of the window above the threshold is reserved for the scales. We use the location of the image blocks and will continue to use it during subsequent tracking. Fig.4 (b) is the schematic diagram that can be calculated to get all the larger than the window threshold, where the window threshold selected in the experiment is based on our manually calibrated target. The scan mode is from top to bottom, from left to right, and the scanning step is one cell.

(a) (b)

Fig. 4. (a) is the current input image sequence, and the yellow solid line box is the selected target, (b) is a multi-scale sliding window satisfying the window threshold.

III. EXPERIMENTAL RESULTS AND ANALYSIS

The experiment selected six image sequences containing various challenging factors (TABLE I) from the literature [26] to verify the validity and robustness of our algorithm. For example, Illumination Variation(IV), Scale Variation(SV), Occlusion（OCC）, Deformation（DEF）, Motion Blur（MB）, In-Plane Rotation(IPR), Out-of-Plane Rotation (OPR), Out-of-View(OV) and Background Clutters(BC), and compared with present well-done TLD, KCF, CT, MIL. for qualitative and quantitative comparison.

TABLE I
THE CHALLENGES IN IMAGE SEQUENCES

Image Sequence	TOTAL FRAMES	Existing Challenges
Bird2	99	OCC,DEF,FB,FB,IPR,OPR
David	770	IV,SV,OCC,DEF,IPR,OPR
Girl	500	SV,OCC,DEF,IPR,OPR
Jump	122	SV,OCC,MB,IPR
Liquor	1714	IV,MB,FB,OPR,OV,BC
Singer2	366	IV,SV,DEF,IPR,BC

D. Evaluation Standard

In order to analyze the performance of the tracking quantitatively, the experiment selected the Centre Location Error (CLE) [27] and Success Rate (SR)as the evaluation criteria. Equation (7), equation (8), and equation (9) can be calculated as follows.

$$\mathrm{CLE} = \sqrt{(x_i - x_{i0})^2 + (y_i - y_{i0})^2} \qquad (7)$$

Where CLE expressed the Center Location Error of correct tracking（x_i, y_i）represents the central target position obtained by the i-th frame of video sequences through the tracking algorithm, and (x_{i0}, y_{i0}) is the target centre position obtained by manually marking i-th frame. Moreover, we take the CLE average of six testing videos as final CLE.

$$\mathrm{SR} = \frac{SN}{N} \qquad (8)$$

$$\mathrm{OR} = \frac{area(R_i \cap R_h)}{area(R_i \cup R_h)} \qquad (9)$$

Where SR means Success Rate, SN is the number of tracking successfully, N is total frames, and we take the SR average of six testing videos as final SR.

Besides, the score is called overlap rate, R_i is the area of rectangle window of the target and $R_h$is manually marked for the target rectangle window. If OR>0.5, SN+1.We test CLE and SR of several target tracking algorithms on six video sequences, partial results shown in TABLE II and TABLE III.

TABLE II
COMPARISON CLE RESULTS OF A PARTIAL SUCCESSFUL TRACKING ALGORITHM

CLE(pixel)	CT	KCF	OURS
Bird2	17.4	12.7	6.7
David	15.8	9.5,	8.2
Girl	21.6	10.1,	12.3
Jump	6.1	9.4	8.9
Liquor	14.7	12.3	6.9
Singer2	13.0	11.2	8.4
Mean CLE	14.75	10.87	8.6

TABLE III
COMPARISON SR RESULTS OF PARTIAL TRACKING ALGORITHM

SR(%)	CT	KCF	OURS
Bird2	59.6	50.5	100.0
David	38.3	59.9	100.0
Girl	25.6	100.0	93.8
Jump	32.0	20.5	20.4
Liquor	20.6	38.4	87.4
Singer2	27.0	5.20	15.9
Mean SR	33.85	45.75	69.58

E. Experimental Results and Analysis

This paper proposed a simple tracker based on kernel SVM classifier and HOG features which are implemented in Python [**Error! Bookmark not defined.**] **Error! Bookmark not defined.**]. Meanwhile, aiming to improve the speed of calculation, we perform Fast Fourier Transform on the input image to process it in the frequency domain and then obtain rotating invariant gradient information as its feature map to distinguish the target and background by the kernel SVM classifier. We experimented on Bird2, David, Girl, Jump, Liquor, Singer2. in different tracking algorithm framework, in which video can be downloaded from the visual tracker benchmark **Error! Bookmark not defined.**], and covers various complex and diverse tracking environments such as deformation, motion blur, fast motion, occlusion, scale variation, illumination variation.

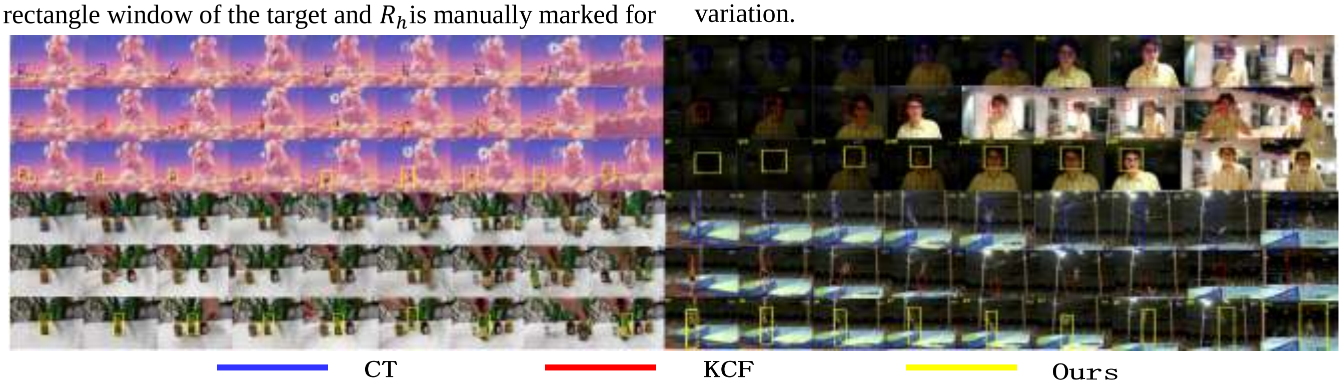


Fig. 5. Partial tracking results for different tracking algorithms on visual tracker benchmark Error! Bookmark not defined.**], blue rectangle, red rectangle, and yellow rectangle respectively tracking algorithm of CT, KCF, and our approaches.**

As shown in Fig. 5 and TABLE II, III, our tracking algorithm extracts the HOG features of the target that eliminates the influence of illumination variation, which has a higher success rate and is even up to 100% on sequence Bird2, David and Liquor. In the sequence Bird2, the target partially or fully obscured, and we use the sliding window to continuously deliver the image block to train SVM discriminator, effectively weaken the impact of occlusion. In sequence David, David walks in the corridor, the illumination and size change firmly. The classifier updating method in this paper ensures the diversity of the positive sample set and achieves stable tracking.

IV. CONCLUSION

We use the HOG features to extract and describe the target. Different from the previous method based on a single positive sample and multiple negative samples for classifier training. In this paper, the extracted HOG features are nucleated and then sent to the kernel SVM classifier to train. The characteristics of the candidate samples are discriminated against and updated continuously by the classifier to achieve feature selection and target tracking. However, there are still many problems in the target recognition technology based on kernel SVM. There is no effective method to find a suitable kernel function and its choice mostly depends on experience. Our future work should focus on the study of effective kernel function selection mechanisms and develop excellent algorithms to reduce the complexity of computing, looking for new ways to solve the challenge problem of online or offline video sequences.

REFERENCES

[1] Lei Z, Huimin L. “*Application of Projection Pursuit Model in residential building comprehensive performance assessment”*[C] Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference on. IEEE, 2010, 3: 851-855.

[2] Stone, Lawrence D., et al. Bayesian multiple target tracking. Artech House, 2013.

[3] Nägeli, Tobias, et al. "*Real-Time Motion Planning for Aerial Videography With Real-Time With Dynamic Obstacle Avoidance and Viewpoint Optimization.*" IEEE Robotics and Automation Letters 2.3 (2017): 1696-1703.

[4] Kakoty N M, Kaiborta M, Hazarika S M. “*Electromyographic grasp recognition for a five-fingered robotic hand*”[J]. IAES International Journal of Robotics and Automation, 2013, 2(1): 1.

[5] Jebelli A, Yagoub M C E, Dhillon B S. “*Design and Control of Underwater Robots with Rotating Thrusters”*[J]. IAES International Journal of Robotics and Automation (IJRA), 2016, 5(4): 284-294.

[6] Xu Y, Yu G, Wang Y, et al. “*A hybrid vehicle detection method based on viola-jones and HOG+ SVM from UAV images*”[J]. Sensors, 2016, 16(8): 1325.

[7] Dadi H S, Pillutla G K M, Makkena M L. “*Face Recognition and Human Tracking Using GMM, HOG and SVM in Surveillance Videos”*[J]. Annals of Data Science, 2017: 1-23.

[8] Ouerhani, Yousri, Ayman Alfalou, and Christian Brosseau. "*Road mark recognition using HOG-SVM and correlation*." Optics and Photonics for Information Processing XI. Vol. 10395. International Society for Optics and Photonics, 2017.

[9] Kim, Sanggi, and Dong Seog Han. "*Real Time Traffic Light Detection Algorithm Based on Color Map and Multilayer HOG-SVM.*" 방송공학회논문지 22.1 (2017): 62-69.

[10] Wei W, Jia Q. “*Weighted feature Gaussian kernel SVM for emotion recognition*”[J]. Computational intelligence and neuroscience, 2016, 2016: 11.

[11] Rao T C S, Sreenivasulu G. A “*Kernel SVM Classifier for Classification of Brain Tumors in Magnetic Resonance Images*”[J]. i-manager's Journal on Image Processing, 2016, 3(3): 34.

[12] Prasad S, Ramkumar B. “*Passive copy-move forgery detection using SIFT, HOG and SURF features*”[C]//Recent Trends in Electronics, Information & Communication Technology (RTEICT), IEEE International Conference on. IEEE, 2016: 706-710.

[13] Wang Q. “*Ore image edge detection using HOG-index dictionary learning approach*”[J]. The Journal of Engineering, 2017, 1(1).

[14] Lampert C H, Blaschko M B, Hofmann T. “*Beyond sliding windows: Object localization by efficient subwindow search*”[C]//Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008: 1-8.

[15] Wang W, Liu Y, Tang X, et al. “*A novel IRA-like codes construction method based on cyclic matrix permutation*”[C]//Image, Vision and Computing (ICIVC), 2017 2nd International Conference on. IEEE, 2017: 909-912.

[16] Doucet A, Godsill S, Andrieu C. “*On sequential Monte Carlo sampling methods for Bayesian filtering*”[J]. Statistics and computing, 2000, 10(3): 197-208.

[17] Siji Joseph, Arun Pradeep *"Object Tracking using HOG and SVM*", International Journal of Engineering Trends and Technology (IJETT), V48(6),321-325 June 2017.

[18] R. Rifkin, G. Yeo, and T. Poggio, “*Regularised least-squares classification,*” Nato Sci. Ser. Sub Ser. III, vol. 190, pp. 131–154, 2003.

[19] Weickert J, Grewenig S, Schroers C, et al. Cyclic schemes for PDE-based image analysis[J]. International Journal of Computer Vision, 2016, 118(3): 275-299.

[20] Siji Joseph, Arun Pradeep "*Object Tracking using HOG and SVM*", International Journal of Engineering Trends and Technology (IJETT), V48(6),321-325 June 2017.

[21] Cheng Y. “*Mean shift, mode seeking, and clustering*”[J]. IEEE transactions on pattern analysis and machine intelligence, 1995, 17(8): 790-799.

[22] Doucet A, Godsill S, Andrieu C. “*On sequential Monte Carlo sampling methods for Bayesian filtering*”[J]. Statistics and computing, 2000, 10(3): 197-208.

[23] Jazwinski A H. “*Stochastic processes and filtering theory*”[M]. Courier Corporation, 2007.

[24] Du W, Piater J. “*Tracking by cluster analysis of feature points using a mixture particle filter*”[C]Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference on. IEEE, 2005: 165-170.

[25] Orellana L H, Rodriguez-R L M, Konstantinidis K T. “*ROCker: accurate detection and quantification of target genes in short-read metagenomic data sets by modeling sliding-window bitscores*”[J]. Nucleic acids research, 2017, 45(3): e14-e14.

[26] Wu Y, Lim J, Yang M H. “*Online object tracking: A benchmark”*[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2013: 2411-2418.

[27] Everingham M, Zisserman A, Williams C K I, et al. “*The PASCAL visual object classes challenge 2007 (VOC2007) results*”[J]. 2007.

[28] R Sindhoori, *"Digital image processing. Multi feature face recognition in PSO -SVM"* SSRG International Journal of Electrical and Electronics Engineering 1.3 (2014): 1-6.

[29] Cortes C, Vapnik V. “*Support-vector networks*”[J]. Machine learning, 1995, 20(3): 273-297.

[30] Jolliffe I. “*Principal component analysis*”[M]. John Wiley & Sons, Ltd, 2002.