



Object Tracking in Occlusion and Contrast Conditions using Patch-wise Sparse Method

Yashesh Joshi^{1*} Hiren Mewada²

¹*Department of Electronics & Communication, Charusat University, Changa, Gujarat, India*

²*Electrical Engineering Department, Prince Mohammad Bin Fahd University, Kingdom of Saudi Arabia*

* Corresponding author's Email: yashesh_joshi@yahoo.co.in

Abstract: Object detection and tracking is one of the active areas of studies because of variation in object location, pose, illumination, scale and occlusion. Singular object tracking (SOT) is an important task in human-computer interaction, anonymous tracking in the robotic application and security analysis. This paper proposes single visual tracking in a complex environment. We propose SOT using patch sparse formulation that is robust to pose, scale, long-term occlusion and low- and high- luminance conditions. A sparse dictionary using the patches from the previous frames is used to track object in the current frame under such a clutter environment. Contrast modeling, patch resizing and likelihood measurement in the proposed sparse framework allows the selection of key patches to tackle long-term occlusion and extreme illumination variation. We evaluate the proposed method on different challenging datasets involving Illumination variation, heavy occlusion and pose variation. The result analysis indicates that the proposed patch wise joint sparse occlusion and contrast (PJS-OC) tracking method outperforms other methods with the maximum precision rate is 99%. Also for the sampling value of 0.5, the proposed tracker shows excellent results compared with the other different sampling values. The proposed sparse formulation provides excellent performance in various quantitative parameters, achieving better precision and speed. The comparative analysis with state-of-art methods using the benchmark database shows the outstanding performance of the proposed method.

Keywords: Object tracking, Sparse representation, Dictionary, Likelihood.

1. Introduction

Object tracking is a procedure to locate the object in the sequence of the frames and it consists of four processing steps, including object initialization in the reference frame, background removal, predict the motion and localizing the object. Object initialization in the reference frame refers to annotating the location of the object to tracked using rectangular bounding box or ellipse, centroid, skeleton, object shape or silhouette. Single object tracking is a well-studied problem within the area of image processing. The ability to track a single object has improved drastically during the last decades, however, it is still considered a complex problem to solve. The importance of object tracking is utilized in various applications such as video surveillance, Biomedical,

activity recognition, and traffic monitoring, human-computer interaction, robot servant [1] etc.

Object tracking can be categorized based on critical factors used in the scheme. The first category is based on object representation scheme, i.e. object's features like color [2], texture [3], statistical features [4] etc. Second category is referred as generative-discriminative scheme. The generative scheme uses image observation approach like template matching [5] and discriminative approach [6] distinguish an object from the background. The third category is online modeling. Sparse modeling and dictionary learning adopt the appearance change of the object and the background. Dictionary approaches mainly reduce to an accession problem. Many optical flow-related approaches, Sum of squared difference (SSD) algorithm [7] is used for object tracking. Also, the Mean shift is an unconventional deterministic approach for visual tracking.

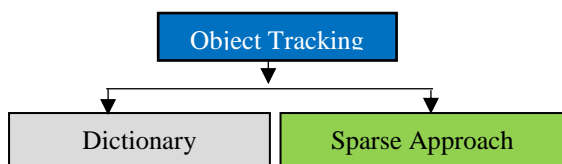


Figure. 1 Online object tracking approach

Sparse approaches mainly reduce to estimation problem. Such problem is estimating the state of the target. In the past Kalman filter [8] oriented approach was used for tracking. But now a days Sequential Monte Carlo algorithm is successfully used for object tracking because of its non-linearity and non-Gaussian cases [9].

However, slight inaccuracy in model results in degradation of the results gradually. To address such problems, different algorithms and different feature oriented methods have been investigated in [10] including outlier rejection [11], supervised boosting [5], key patch representation [6] and the hierarchical sparse formation of the extracted features [11]. All these methods have not considered the correlation of visual appearance information and neighboring pixel information resulting decline of tracking accuracy in long term occlusion and high illumination variation.

In this paper, we propose a sparse based computationally efficient and robust algorithm. The proposed algorithm identifies and uses the key patches from the past frames to track object under the occlusion. The contrast modeling and likelihood of the pixels help to determine the key patches from the image. Further scaling operation helps to increase the frame per second (FPS) rate.

We compare our results with seven well-known trackers, namely OAB, MIL, Frag, IVT, APG, MTT and PJSS. We also report of two recent sparse trackers denoted as MTT and PJSS. Also we evaluate our method with different sampling values. The proposed method is denoted by patch wise joint sparse occlusion and contrast (PJS-OC) tracking.

The summarization of the proposed work is as follows. First, sparse codes of neighborhood image patches are used in an object to represent the appearance model. For the sparse approach, we propose the joint sparse representation method to improve the performance of object under heavy occlusion and contrast situation. Second, the sampling rate has been decreased gradually to achieve a better frame rate. Also the precision is increased by increase the iteration of patches. The result analysis indicates that the proposed sparse method outperforms other state of art methods. The paper is organized as follows: Section 2 presents the work related to SOT. Then a mathematical representation of a sparse based tracking method is

explained in section 3. In section 4, the proposed method to handle the occlusion and uneven contrast in video tracking is explained. Finally, a comparative analysis of the proposed method and corresponding conclusion and future work are discussed in section 5 and 6 respectively.

2. Related works

Many objects tracking related work have been reported in computer vision application. Early works, Sum of square difference (SSD) algorithm with gradient decent was most commonly used for object tracking to minimize SSD [12]. A Mean shift algorithm with Kalman filter have been successfully tracked object from a video sequence. Based on Mean shift algorithm, Bhattacharya coefficient is used in covariance matrix to update Kalman filter error. Also, it determines whether an object is occluded or not [13].

Now a days sparse related approach widely used in computer vision because of its discriminative ability. The general sparse representation base approach is mainly used in computer vision such as action recognition [14], visual tracking [15], Face detection [16] and Super resolution [17]. Sparse representation methods have been designed in the computer vision for object tracking. Sparse representation requires a few basis vectors for a linear combination. W Hu. et al. [18] proposed a multi feature sparse algorithm for occlusion handling during object tracking. Such algorithm automatically targets on an occluded object by using trivial templates. H Yang et al. [19] used the detection confidence to form the templates set and identify the targets. It is mainly utilized for sparse representation of the appearance model.

Babenko et al. [20] proposed to create a series of templates around the estimation of target and multiple instances learning framework for tracking. MIL tracker handles the various appearance and partial occlusion challenges to some extent. This approach could perform well for short term occlusions. But clearly fails in a challenging environment like heavy occlusions. R. Venkatesh Babu et. al. [21] proposed a new approach with the combination of color based mean shift and sum of squared differences (SSD) tracker. The mean shift tracker beats the fast model change in the SSD tracker. The drawback of MS tracker is to unable of handling the large movement of objects.

Ashwini et al. [22] proposed a new model based on the object and surrounding information to create new dictionary of overlapping patches. Candidate patches are sparsely represented in the new dictionary

by ℓ_1 minimization. The drawback of this approach is to fail to track under severe occlusion Ruchandani et al. [23] represented a new template update strategy to update the patches of the object. Such approach performed better than the state of art algorithms. But this method fails in occlusion and Illumination variation condition. To minimize descriptors obtained from the patches, a principle component analysis based dimension reduction in the sparse domain was presented in [11,24].

In this work, we propose a high speed and efficient algorithm based on patch wise joint sparse representation approach. The main purpose of our method is to represent the patch wise sparse approach and to determine the optimum result for tracking.

3. Methodology

3.1 Sparse representation

A sparse representation is mostly used for various computer vision applications such as Face detection, object tracking, image classification, etc. Fig. 2 shows the flow chart of SOT using sparse approach. A candidate dictionary y and the vector elements of the dictionary D can be represented as given below,

$$y = Db \tag{1}$$

Where y is candidate vector, $D = [d_1, d_2, \dots, d_n] \in \mathbb{R}^{l \times n}$, $b = [b_1, b_2, \dots, b_n]^T$ is the coefficient vector of dictionary D .

From Eq. (1) if $l \ll n$ then there is no unique solution for b . So such a model can be solved by l_1 minimization least square problem,

$$\min \|Db - y\|_2^2 + \beta \|b\|_1 \tag{2}$$

Where b is a sparse coefficient vector corresponding to image vector y and β is a scalar positive constant that controls the sparsity of b . also $\|\cdot\|_0$, $\|\cdot\|_1$ and $\|\cdot\|_2$ are l_0 , l_1 and l_2 norms respectively.

3.2 Dictionary creation

Dictionary training is a successful approach to dictionary design. The dictionary has been mainly affected in sparse representation models and its algorithms. For template related tracking in l_1 minimization approach, tracking is obtained by comparing the candidate template set to object template sets through sparse representation [15].

Two dictionaries are used in our approach – One is static and the other is dynamic. The static dictionary is originated with the object selected in the

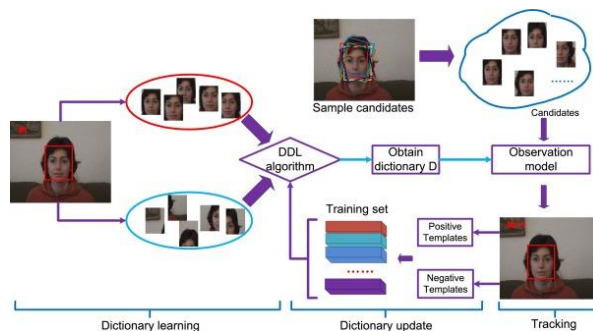


Figure. 2 Flow chart of object tracking [23]

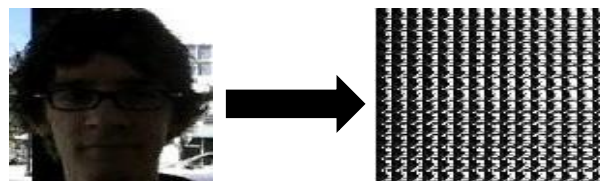


Figure. 3 Patch dictionary of an object [21]

first frame and the dynamic dictionary is updated based on the appearance changes during object tracking. The original dictionary size and the object size create fragment size. Suppose the object size is $P \times Q$ and dictionary size is $c \times d$ then the patch size will be $(P-c+1) \times (Q-d+1)$. Fig. 3 shows dictionary generated by overlapping fragments of the object in raster scan order.

Each patch is a set of $D_z = [d_{z1}, d_{z2}, \dots, d_{zn2k}] \in \mathbb{R}^{l \times n}$. Also template is vectored into a single column vector $D_{zj} \in \mathbb{R}^l$. The size of object array is same as dictionary array, but only a sufficient number of candidate fragment estimate the motion of the object and track such an object. In general, candidate patch is described as a linear combination of the dictionary patches. For the k^{th} candidate Eq. (1) can be rewritten as

$$y_k = [D_1, D_2] [B_{1k}^T B_{2k}^T]^T \tag{3}$$

Where D_1 and D_2 are represented as static and dynamic dictionaries respectively. Eq. (3) can be solved by sparse representation using l_1 minimization.

3.3 Motion estimation

Gaussian mixture model comprises of various Gaussian component functions. Such algorithm performs well for extraction process of background. Pixel plays an important role in Gaussian distribution. Each pixel in the current frame is compared with GMM model. Motion estimation is mainly depending on weighted elements. Same value of pixels with lower deviation and weighted element is considered as motion estimation. The lower value of weight element is considered as foreground in an image [25].

A blob is nothing but a region of connected pixel. Blob analysis is used to calculate area and perimeter. The function of blob analysis is to detect the cluster of pixels under moving object condition. It can also be used to find blob size, number, location, and blob.

4. Proposed system

The purpose of the proposed method is to track the target object based on the set of candidates created by means of particles in continuous frame. To track the target object, the first step is to find the location of target candidates that is maximum possibly generated by the appearance model. Based on target observed in previous frame the proposed method composed the dictionary patches which is used for target appearance in the current frame. The target candidates have a minimum error over its patches. The appearance model updated in next frame only when high quality candidate is determined in the target object. During the process of updating dictionary, the occluded image patches remove from consideration patches. Occlusion generated by the scene objects So occlusion handling is a complex problem within a challenging environment. Contrast indicates the range of brightness value present on an image. Contrast is the difference in illumination which makes object different from real images. The proposed method flow diagram is shown in Fig. 4.

4.1 Appearance dictionary

The target appearance is a linear combination of a few elements (atoms) available from the dictionary. Based on updating dictionary, the best candidate can be found by the target appears in the dictionary. In different words, the dictionary atoms are linearly spanned on low dimension subspace to track target object. An appearance model of linear subspace can handle the challenging attributes of illumination and pose variation. Fig. 5 shows that the pose varying related face image dictionary is able to describe model of pose changes. A patch wise appearance model with dictionary and combination of best selected candidate patches is shown in Fig. 5. The target appearance model can be expressed as

$$Z = \Phi W + V + \varepsilon \tag{4}$$

Where $z \in \mathbb{R}^{M \times 1}$ is a pixel intensity vector of the target, $\Phi \in \mathbb{R}^{M \times n}$ is a set of orthogonal basis vector of the space spanned by the target template, $W \in \mathbb{R}^{n \times 1}$ vector coefficient generated by representing z , $V \in \mathbb{R}^{M \times 1}$ is a sparse vector, $\varepsilon \in \mathbb{R}^{M \times 1}$ represent noise term.

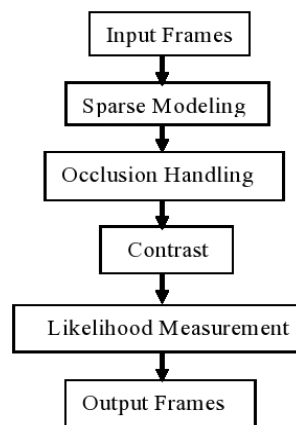


Figure. 4 Block diagram of proposed visual tracking model



Figure. 5 Appearance model

There are mainly two problems with proposed appearance model, first is feature extraction from appearance model and the second is a patch wise extraction from the target. The feature oriented sparse trackers are SIFT, gray scale, covariance matrix and block division based tracker. Most element extraction strategies include just straight (or around direct) tasks, and the gray scale features of linear transformation have no effect on face recognition by applying sparse representation. Then again, highlight extraction strategies, for example, SIFT or Haar are followed by the loss of information, which decrease the discriminative capacity of the tracker. We use grayscale features because of low computational complexity. The target representation is also a major concern in feature based tracking. The dictionary is a combination of the vectored images of target objects in previous frames. Wright et. al [25] suggested that the performance of face recognition decreases under heavy occlusion condition. Likewise, the effect of an ongoing dataset with existing method shows that trackers with a patch wise see most existing approaches are heartier occlusion than strategies with an all other method perspective. The benefits of the proposed target appearance model are twofold: increase the occlusion handling performance during tracking and improve the speed with change in the size of dictionary.

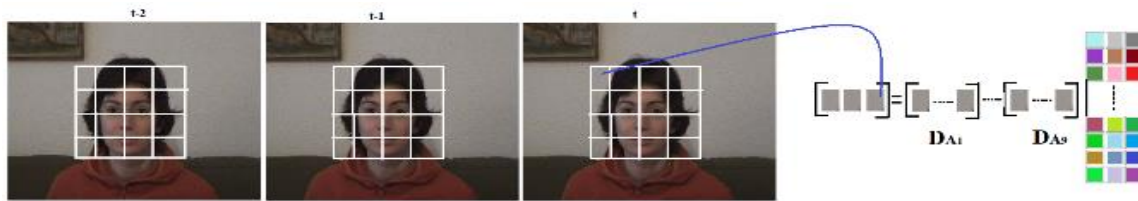


Figure. 6 Sparse model

4.2 Sparse representation

Taking into account that the object appearance changes easily, it is conceivable to accept that the object target over a couple of sequential frames, have a place with a similar subspace. The model of sparse representation can be used as a target appearance of an object and the likelihood of a better candidate available in the image. The appearance model is considered as a number (N) of training samples and dictionary of the target appearance in given sample [26].

In Fig. 6, sparse representation can be used as a target appearance of an object and the likelihood of a better candidate available in the image. The appearance model is considered as a number (N) of training samples and dictionary of the target appearance in given sample [26].

The sparse representation of image block b can be represented in terms of dictionary

$$\min \|b - Da\|_F^2 + \beta \|a\|_1 \quad (5)$$

Where a is a sparse coefficient, b is a image block, β is a scalar positive coefficient. Fig. 6 represents target object dictionary have been distributed in many patches of the targeted object. At the point when an object is somewhat moved away from the target, the appearance of its patches will generously increment. As delineated in Figure over, this is on the grounds that the moved object is never again lined up with the target, subsequently the inadequate description of any of its patches is more averse to utilize particles from its comparing patch format. Since the remaking mistake is processed utilizing the coefficients of the comparing patch format, the log probability of a skewed up-and-comer will be very little. The appearance model of sparse representation finds the best patches from the target patches. The best candidate patch can be evaluated by a particle of gathering a best candidate available with relating patches from the target layout even if there should be an occurrence of a moderate size change therefore improving its probability. The joint sparse

representation model minimizes the likelihood information of the best candidate available from the target object.

As per above representation, we have defined the likelihood measure for a lot of examined object focused inside a rectangular locale around the target, for different techniques as shown in Fig. 7. The proposed method is suitable to find the likelihood of the best candidate available from the target patches.

4.3 Occlusion handling model

We propose a new occlusion model based on joint sparse representation model. By using occlusion likelihood, we evaluate whether the current patch is occluded or not. If the occlusion likelihood is greater than 0.5 then the target is considered to be occluded. During the process of updating dictionary, the occluded patches are removed from the best candidate available from the target object. We target two state Markov chain to evaluate the occlusion state of most available patch as shown in Fig. 7. Maximum A Posteriori (MAP) estimation is used to update the likelihood model during online visual tracking. The progress possibility of this model is refreshed internet during following, by utilizing Maximum a Posteriori (MAP) estimation. Evacuating the patch superscript (I) from impediment states $o(i)$ and reestablishing the edge subscript Markov chain of occlusion state. Two-state Markov chain with versatile change possibility used to characterize impediment earlier of each patch.

States o and o speak to blocked and non-occluded states, individually. Suitable earlier for these parameters is the beta circulation, to be specific (a ; b), and (c ; d). By changing the parameters of these appropriations we can completely mirror any earlier information about the progress likelihood that decide patch impediment. With information on occlusion states in past casings and earlier appropriation of the progress likelihood can be detected target utilizing MAP estimation technique.

4.4 Contrasts model

Contrast is nothing but the difference in color or

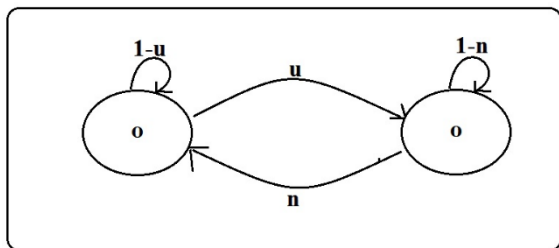


Figure. 7 Occlusion model

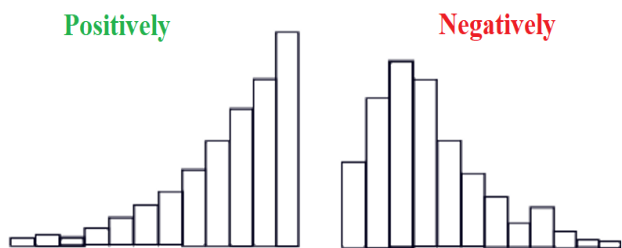


Figure. 8 Contrast model

luminance which makes an object detectable from other objects within the common field of view. A graphical description of the color scheme of an image is represented by an image histogram in the form of a histogram. In different words, the amount of pixels for each color in an image represents the image histogram. Based on the histogram for a particular image, initially it is easy to judge the contrast of the color image.

The proposed method is more reliable in a contrast situation. If the mean of the histogram is less than 0.2, then it is a low contrast image. This low contrast image mainly focuses on a weighted parameter to track an object more robustly. The weighted parameter is mainly used for calculating a patch during the process of tracking. The flat part of the chart represents the color scheme of an image, while the vertical axis indicates the absolute number of pixels in that specific contrast shown in Fig. 8. The left half indicates the dull region of the image. The center part represents the average of the color image. And the right side of the image describes the light region of the image. The vertical axis speaks to the dimensions of the territory (entire number of pixels) that is caught in each unmarked zone. In this manner, the histogram for a dull image could have the majority of its statistics specialized in the left aspect and focal factor of the diagram. On the other side, the histogram for an exceptionally brilliant picture with barely any dim territories as well as shadows will have a large portion of its information focused on the right side and focal point of the diagram.

4.5 Likelihood measurement

The objective picture was seen as an assortment

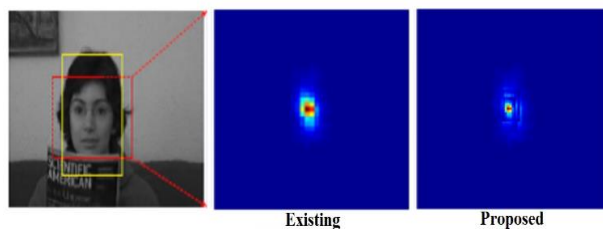


Figure. 9 Likelihood model

of non-covering patches. This brought about a probability measure which is determined as a total over fixed reproduction blunders. Moreover, an impediment identification plot was presented by exploiting patch recreation mistakes, and an earlier likelihood of impediment got from a versatile Markov chain of occlusion states.

Broad test results indicated the viability of the proposed approach on an assortment of video sequences. Specifically, our perceptions show that analyzing the target as an assortment of smaller patches decreased the impacts of impediment by confining its belongings to the blocked patches. For the likelihood model, we compare our proposed algorithm result with Zarezade et al. [26] as shown in Fig. 9.

The advantage of likelihood measurement is that the impeded patches can be delayed from this system. Likewise, we have seen that impressive performance was achieved by and large improved the tracker's presentation against appearance change, in spite of the fact that such improvement was speedy in terms of appearance changes and less size.

5. Result and analysis

Visual tracking is still a challenging task after more than three decades. The tracking speed and accuracy are the highly demanding areas for computer vision-based tracking. Nowadays, a patch-wise sparse method achieves better results compared to the other state-of-the-art tracker. So we are targeting on a joint sparse method to achieve high speed as well as better accuracy.

In order to establish the performance of the proposed tracker, we handle experiments on various challenging datasets which affect occlusion and contrast variation for tracking. In this section, we describe how the proposed tracker performed in complex environments for various challenging video sequences.

5.1 Input datasets

The full benchmark carries a hundred sequences from recent literatures. The datasets are usually called

Table 1. Problems full names

Attribute	Description	Attribute	Description
IV	Illumination Variation	PC	Pose Change
SV	Scale Variation	ST	Similar Target
OCC	Occlusion	CM	Camera Motion
DEF	Deformation	PO	Partial Occlusion
MB	Motion Blur	SO	Severe Occlusion
FM	Fast Motion	BC	Background Clutters
IPR	In-Plane Rotation	LR	Low Resolution
OPR	Out-of-Plane Rotation	OV	Out-of-View

Table 2. Methods full names

Method	Full names
OAB	Online AdaBoost Tracking
APG	Accelerated Proximal Gradient
Frag	Fragment-based Tracking
MIL	Multiple Instance Learning
PJS-S	Patch wise Joint Sparse Tracking M-FOCCUS
IVT	Incremental Visual Tracker
MTT	Multi-task sparse learning
PJS-M	Patch wise Joint Sparse Tracking Simultaneous Orthogonal Matching Pursuit (SOMP)

Girl, Skating, Man, Trellis, Faceocc2, which are available along with their ground truth at [35]. These datasets cover the challenging environment during the object tracking process. When multiple objects appearance during the object tracking process the target object represented as *dot+id_number* (e.g. Jogging.1 and Jogging.2). In this section we consider different challenging datasets for different tracker to become more. Table 1 shows list of challenges used as attribute for different datasets while Table 2 indicates full name of different tracker.

1) **David**: This dataset encompasses challenges which include pose exchange, partial occlusion, illumination variant and out of plane rotation. OAB moved away from target prior to frame 124 for illumination variation and pose change. As target rotates approximately to 90, APG, Frag and MIL trackers additionally misplaced the target. All

trackers drifted away from the target when target back to its initial position except PJS-OC tracker.

2) **Skating1**: In this video sequence person undergoes various visual changes such as illumination, pose variation and occlusion. Around frame 73, as the target goes through rapidly pose change, all trackers lost the target except OAB, PJSS and PJS-OC. Finally, after an extreme illumination exchange, simplest PJS-S tracker failed. Only PJS-OC was able to track the target throughout the sequences.

3) **Bird**: The main challenge in this video sequence is to large scale changes, motion and shape deformation with occlusion. When the bird is hidden by clouds, all trackers slightly drifted from the target. PJS-OC was able to track the bird throughout the all video sequences.

4) **Faceocc2**: In this video sequence, the main challenge is to track the target in various conditions like occlusion, rotation, pose variation. The target is not extremely occluded before frame 335. At that time all trackers tracking targeted successfully. After frame 335, APG, IVT, PJSS trackers slightly drifted from the target. Only PJS-OC was able to overcome these challenges successfully.

5) **Trellis**: This dataset is same as David dataset, however reveals sever illumination changes and large pose variation are demanding situation. After frame 212, due to large variation in illumination, all trackers drifted away from the target object. The only PJSS and PJS-OC is successfully track the target object. The Centroid error, Precision, VOC measure and Frame per second (FPS) are typical measurement parameters are plotted generally as a function of frame number.

Fig. 10 shows different tracker performance with our proposed algorithm named PJS-OC with different video sequence. We compare our result PJS-OC (Occlusion and Contrast) proposed method with different trackers namely Frag [27], OAB [28], MIL [29], IVT [30], APG [31], MTT [32], PJS [33]. The different parameters are describe as given below.

FPS: The full form of FPS is Frame Per Second. It is used to measure the number of consecutive frames execute with respect to time.

$$FPS = \frac{\text{number of frame}(n)}{\text{time}(t)} \quad (6)$$

Precision: Precision is described as a share between the quantity of legitimate frames detailed by using the module and the wide variety of all frames categorized as valid frames. where tp represent true positive, fp represent false negative.



Figure. 10 : (a) Bird datasets, (b) girl datasets, (c) skating datasets, (d) man datasets, (e) trellis datasets, and (f) faceocc2 datasets

$$Precision = \frac{tp}{tp + fp} \quad (7)$$

VOC (overlap score): A prediction of a part became taken into consideration true or false in line with the overlap test, as used within the detection part, i.e. for a real prediction the bounding container of the element overlaps the ground reality by using at the

least 50% [34]. For each component type, the average precision turned into used because the quantitative degree.

Centroid Error: The centroid error measures the difference between the objectives, expected middle from the tracker and the ground truth reality. In other word centroid error calculate the Euclidian distance between the middle of tracker and the ground truth bounding box.

Table 3. Frame per second (FPS) vs scale

Scale	1	0.9	0.8	0.7	0.6	0.5
Parameters	FPS					
Bird2	4.40	3.30	3.10	2.98	2.88	1.55
Crossing	6.64	6.64	6.63	6.57	6.46	1.69
David	6.78	6.78	6.50	6.23	5.13	1.70
FaceOcc1	4.00	4.00	3.96	3.23	3.92	1.43
FaceOcc2	4.65	4.65	4.67	4.51	4.25	1.56
Girl	6.16	6.16	6.01	6.12	5.99	1.67
Man	4.55	4.35	4.26	3.99	3.51	1.73
Skating	3.87	3.87	3.87	3.70	3.67	1.42
Trellis	3.93	3.93	3.90	3.79	3.65	1.31

Table 4. Precision vs scale

Scale	1	0.9	0.8	0.7	0.6	0.5
Parameters	Precision					
Bird2	0.88	0.89	0.89	0.95	0.98	0.99
Crossing	0.92	0.95	0.96	0.97	0.98	0.99
David	0.88	0.83	0.87	0.90	0.92	0.92
FaceOcc1	0.53	0.92	0.94	0.97	0.98	0.98
FaceOcc2	0.66	0.87	0.89	0.92	0.94	0.94
Girl	0.33	0.65	0.65	0.84	0.87	0.87
Man	0.93	0.96	0.96	0.96	0.95	0.99
Skating	0.39	0.96	0.95	0.97	0.97	0.98
Trellis	0.83	0.92	0.94	0.96	0.96	0.96

Table 5. VOC measure vs scale

Scale	1	0.9	0.8	0.7	0.6	0.5
Parameters	VOC Measure					
Bird2	0.60	0.76	0.80	0.85	0.87	0.87
Crossing	0.71	0.48	0.57	0.66	0.71	0.71
David	0.60	0.69	0.75	0.80	0.83	0.83
FaceOcc1	0.54	0.84	0.87	0.90	0.91	0.91
FaceOcc2	0.65	0.78	0.82	0.85	0.87	0.87
Girl	0.45	0.61	0.61	0.74	0.78	0.78
Man	0.55	0.56	0.63	0.72	0.76	0.76
Skating	0.65	0.74	0.78	0.83	0.85	0.85
Trellis	0.68	0.73	0.77	0.84	0.82	0.86

Table 6. Centroid error vs scale

Scale	1	0.9	0.8	0.7	0.6	0.5
Parameters	Centroid Error					
Bird2	79.34	6.27	4.87	3.47	2.77	2.77
Crossing	22.69	6.19	4.79	3.39	2.69	2.69
David	34.25	6.25	4.84	3.44	2.76	2.76
FaceOcc1	70.75	6.37	4.95	3.54	2.82	2.82
FaceOcc2	51.30	6.36	4.95	3.54	2.82	2.82
Girl	37.16	6.31	6.31	3.48	2.79	2.79
Man	22.54	6.19	4.79	3.38	2.70	2.70
Skating	23.84	6.28	4.87	3.46	2.78	2.78
Trellis	54.91	6.54	4.79	3.45	2.92	2.92

Table 7. Accuracy comparison of different tracker

	Fr ag	O AB	MI L	IV T	AP G	M TT	PJ SS	PJS - OC
Crossing	37	69	86	21	45	51	95	99
David	10	18	5	33	33	29	46	92
FaceOcc2	45	67	62	52	50	70	63	94
Skating	15	20	9	4	5	5	41	98
Trellis	22	5	17	22	13	12	72	96

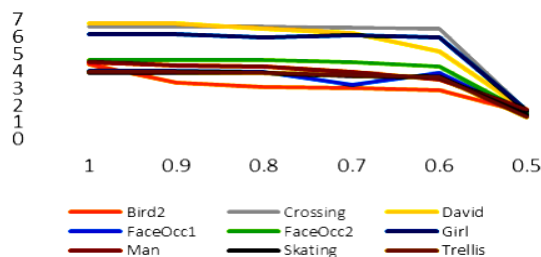


Figure. 11 FPS vs scale plot

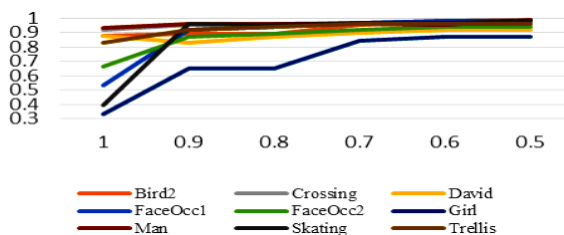


Figure. 12 Precision vs scale plot

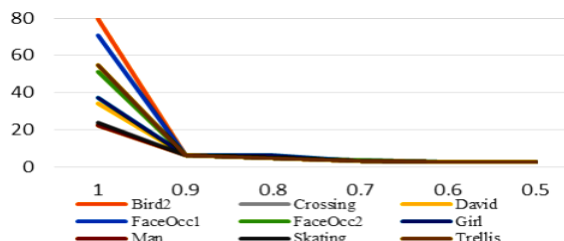


Figure. 13 Centroid error vs scale plot

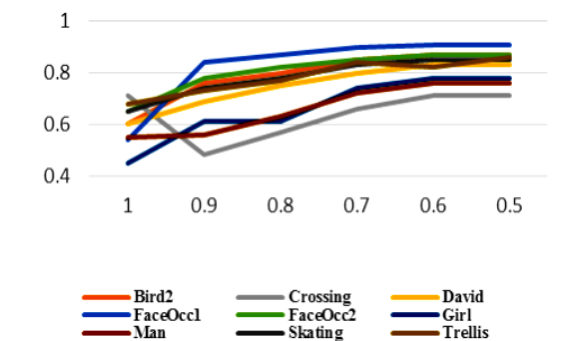


Figure. 14 VOC measure vs scale plot

Fig. 10 indicates sample frames and the bounding containers found with the aid of every tracker in

every dataset. The above video sequences indicate that the newly proposed method have been

performing well in challenging environment in which other trackers have lost the target. The main advantage of proposed algorithm is that to track objects in the heavy occlusion scenario. Fig. 11 and 12 show the frame per second and precision with different the sampling rate. From such figure we observe that as the sampling rate is decrease the frame rate is also decreasing and speed is also increased for entire video sequence. Fig. 13 indicates the centroid error of object during tracking process. Same as Fig. 14 represent the overlap ratio of different object for different sampling value. Table. 3, Table 4, Table 5, and Table 6 provides a comparison of frame per second, precision, VOC, centroid error for different sampling value 0.5 to 1 respectively. The proposed tracker is best for all sampling value. Also for all sequences the proposed tracker for sampling value of 0.5 achieve first or second place. The main advantage of the patch wise sparse method is to track target bounding box accurately align with the object. Also for lower sampling rate for video sequence the speed has considerably increased. The average precision in all datasets for the sampling rate of 0.5 is 96%, which is considerably higher than other sampling rate, i.e. 0.6 to 1. The frame rate of David, Crossing, Girl datasets are higher for sampling value of 1. Table. 7 indicates a comparison of different trackers with different datasets. The proposed algorithm is far better than other algorithms in terms of accuracy. As the proposed algorithm achieves higher accuracy with increasing speed of given datasets. For crossing datasets MIL, PJSS and our proposed algorithm performed well. Above observation indicate that as sampling rate decrease, speed is increased gradually and accuracy is also increased for all challenging datasets.

6. Conclusion

In this paper, we targeted the joint sparse representation approach to track the target object in sequential frames. The non-overlapping patches mainly represented the target image. The likelihood measures evaluated as a sum of reconstruction error. An adaptive Markov chain was introduced the patch reconstruction errors and likelihood of occluded objects. Additionally, the occlusion model was applied by taking the benefits of the patch reconstruction error and the initial likelihood of occlusion derived by the Markov chain using the occlusion states. The result of proposed method showed the robustness of the joint sparse method on different challenging video sequence. The effect of occlusion was decreased by using a collection of

smaller patches of the target. This is further significant for refreshing the dictionary, the fact that the occluded patches can be avoided from this method. Also, we have determined that implementing the temporal comparison assumption commonly increased the performance of tracker against appearance change, despite the fact that this improvement was less significant for video arrangements with quick changes in appearance. The average precision of the proposed method is 96% against the other state of art methods. Also for the sampling value of 0.5, our method achieved the highest speed with maximum precision. Numerous challenges may be followed for future work. For future work, improved tracker speed is challenging aspect to implement in real time application. The tracker performance can be improved by using low weight reconstruction error patches.

Conflicts of Interest

The authors have no conflict of interest.

Author Contributions

The contribution of author is given below:
Writing and Implementation: Yashesh Joshi
Conceptualization, supervision, review and editing: Hiren Mewada.

References

- [1] J. Patoliya and H. Mewada, "Automated Tracking and Real Time Following of Moving Person for Robotics Applications", *Journal of Automation, Mobile Robotics and Intelligent Systems*, Vol. 13, No. 4, 2020.
- [2] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet. "Color-based probabilistic tracking", In: *Proc. Eur. Conf. Comput. Vision.*, pp. 661-675, 2002.
- [3] S. Avidan, "Ensemble tracking", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 2, pp. 261-271, 2007.
- [4] D. Israni and H. Mewada, "Identity Retention of Multiple Objects under Extreme Occlusion Scenarios using Feature Descriptors", *Journal of Communications Software and Systems*, Vol. 14, No. 4, pp. 290-301, 2018.
- [5] V. Ssrikrishnam, T. Nagara, and S. Chaudhari, "Fragment based tracking for scale and orientation adaption", In: *Proc. of Indian Conf. on Computer Vision, Graphics & Image Processing*, pp. 328-335, 2008.
- [6] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking",

- In: Proc. Eur. Conf. Computer Vision*, pp. 234-247, 2008.
- [7] G. Zhu, S. Zhang, X. Chen, and C. Wang, "Efficient illumination insensitive object tracking by normalized gradient matching", *IEEE Trans. On Signal Processing*, Vol. 14, No. 12, pp. 944-947, 2007.
- [8] Q. He, C. Wei, and Y. Xu, "An improved adaptive Kalman filtering algorithm for balancing vehicle", In: *Proc. of IEEE Conf. on Chinese Automation Congress*, pp. 5721-5725.
- [9] M. Li, L. Peng, T. Wu, and Z. Peng, "A bottom up and top – down integration framework for online object tracking" *IEEE Transaction on Multimedia*, 2020.
- [10] S. Balaji and S. Karthikeyan, "A survey on moving object tracking using image processing", In: *Proc. of IEEE International Conf. on Intelligent Systems and Control*, pp. 469-474, 2017.
- [11] B. Wang, F. Chen, and P. Deng, "Online object tracking based on sparse subspace representation", In: *proc. of Chinese Control and Decision Conf.* pp. 3975-3980, 2014
- [12] L. Chen, F. Zhou, Y. Shen, X. Tian, H. Ling, and Y. Chen, "Illumination insensitive efficient second order minimization for planar object tracking", In: *Proc. of IEEE International Conf. on Robotics and Automation*, pp. 4429-4436, 2017.
- [13] M. Turhan and D. Hanbay, "Mean shift based object tracking supported with adaptive Kalman filter", In: *Proc. of IEEE Conf. on Signal Processing & Communication Applications*, pp. 2670-2673, 2015.
- [14] Y. Shan, Z. Zhang, P. Yang, and K. Huang, "Adaptive slice representation for human action classification", *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 25, No. 10, pp. 1-13, 2015.
- [15] M. Yang, Y. Wu, and G. Hua, "Context aware visual tracking", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 7, pp. 1195-1209, 2009.
- [16] K. Jeong, J. Choi, and G. Jang, "Semi local structure patterns for robust face detection", *IEEE Trans. On Image Processing*, Vol. 22, No. 9, pp. 1400-1403, 2015.
- [17] W. Sheng and D. Fang, "Angular super resolution for phased antenna array by phase weighting", *IEEE Trans. On Aerospace and Electronic Systems*, Vol. 37, No. 4, 2001.
- [18] W. Hu, W. Li, X. Zhang and S. Maybank, "Single and multiple object tracking using a multi feature joint sparse representation", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 37, No. 4, pp. 816-833, 2015.
- [19] H. Yang, S. Qu, C. Chen, and B. Yang, "Multiple objects tracking with improves sparse representation and rank based dynamic estimation", *IEEE Access*, pp. 42264-42278, 2018.
- [20] B. Babenko, M. Yang, and S. Belongie. "Robust object tracking with online multiple instance learning." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1619–1632, 2011.
- [21] R. Venkatesh Babu, P. Perez, and P. Boutheymy, "Robust tracking with motion estimation and local Kernel-based color modeling", Elsevier, Vol. 25, No. 8, pp. 1205-1216, 2007.
- [22] M. Ashwini, R. Babu, and K. Ramakrishnan, "Context aware real time tracking in sparse represented framework", In: *Proc. of IEEE International Conf. on Image Processing*, pp. 2450-2454, 2013.
- [23] A. Ruchandani and R. Babu, "Local appearance based robust tracking via sparse representation", In: *Proc. of Indian Conf. on Computer Vision, Graphics and Image Processing (ICVGIP)*, 2012.
- [24] A. Achmad, L. Nurtanio and F. Mayasari, "Blob modification in counting vehicles using Gaussian mixture models under heavy traffic", *Asian Res. Publ. Netw. ARPANet*, Vol. 10, No. 16, pp. 7157-7163, 2015.
- [25] D. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation", *IEEE Trans. on Pattern Anal.*, Vol. 31, No. 2, pp. 210-227, 2009.
- [26] M. Nalamati, A. Kapoor, M. Saqib, N. Sharma, and M. Blumenstein, "Drone detection in long range surveillance video", In: *Proc. of IEEE International Conf. on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1-6, 2019.
- [27] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragment based tracking using the integral histogram", In: *Proc. of Computer Vision and Pattern Recognition*, pp. 798-805, 2006.
- [28] H. Grabner, M. Grabner, and H. Bischof, "Real-time tracking via on-line boosting", In: *Proc. of British Machine Vision Conf.* pp. 6.1-6.10, 2006.
- [29] B. Babenko, M. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning", *IEEE Transaction Pattern Anal.*, Vol. 33, No. 8, pp. 1619-1632, 2011.
- [30] D. Ross, J. Lim, R. Lin, and M. Yang, "Incremental learning for robust visual tracking", In: *Proc. of Int. Journal of Computer Vision*, pp. 125-141, 2008.

- [31] C. Bao, Y. Wu, H. Ling, and H. Ji, "Real time robust 11 tracker using accelerated proximal gradient approach", In: *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1830-1837, 2012.
- [32] T. Zang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via multi task sparse learning", In: *Proc. of IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2042-2049, 2012.
- [33] A. Zarezade and H. Robiee, "Patchwise joint sparse tracking with occlusion detection", *IEEE Trans. on Image Processing*, Vol. 23, No. 10, pp. 4496-4510, 2015.
- [34] Naresh, P. Parate, and R. Venkateshbabu, "Fragment based real time object tracking: a sparse representation approach", In: *Proc. of IEEE International Conf. on Image Processing*, pp. 433-436, 2012.
- [35] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," In: *Proc. of IEEE Conf. Comput. Vision and Pattern Recognition*, 2013.