Robust Visual Tracking using Sparse Principle Component Analysis and Haar-like Features

Ashish Tiwari[1] Snehlata Raisagar[2] Kajal Tiwari[3]
[1]Assistant Professor [2]M.Tech Scholar
[1,2]Shri Shankaracharya Group of Institutions, Bhilai, India [3]Chhatrapati Shivaji Institute of Technology Durg, India

Abstract— Object tracking is an important problem with wide ranging applications. The aim of object tracking is to detect object and track its motion in the video. Issues of concern are to be able to map objects correctly between two frames, and to be able to track through occlusion. Factors such as pose variation, illumination change, occlusion, and motion blur make it difficult to develop effective and efficient appearance models for robust object tracking. A new tracking method combining on sparse representation, PCA and Haar-like features is proposed in this paper. Occlusion can be evaluated in PCA based analysis in Bayesian inference framework. Compressed features using sparse representation matching method is used to locate the target object if the level of occlusion satisfies inequality criteria. Most of the unimportant samples are removed before computing the compressed features; hence the computational complexity of proposed algorithm is maintained in optimum level. Experiments show that the proposed method with performs better against illumination change, occlusion and appearance variation, and outperforms several latest important tracking methods in terms of tracking performance.

Key words: PCA and Haar-like, Robust Visual Tracking

I. INTRODUCTION

Video surveillance is a famous research subject in computer vision that tries to perceive, distinguish and track objects/individuals over an arrangement of pictures and it additionally makes an endeavour to comprehend and portray object behaviour by supplanting the developing old standard system for watching cameras by human observer. Video sensors are exceptionally helpful to record the earth of a moving vehicle because of their high determination contrasted with size and cost. A diagram of vision-based ways to deal with obstruction discovery is introduced. The effective registering machines, the openness of high gauge and practical cameras and the extending necessity for robotized video examination has made a considerable measure of enthusiasm for object tracking. There are three key strides in video examination, moving object detection, tracking of such objects in each frame, and investigation of item tracks to perceive their conduct. In this way, the utilization of object tracking is the assignments of movement based recognition. Object detection includes finding objects in the frame of a video succession. Each following technique requires an object detection system either in each frame or when the article first shows up in the video. Following and observation applications require the division of objects from the scene to empower identification and tracking. This can be achieved by using techniques such as motion detection, or optical flow. Most optical flow techniques are either gradient based methods [1], [2], or block matching based methods [3]. Gradient based strategies have been favoured because of velocity and execution consideration. These techniques examine the adjustment in intensity and gradient (utilizing incomplete spatial and transient subsidiaries) to decide the optical flow. Procedure of finding an object or numerous objects after some time utilizing a camera is called object tracking. Automatic recognition, tracking, and counting of a inconstant number of objects are critical task for an extensive variety of home, business, and modern applications, for example, security, reconnaissance, administration of access focuses, urban arranging, movement control, and so on. Object identification and tracking are vital and testing errands in numerous computer vision applications, for example, surveillance, vehicle navigation and self-governing robot navigation. In any case, these applications were not even now having critical influence in consumer electronics. The fundamental reason is that they require solid prerequisites to accomplish tasteful working conditions, specific and costly equipment, complex establishments and setup methodology, and supervision of qualified labourers. Some works have concentrated on creating programmed location and following calculations that minimizes the need of supervision. They regularly utilize a moving object work that assesses every speculative object design with the arrangement of accessible location without to explicitly calculate their data affiliation. In this way, a significant reduction in computational expense is accomplished. Additionally, the probability capacity has been intended to represent boisterous, false and missing detection. The field of machine (computer) vision is worried with issues that include interfacing computers with their encompassing surroundings. One such issue, surveillance, has a target to screen a given domain and report the data about the watched action that is of huge interest. In this admiration, video surveillance more often than not uses electro-optical sensors (camcorders) to gather data from the earth. There are principle three stages in the arrangement of video surveillance:

- Locating fascinating moving objects in a scene.
- Following of found intrigued objects from every last frame.
- Analysis of trajectory path of object to estimate their behavior in next video frames.

Up to now, the video surveillance framework was for the most part utilized just for expansive scale or high secured organizations or military. In spite of the fact that, it will be valuable to control the expanded wrongdoing rate, particularly in innovative urban areas, took better safety measures to lessening criminal exercises in securing delicate spots, for example, airplane terminals, at the outskirts of the nation, secured government workplaces, and so on. Indeed, even individuals likewise by and by need to look for their customized security frameworks to control their homes or to make secure their important things. In a run of the mill surveillance framework, these camcorders are mounted in altered positions or on dish tilt gadgets and transmit video streams to a specific area, called observing room. At that point, they got video streams are checked on showcases and followed by human administrators. Be that as it may, the human administrators may confront numerous issues, while they are checking these sensors. One issue is because of the way that the administrator must explore through the cameras, as the suspicious object moves between the restricted field of perspective of cameras and ought not miss whatever other object while taking it. Therefore, checking turns out to be increasingly testing, as the quantity of sensors in such a surveillance system increments. In this way, surveillance frameworks must be robotized to enhance the execution and dispense with such administrator mistakes.

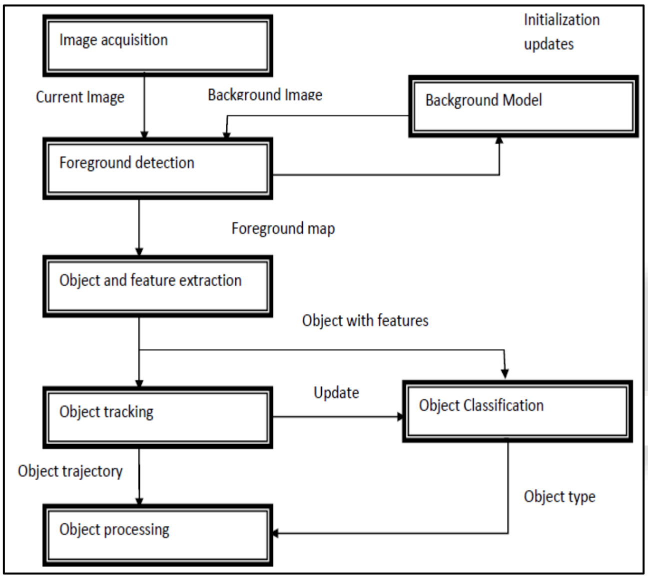

Figure 1 Video analysis block diagram

Figure 1portray the framework square chart of Moving objects following. In a perfect world, a robotized surveillance framework ought to just require the objectives of an application, in which constant translation and vigor is required. At that point, the test is give strong and ongoing performing surveillance frameworks at a moderate cost. With the reduction in expenses of equipment for detecting and figuring, and the expansion in the processor speeds, surveillance frameworks have turned out to be industrially accessible, and they are presently connected to various distinctive applications, for example, activity observing, airplane terminal and bank security, and so forth. Be that as it may, machine vision calculations (particularly for single camera) are still seriously influenced by numerous inadequacies, similar to impediments, shadows, climate conditions, and so forth. As these costs diminish nearly every day, multi-camera arranges that use 3D data are turning out to be more accessible. In spite of the fact that, the utilization of different cameras prompts better treatment of these issues, contrasted with a solitary camera, shockingly, multi-camera surveillance is still not a definitive arrangement yet. There are some testing issues inside the surveillance calculations, for example, foundation demonstrating, highlight extraction, following, impediment taking care of and occasion acknowledgment.

In addition, machine vision calculations are still not sufficiently strong to handle completely mechanized frameworks and numerous exploration concentrates on such upgrades are as yet being finished. This work concentrates on building up a structure to recognize moving objects and create solid tracks from surveillance video. The issue is the greater part of the current calculations takes a shot at the dark scale video. Be that as it may, subsequent to changing over the RGB video edges to dark at the season of transformation, data misfortune happens. The principle issue comes when foundation and the frontal area both have roughly same dim qualities. At that point it is troublesome for the calculation to discover which pixel is forefront pixel and which one foundation pixel. Once in a while two diverse hues, for example, dim blue and dull violet, shading when changed over to dim scale, their dim qualities will come exceptionally close to each other, it can't be separated what esteem originates from dim blue and which originates from dim violet. Be that as it may, if shading pictures are taken then the foundation and frontal area shading can be effortlessly separated. So without losing the shading data this adjusted foundation model will work straightforwardly on the shading edges of the video.

The problem and difficulty of object tracking depend on several factors, such as the amount of prior knowledge about the target object and the number and type of parameters being tracked (e.g. location, scale, detailed contour). Although there has been some success with building trackers for specific object classes (e.g. faces [4], humans [5], mice [6], rigid objects [7]), tracking generic objects has remained challenging because an object can drastically change appearance when deforming, rotating out of plane, or when the illumination of the scene changes.

A typical tracking system consists of three components: (1) an appearance model, which can evaluate the likelihood that the object of interest is at some particular location; (2) a motion model, which relates the locations of the object over time; and (3) a search strategy for finding finding the most likely location in the current frame. The contributions of this paper deal with the first of these three components; we refer the reader to [8] for a thorough review of the other components. Although many tracking methods employ static appearance models that are either defined manually or trained using only the first frame [5], [7], [9], [10], [11], [12], [12], these methods are often unable to cope with significant appearance changes. These challenges are particularly difficult when there is limited a priori knowledge about the object of interest. In this scenario, it has been shown that an adaptive appearance model, which evolves during the tracking process as the appearance of the object changes, is the key to good performance [13], [14], [15]. Training adaptive appearance models, however, is itself a difficult task with many questions yet to be answered. Such models often involve many parameters that must be tuned to get good performance (e.g. “forgetting factors” that control how fast the appearance model can change), and can suffer from drift problems when an object undergoes partial occlusion.

The proposed work focus on the problem of tracking an arbitrary object with no prior knowledge other than its location in the first video frame (sometimes referred to as "model-free" tracking). Goal of work is to develop a more robust way of updating an adaptive appearance model; we would like our system to be able to handle partial occlusions without significant drift, and for it to work well with minimal parameter tuning. To do this, we turn to a discriminative learning paradigm with sparse representation and principle component analysis.

II. Related Research

Babenko B, Yang M-H, Belongie S (2011) presented a paper titled "Robust object tracking with online multiple instance learning". In this work, they dealt with the tracking of single object in a sequence of frames either from a live camera or a previously saved video. A moving object is detected frame-by-frame with high accuracy and efficiency using Median approximation technique. As soon as the object has been detected, the same is tracked by kalman filter estimation technique along with a more accurate Template Matching algorithm. The templates are dynamically generated for this purpose. This guarantees any change in object pose which does not be hindered from tracking procedure. The system is capable of handling entry and exit of an object. Such a tracking scheme is cost effective and it can be used as an automated video conferencing system and also has application as a surveillance tool. Performance can be improved on tracking multiple objects at the same time on improving tracker accuracy during camera motion. These algorithms can be further extended for the use in real-time applications and object classifications [13].

Jia X, Lu H, Yang M-H (2012). "Visual tracking via adaptive structural local sparse appearance model". In this work they developed a simple yet robust tracking method based on the structural local sparse appearance model. This representation exploits both partial information and spatial information of the target based on a novel alignment-pooling method. The similarity obtained by pooling across the local patches helps not only locate the target more accurately but also handle occlusion. In addition, we employ a template update strategy which combines incremental subspace learning and sparse representation. This strategy adapts the template to the appearance change of the target with less possibility of drifting and reduces the influence of the occluded target template as well. Both qualitative and quantitative evaluations on challenging benchmark image sequences demonstrate that the proposed tracking algorithm performs favorably against several state-of-the-art methods [16].

Tianzhu Zhang, Bernard Ghanem, Si Liu, Narendra Ahuja(2012) presented a paper titled "Robust Visual Tracking via Multi-Task Sparse Learning". In This paper they formulate object tracking in a particle filter framework as a multi-task sparse learning problem, which we denote as Multi-Task Tracking (MTT). Since wemodel particles as linear combinations of dictionary templates that are updated dynamically, learning the representation of each particle is considered a single task in MTT. By employing popular sparsity-inducing `p,q mixed norms ($p \in \{2, \infty\}$ and $q = 1$), they regularize the representation problem to enforce joint sparsity and learn the particle representations together. As compared to previous methods that handle particles independently, our results demonstrate that mining the interdependencies between particles improves tracking performance and overall computational complexity. Interestingly, we show that the popular L1 tracker [15] is a special case of our MTT formulation (denoted as the L11 tracker) when $p = q = 1$. The learning problem can be efficiently solved using an Accelerated Proximal Gradient (APG) method that yields a sequence of closed form updates. As such, MTT is computationally attractive. They test our proposed approach on challenging sequences involving heavy occlusion, drastic illumination changes, and large pose variations. Experimental results show that MTT methods consistently outperform state-of-the-art trackers[17].

Hanxi Li, Chunhua Shen, and Qinfeng Shi(2012) presented a paper titled "Real-time Visual Tracking Using Sparse Representation". In this work they presented a sparse dictionary learning method, specifically tuned to solve the tracking problem. Recently, sparse representation has drawn much attention because of its genuineness and strong mathematical background. In this paper they presented an online method for dictionary learning which is desirable for problems such as tracking. Online learning methods are preferable because the whole data are not available at the current time. The presented method tries to use the advantages of the generative and discriminative models to achieve better performance. The experimental results show this method can overcome many tracking challenges [18].

Wang D, Lu H and Yang M-H (2013) presented a paper titled "Online object tracking with sparse prototypes". In this paper, they propose a novel online object tracking algorithm with sparse prototypes, which exploits both classic principal component analysis (PCA) algorithms with recent sparse representation schemes for learning effective appearance models. They introduce l1 regularization into the PCA reconstruction, and develop a novel algorithm to represent an object by sparse prototypes that account explicitly for data and noise. For tracking, objects are represented by the sparse prototypes learned online with update. In order to reduce tracking drift, we present a method that takes occlusion and motion blur into account rather than simply includes image observations for model update. Both qualitative and quantitative evaluations on challenging image sequences demonstrate that the proposed tracking algorithm performs favorably against several state-of-the-art methods [19].

Xiao ZY, Lu HC and Wang D(2014) presented a paper titled "L2-RLS based object tracking". In this paper, they presented a robust and fast tracking algorithm in which object tracking is achieved by solving $\ell2$-regularized least square ($\ell2$-RLS) problems in a Bayesian inference framework. First, the changing appearance of the tracked target is modeled with PCA basis vectors and square templates, which make the tracker, not only exploit the strength of subspace representation but also explicitly take partial occlusion into consideration. They can together represent both the intact and corrupted objects well. Second, They adopt the ℓ2-regularized least square method to solve the proposed representation model. Compared with the complex ℓ1-based algorithm, it provides a very fast performance without the loss of accuracy in handling the tracking problem. In addition, a novel likelihood function and a refined update scheme further help to improve the robustness of proposed tracker [20].

Review of the above papers show that target can be modeled as PCA basis vector to improve target appearance and take partial occlusion into consideration. Sparse representation has been repetitively used to create a low dimensional space to reduce the computational complexity. Harr-like features have been used to represent target. These features have been proved robust in appearance change but sensitive to large occlusion. A better an improved method can be developed by considering pros of above mentioned methods.

III. Problem Statement

The object tracking is a challenging problem that has received an enormous amount of attention by many researchers [13, 14, 15, and 1] and hence many algorithms have been presented. Numerous of them are generative algorithms and discriminative algorithms.

Generative tracking algorithms typically learn a model to represent the target object and then use it to search for the image region with minimal reconstruction error. Black et al. [6] learn an off-line subspace model to represent the object of interest for tracking. The IVT method utilizes [7] an incremental subspace model to adapt appearance changes. Recently, sparse representation has been used in the l1-tracker where an object is modeled by a sparse linear combination of target and trivial templates [1]. However, the computational complexity of this tracker is rather high, thereby limiting its applications in real-time scenarios. Li et al. [2] further extend the l1-tracker by using the orthogonal matching pursuit algorithm for solving the optimization problems efficiently. Despite much demonstrated success of these online generative tracking algorithms, several problems remain to be solved. First, numerous training samples cropped from consecutive frames are required in order to learn an appearance model online. Since there are only a few samples at the outset, most tracking algorithms often assume that the target appearance does not change much during this period. However, if the appearance of the target changes significantly at the beginning, the drift problem is likely to occur. Second, when multiple samples are drawn at the current target location, it is likely to cause drift as the appearance model needs to adapt to these potentially mis-aligned examples [5]. Third, these generative algorithms do not use the background information which is likely to improve tracking stability and accuracy.

Discriminative algorithms pose the tracking problem as a binary classification task in order to find the decision boundary for separating the target object from the background. Avidan [21] extends the optical flow approach with a support vector machine descriptor for object tracking. Collins et al. [22] demonstrate that the most discriminative features can be learned online to separate the target object from the background.

Grabner et al. [23] propose an online boosting algorithm to select features for tracking. However, these trackers [21–23] only use one positive sample (i.e., the current tracker location) and a few negative samples when updating the descriptor. As the appearance model is updated with noisy and potentially misaligned examples, this often leads to the tracking drift problem. Grabner et al. [8] propose an online semi-supervised boosting method to alleviate the drift problem in which only the samples in the first frame are labeled and all the other samples are unlabeled.

The generative tracking techniques generally just concentrate on the target itself, so the background in the target area will cause the tracking drift. In spite of the fact that it can adjust to changes in posture and illumination, it is inclined to drifting because of the interference from the background. The discriminative tracking strategies have collaborated with background data, they are as yet hard to utilize the unlabeled information and update the classifiers amid tracking. Background data is essential for a tracker with a decent capacity to recognize foreground. Object tracking can be improved by incorporating advantages of different methods. A robust tracking algorithm can be implemented by using Haar-like features in compressed domain and PCA based match algorithm in Bayesian inference framework.

IV. Proposed Method

A new method is proposed which combines the Haar-like features and PCA optimization in Bayesian inference framework. The target image is represented as:

$$Y = Bc_b + e$$

Where B is the member of matrix of PCA basis vector represents feature vector dimension and represents number of PCA basis vectors represents error and generated by square templates generated offline.

A cost function is defined as:

$$c^* = \arg \min_{c} f(c)$$

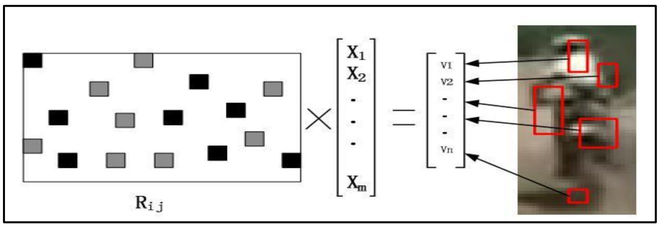

Fig. 2: Graphical representation of compressing a high-dimensional vector x to a low dimensional vector v. In the matrix R, dark, gray and white rectangles represent negative, positive, and zero entries, respectively.

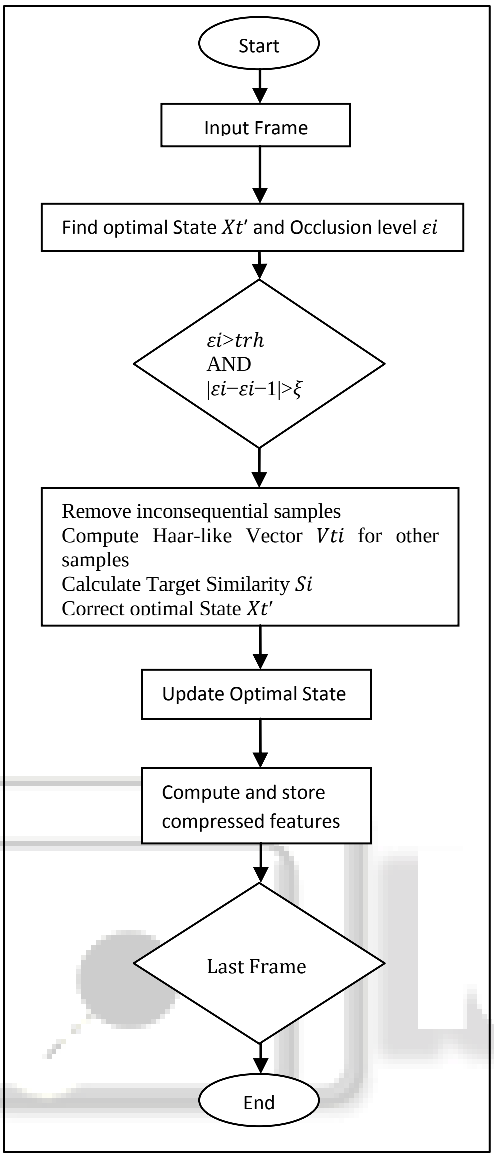

Fig. 3: Proposed Algorithm

Object tracking as a Bayesian inference task in a Markov model with hidden state variables. Hidden state variable is represented as and can be estimated recursively using image observation vector $Y_t = \{Y_1, Y_2 \ldots. Y_t\}$ as:

$$p(X_t|Y_t) = p(Y_t|X_t) \int p(X_t|X_{t-1}) p(X_{t-1}|Y_{t-1}) dX_{t-1}$$

Proposed algorithm uses the sparse decomposition method to reduce the matrix dimension. Final tracking framework is proposed to avoid problems in Haar like feature and sparse representation based tracking like L1 and L2 trackers.

The tracking method is represented in flow chart in figure 2.

V. RESULTS

We performed our experiments on 5 publicly available video sequences, as well as 1 of our own. For all sequences we labeled the ground truth center of the object for every 5 frames, and interpolated the location in the other frames. Figure 4 shows results of object tracking using one publically available database and one database of our own.

VI. CONCLUSION

Objective evaluations are performed in the tracking outputs. Proposed method also quantitatively compared with different model on different image Sequences. The test algorithms include: incremental visual tracker (IVT) variance ratio tracker (VRT) online boosting tracker (BoostT), $_1$ tracker (L1T) and proposed algorithm Project work included the calculation of success rate and proposed method shows better performance.

Object tracking has very important place in computer vision and has many practical applications. The problem and its difficulty depend on several factors, such as the amount of prior knowledge about the target object and the number and type of parameters being tracked (e.g. location, scale, detailed contour).

(a)

(b)

Fig. 4: some is tracking outputs. (a) Tracking of Ant (b) Tracking of Toy

Although there has been some success with building trackers for specific object classes (e.g. faces, humans, mice, rigid objects), tracking generic objects has remained challenging because an object can drastically change appearance when deforming, rotating out of plane, or when the illumination of the scene changes.

In this paper, a robust online object tracking algorithm is presented, in which object tracking is achieved by combining the Sparse representation, Principle Component Analysis and a novel feature matching algorithm based on compressed multi-scale Haar-like features in a Bayesian inference framework. Haar like features used to avoid wrong tracking in case of varying appearance and avoid drift. Both qualitative and quantitative evaluations demonstrate that the proposed tracking algorithm performs better when the target object undergoes pose variation or rotation. Furthermore, by removing most of the insignificant samples before computing the compressed features, it reduces computational complexity. A comparison is presented between the proposed tracker with four other trackers on different challenging sequences to validate robustness. Overall, our tracker achieves better performance than other state-of-the-art algorithms in terms of accuracy.

REFERENCES

[1] Idaku Ishii, Taku Taniguchi, Kenkichi Yamamoto, and Takeshi Takaki,"High-Frame-Rate Optical Flow System" IEEE Transactions On Circuits And Systems For Video Technology, VOL. 22, NO. 1, Pp. 105-112, January 2012, pp 105-112.
[2] D. Hari Hara Santosh, P. Venkatesh, P. Poornesh, L. Narayana Rao and N. Arun Kumar, "Tracking Multiple Moving Objects Using Gaussian Mixture Model" International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-2, May 2013.
[3] Xiaofei Wang, Xiaomin Yang, Xiaohai Hea, Qizhi Teng and Mingliang Gao, "A high accuracy flow segmentation method in crowded scenes based on streakline" Optik 125 (2014) 924– 929.
[4] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in CVPR, vol. 1, 2001, pp. 511–518.
[5] Kalal, Z., Matas, J., Mikolajczyk, K.: P-n learning: bootstrapping binary classifier by structural constraints. In: CVPR, pp. 49–56 (2010).
[6] Candes, E., Tao, T.: Decoding by linear programing. IEEE Trans. Inform. Theory 51,4203–4215 (2005).
[7] Mei, X., Ling, H.: Robust visual tracking and vehicle classification via sparse representation. PAMI 33, 2259–2272 (2011).
[8] Li, H., Shen, C., Shi, Q.: Real-time visual tracking using compressive sensing. In: CVPR, pp. 1305–1312 (2011).
[9] Avidan S (2004) Support vector tracking. IEEE Trans Pattern Anal Mach Intell 26(8):1064–1072.
[10] Babenko B, Yang M-H, Belongie S (2011) Robust object tracking with online multiple instance learning. IEEE Trans Pattern Anal Mach Intell 33(8):1619–1632.
[11] Donoho, D.: Compressed sensing. IEEE Trans. Inform. Theory 52, 1289–1306 (2006).
[12] Candes, E., Tao, T.: Near optimal signal recovery from random projections and universal encoding strategies. IEEE Trans. Inform. Theory 52, 5406–5425 (2006).
[13] Ross, D., Lim, J., Lin, R., Yang, M.-H.: Incremental learning for robust visual tracking. IJCV 77, 125–141 (2008).
[14] Grabner, H., Leistner, C., Bischof, H.: Semi-supervised On-Line Boosting for Robust Tracking. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 234–247. Springer, Heidelberg (2008).
[15] P. Doll´ar, Z. Tu, H. Tao, and S. Belongie, “Feature mining for image classification,” in CVPR, 2007.
[16] Jia X, Lu H, Yang M-H (2012) Visual tracking via adaptive structural local sparse appearance model. In: 2012 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 1822–1829.
[17] Tianzhu Zhang, Bernard Ghanem, Si Liu, Narendra Ahuja "Robust Visual Tracking via Multi-Task Sparse Learning" IEEE 2012 PP 2044-2049.
[18] Hanxi Li, Chunhua Shen, and Qinfeng Shi "Real-time Visual Tracking Using Sparse Representation" CVPR IEEE 2012 pp 304-309.
[19] Wang D, Lu H, Yang M-H (2013) Online object tracking with sparse prototypes. IEEE Trans Image Process 22(1):314–325.
[20] Xiao ZY, Lu HC, Wang D, “L2-RLS based object tracking,” IEEE Trans Circ Syst Video Technol 24(8): 1301–1309 (2014).
[21] Grabner H, Bischof H (2006) On-line boosting and vision. In: 2006 IEEE computer society conference on computer vision and pattern recognition, vol 1. IEEE, pp 260– 267.
[22] Kwon J, Lee KM (2010) Visual tracking decomposition. In: 2010 IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp 1269–1276.
[23] Mei X, Ling H (2009) Robust visual tracking using Babenko minimization. In: 2009 IEEE 12th international conference on computer vision. IEEE, pp 1436–1443.