International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Tracking Moving Objects: A Comparative Study

Walaa Omar El-Farouk Badr[1], Dr. Hossam El-Din Mostafa[2]

[1]Electronics and Communication Department, The High Institute of Engineering and Technology in Mansoura.

[2]Electronics and Communication Department, Faculty of Engineering, Mansoura University

Abstract: ***Visual tracking is considered to be one of the most important challenges in computer vision with numerous applications such as object recognition and detection.Inthe present paper, four tracking techniques will be introduced circulant structure with kernels (CSK), Kernelized correlation filters (KCF), Adaptive color attributes (ACT), and distractor – awareness tracker (DAT) for the visual object tracking (VOT14) challenge datasets. Performance evaluation for each method was calculated using three measures; center location error (CLE), distance precision (DP), and speed in frames per second (FPS). Results have shown that KCF tracker is the fastest technique. It achieves the best results in most sequences and the highest precision at lower threshold. It is used in time-critical application with satisfactory performance. Each tracker performs favorable and competitive results in some sequence and fails in others. So it is noted that the choice of the tracker is application dependent.***

Keywords: Visual tracking; circulant; correlation filter; distractor; distance precision; precision plot

1. Introduction

One of the main goals of computer vision is to enable computers to replicate the basic functionsof human vision such as motion perception and scene understanding. To achieve thegoal of intelligent motion perception, much effort has been spent on visual object tracking. Essentially, the core of visual object tracking is to robustly estimate the motion state (i.e., location, orientation, size, etc.) of a target object in each frame of an input image sequence

Object tracking is considered to be one of the most important problems of image analysis with numerous applications. It is concerned with low-level visual processing and high-level image analysis. Moreover, it is also widely used in image understanding, human- computer interaction, surveillance, and robotics. This problem is such a challenging one as it needs to deal with appearance variations caused by numerous factors such as illumination variations, pose angle, partial occlusion, background clutter, shape deformation, motion blur, scale variation and out -of - plane rotation [1].

Visual object tracking methods include image input, appearance feature description, context information integration, decision and model update as shown in Fig.1 [2]. Most trackers either depend on intensity or texture information [3, 4], while others depend on color information that is limited to simple color space transformation [5]. In contrast to visual tracking, color features are providing excellent performance for different application such as object recognition and detection. Using color information for visual tracking is a very difficult problem due to variation in illumination, shadows, shading, camera and object geometry. So, it is a must to choose the suitable color transformation for visual object tracking.

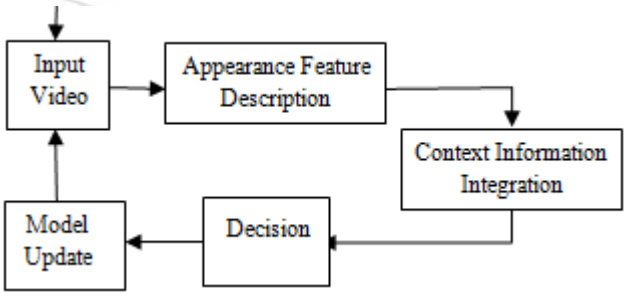

Figure 1: The flowchart of visual tracking

To tackle these challenges; firstly, the objective is to determine the position of the object in the first frame either manually or by using reference model (ground truth). Secondly, it is a must to detect the locations of object in an image sequence with separating the target object from the background. The object is tracked in each frame of the video by several approaches. In this paper, there is a comparison between several methods of visual tracking that provide highperformance among the top visual trackers such as: tracking by detection approach circulant structures with kernels (CSK), KCF approach, Adaptive color attributes (ACT) and distractor awareness (DAT).The aim of these methods is to track the object in each frame by trying to find out the region in the frame whose interior generates a sample distribution over the target object model which has the best match with the reference model distribution.

2. Previous Work

Due to the importance of visual tracking, many approaches have been proposed to handle its problems. There exist two main approaches namely discriminative and generative methods that are used to handle the different problems of visual tracking. The generative methods handle the problem by searching for regions that are most similar to the target model [4, 6]. Recent benchmark evaluation shows that the generative models are outperformed by discriminative approaches which incorporate binary classifier to distinguish the object from its surrounding background [7, 8]. The models in these methods are based on templates, subspace models, HOG features [9] and Haar-like features [4, 3].

However, rectangular initialization bounding boxes include background information. Another method uses segmentation methods in order to improve the generative methods, but these methods still suffer from missing the advantages of discriminative methods to distinguish the object from its surrounding background [10, 7]. Another problem with using template-based methods is that the objective function is not enough to achieve the optimum solution [11]. So, an alternative is the use of histogram to describe the object. Histogram-based (kernel-based) descriptors integrate information over a large patch of the image. So they are not sensitive to spatial structure and give best results due to they are very fast. Another approach that uses multiple kernel to overcome the problem of losing of spatial information, which happens when building the histogram and improve the results of single histogram descriptor [12, 11], but it requires other mechanism to determine the number and shape of the kernels. If the number of kernel is too small, other additions are statistics analysis, and using feature selection [13, 14].Distribution fields (DFs)uses a histogram that contains robust information and preserves the spatial information of the object by using a distribution at every pixel. It can be shown that it is a combination of histogram-based descriptors and template-based descriptors [11].

On the other side, the discriminative approaches pose the problem by differentiating the target from the background by using tracking as a binary classification problem.. It has also been exploited to handle appearance changes during visual tracking, where a classifier is trained and updated online to distinguish the object from the background. This method is also termed as tracking by detection, in which a target object identified by the user in the first frame is described by a set of features. A separate set of features describes the background, and a binary classifier separates target from background in successive frames. To handle appearance changes, the classifier is updated incrementally over time. They also exploit visual information from the target and the background. Due to the success of discriminative approaches, many classifiers can be explored such as: SVMs, RVM [3] and several methods which depend on boosting [15] in order to distinguish the foreground from the background by an ensemble of classifiers. Some trackers use a tracking method that recognizes the object representation by partial least squares analysis and using more than one appearance model which is initialized in the first frame [14].

3. Tracking Techniques

3.1 The CSK Tracker

Tracking by detection has been proved to be a successful method. This stems directly from the development of discriminative methods in machine analysis, and their application to detect with offline training. It provides the highest speed among the visual trackers due to the circulant structure of the kernel.This method explores a dense sampling strategy by training a Gaussian kernel classifier with all subwindows (samples).It allows more efficient training. The reason is that the kernel matrix in this case becomes highly structured and circulant.This algorithm could be operated directly on the pixel values and without using feature extraction due to the using of fast Fourier transform.

Steps of Tracking

- Initializing the target object in the first frame manually or using the first position of the object from the ground truth.
- Training images must be pre-processed with cosine windows
- Calculating the response of the classifier at all locations(subwindows) by using dense gauss kernel and FFT using equation of detection as follows,

$$response = real\left(ifft2\left(alphaf .* fft2(k)\right)\right) \quad (1)$$

where $alphaf$ is a classifier coefficient in Fourier transform and (.∗) called product–wise operationin matlab and$fft2(\,)$, $ifft2(\,)$ are fast Fourier transform, inverse fast Fourier transform respectively in matlab.

- Finding the maximum response.
- Getting subwindow at the current position of the target so as to train the classifier
- Training new models in order to determine new alpha and new position where alpha is kernel regularized least square solution(KRLS) according to the equation,

$$alphaf_{new} = yf ./fft2(k) + \lambda \quad (2)$$

whereyf is a classifier response in Fourier transform and λ is regularization parameters, $\lambda = 10^{-2}$ in the present work.

- Finding Gaussian kernel (k) by using dense gauss kernel function as follows,

$$k = \exp\left(-\frac{1}{\sigma^2}\left(\|x^2\| + \|z^2\| - 2F^{-1}(F(x) \odot F^*(z))\right)\right) \quad (3)$$

where$\sigma = 0.2$ is a gaussian kernel bandwidth, x is training image at current frame and z is test image at next frame.$F(.)$, $F^{-1}(.)$denote fourier transform and inverse fouruer transform respectively.See reference paper [4] for more details.

3.2 The KCF Tracker

It is the new version of CSK tracker but it deals with multi-channel HOG features for best performance. It depends on Gaussian Kernel correlation.The input patches are weighted by a cosine window that smoothly removes discontinuities at the image boundaries caused by the cyclic shift. The region of tracking has three times the size of the target to provide additional negative samples and some context. Due to the training samples which consist of shifts of a base sample, it is a must to specify a regression target for each one in y [9].

The tracker implements three functions as follows:

- Training function: it trains the image patch at the initial position of the target

  $Alpha = train(x, y, sigma, lambda)$,

[9] where x is the train image patch, sigma is feature bandwidth, and lambda is regularization factor.

- Detecting function: it detects over the patch at the previous position and the target position is updated to

the one that has the maximum value. So, train has a new model at the new position.

$$response = detect(alphaf, x, z, sigma),$$

where z is the test image patch.

- Kernel correlation function, as it is called by the two previous functions, will compute Gaussian kernel correlation between x, z

$$K = kernal\ correlation(x, z, sigma),$$

where k can be written as k^{xz},

$$k^{xz} = \exp(-\frac{1}{\sigma^2}(\|x^2\| + \|z^2\| - 2F^{-1}(\sum_c \hat{x}_c^* \odot \hat{z}_c))) \quad (4)$$

where $\sigma = 0.5$,$F^{-1}$is the Fourier inverse, $\hat{x}_c^*$ is the conjugate of Fourier train patch for channel c and $\hat{z}_c$ Fourier test patch for channel c. The operations are only element-wise operations in Fourier domain due to the diagonalization, which result in the tracker to be the faster one.One challenge for the system is that happens due to the absence of a failure recovery mechanism. For more details of the tracker see reference paper [9].

3.3 The ACT Tracker

It is the extension of the CSK tracker with color attributes, which have shown excellent performance and results for object recognition.Color attributes, or color names (CN), are linguistic color labels assigned by humans to represent colors in the world. In a linguistic study performed by Berlin and Kay [16], it was concluded that the English language contains eleven basic color terms: black, blue, brown, grey, green, orange, pink, purple, red, white and yellow. In the field of computer vision, color naming is an operation that associates RGB observations with linguistic color labels. The mapping provided by [17] is used, this mapping is automatically recognized from images retrieved with Google-image search. This maps the RGB values to a probabilistic 11 dimensional color representation which sums up to 1.

Nevertheless, the high dimensionality of color attributes results in an increasing in the time performance and computational overhead, which could limit its application in real time surveillance. So, in order to overcome this problem, the ACT tracker proposes an adaptive dimensionality reduction technique which reduces the eleven dimension of the color attributes to two only [18].

The ACT updates the MOSSE trackerfrom linear kernels classifier and one dimensional feature to Gaussian classifier and multi-channel color features to be sub-optimal.Since the visual tracking is sensitive to appearance changes, so it is necessary for the target model to be updated over time through equation,

$$\dot{x}^p = (1 - \gamma)\dot{x}^{p-1} + \gamma x^p \quad (5)$$

where$\dot{x}^p$ is the updated learned target appearance, γ is the learning rate for the appearance model update, and p is the index of the current frame. The advantage of this model that it is not needed to store all the previous appearance but only the current model in each new frame can be saved.

Finally, the tracker proposes an adaptive dimensionality reduction technique that preserves useful information while reducing the number of color dimensions.

3.4 The DAT Tracker

The tracker presents a discriminative object model to differentiate the object of interest from the background. Also, it relies on standard color histograms. In contrast, it extends this model to identify and suppress distracting region in advance to improve the tracking performance. It proposes an efficient scale estimation scheme which gives the chance and allows obtaining accurate tracking results.

There is a difference between supporting and distracting regions.Supporting regions have different appearance than the target but co-occur with it, providingvaluable cues to overcomeocclusions.Distractors, on the other hand, exhibit similar appearance and may therefore get confused with the target. So, it needs to track these distracters in addition to the target in order to prevent drifting. DAT tracker adapts the object representation such as that potentially distracting region which is suppressed in advance with the background. So it combines object background model with distractor aware model to give the final object model as follow,

$$P\left(x \in \frac{O}{b_x}\right) = \lambda_p P\left(x \in \frac{O}{O}, D, b_x\right) + (1 - \lambda_p) P\left(x \in \frac{O}{O}, S, b_x\right) \quad (6)$$

whereλ_p=0.5, is a predefined weighting parameter [19].

Thus, applying this model causes high likelihood scores while decreasing the effect of distracting region. So no explicit tracking of distractors is required. It uses tracking-by-detection principle to localize the object of interest in a new frame and obtain the new location as follow,

$$\dot{O}_t = \arg_{O_{t,i}} \max(s_v(O_{t,i}) s_d(O_{t,i})) \quad (7)$$

where$s_v(.)$, $s_d(.)$ denote vote score and distance score respectively.After localizing the object, it performs scale estimation models to adapt the scale of the current object hypothesis $\dot{O}_t$according to,

$$O_t = \lambda_s O_t^S + (1 - \lambda_s)\ \dot{O}_t \quad (8)$$

whereλ_s=0.2 scale update parameter, for more details see reference paper [19].

4. Experimental Results

4.1 Datasets

The presented approaches were implemented usingMatlab versionR2015a (8.5) on IntelCore(TM) i5-3230M CPU 2.60 GHz with 4GB RAM. The VOT14 datasets had been utilized. This dataset contains videos which have been collected from well-known tracking evaluations: such as the Amsterdam Library of Ordinary Videos (ALOV) [20]. The VOT committee proposed a sequence selection methodology in order to compile datasets which cover various real-life

visual phenomena. As a result, the datasets consist of 25(VOT14) [21] sequences pose challenging situations such as illumination changes, object deformations and appearance changes, abrupt motion changes, significant scale variations, camera motion, and occlusions. These challenging dataset are considered to be the largest model free tracking benchmarks till now.

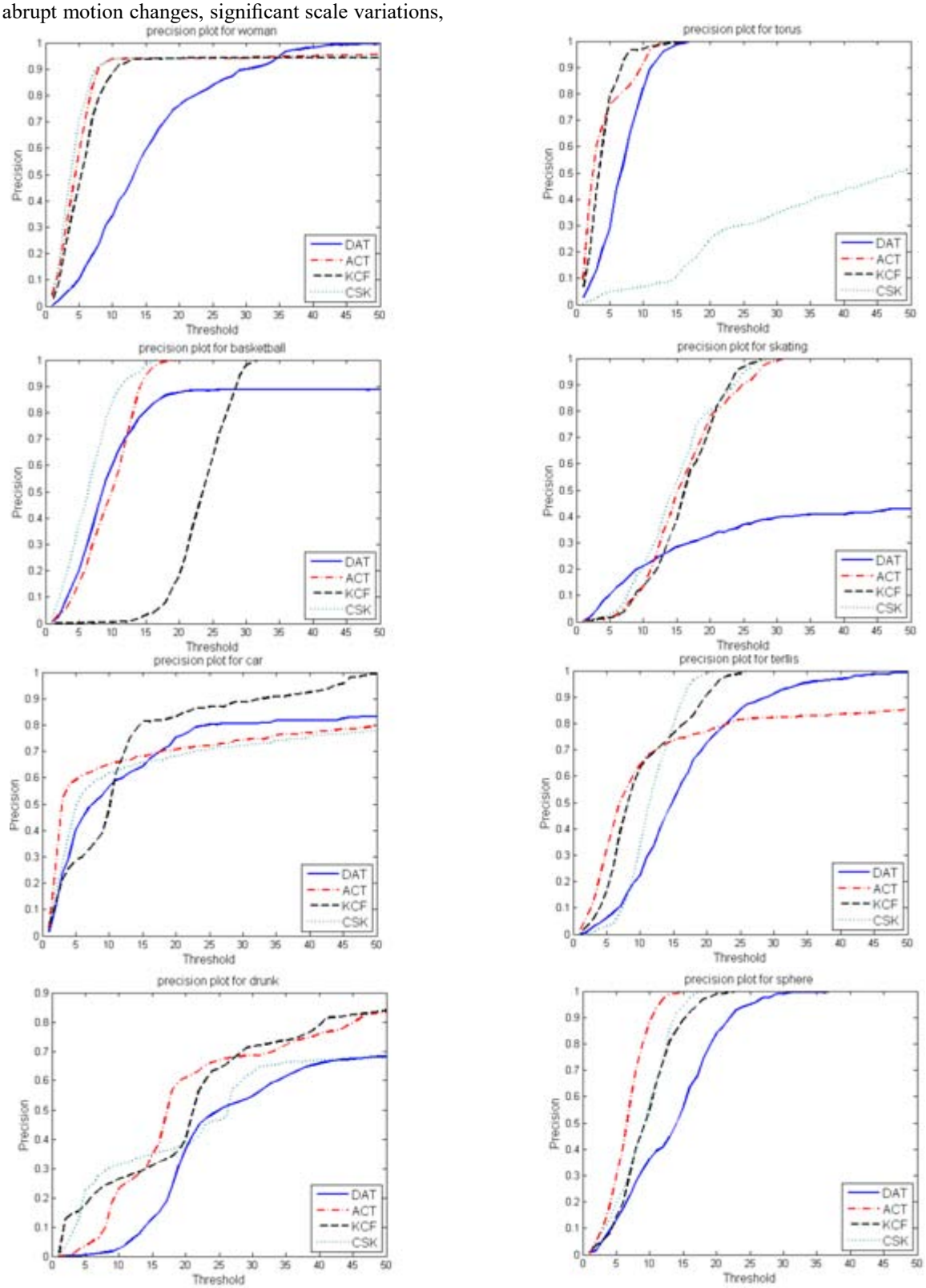

Figure 2: Precicion plots of different sequence of the VOT14 dataset

4.2 Evaluation methodology

To compare between the presented algorithms, Results were compared using two evaluation metrics, center location error (CLE), and distance precision (DP). CLE is computed as the average Euclidean distance between the estimated center location of the target and the ground-truth. DP is the relative number of frames in the sequence where the center location error is smaller than a certain threshold [18]. The DP values were reported at a threshold of 20 pixels [6, 18]. The results were summarized in Table (1) using the median CLE and DP values over all sequences. Also, the speed of the trackers

was taken into consideration in median frames per second (FPS). Figure (2) shows the precision plots for 8 different sequences taken from VOT14 datasets.

A precision plot shows the ratio of successful frames whose tracker output is within the given threshold (x-axis of the plot, in pixels) from the ground-truth, measured by the center distance between bounding boxes.

In the precision plots, the distance precision is plotted over a range of thresholds as shown.The trackers wereranked using the DP scores at 20 pixels. A higher precision at low thresholds means that the tracker is more accurate, while a lost target will prevent it from achieving perfect precision for a very large threshold range. When a representative precision score is needed, the chosen threshold is 20 pixels, as done in previous works.

From the results in Table (1), KCF tracker is the best in case of speed (runs at hundreds of frames per seconds)so it can be used in time critical applications such as visual surveillance or robotics. It achieves the best DP equally with ACT if it is compared with the other trackers. The ACT is the best in case of mean CLE at the cost of lower frames rates. It is also found that each tracker is the best in some sequences only. This is due to the attributes of the sequence such as illumination variations, pose angle, partial occlusion, background clutter, shape deformation, motion blur, scale variation and out -of -plane rotation. But the ACT and KCF are more stable and reach the value 1 of the precision value at low threshold than the two others in most sequences.ACT tracker provides significant performance due to the using of color attributes but with lower speed. KCF is the fastest tracker due to its circulant structure and its diagonalization by the DFT.

5. Conclusion

A comparison between four tracking algorithms have been presented. It is noted that the KCF tracker runs at hundredof FPS so it is suitable for real time-critical applications and can be implemented with only a few lines of code. It achieves competitive DP in the most of sequences of the dataset VOT14. It is noted that all trackers fail in at least one sequence. Hence, the choice of a tracker depends on attributes of the video and the application under consideration.

References

[1] Q. Wang and F. Chen, “Object Tracking via Partial Least Squares Analysis”, IEEE Trans. Image Processing, vol. 21, no. 10, pp. 4454-4465, Oct 2012.
[2] H. Yang, L. Shao, F.Zheng, L.wang, and Z. Song, “Recent advances and trends in visual Tracking: A review”, Neurocomputing, vol. 74, no. 18, pp. 3823-3831, Nov 2011.
[3] S. Hare, A. Saffari, and P. Torr, “Struck: Structured output tracking with kernels”, In ICCV, 2011.
[4] K. Zhang, L. Zhang, and M. Yang, “Real-time compressive tracking”, In ECCV, 2012.
[5] S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan, “Locally orderless tracking”, In CVPR, 2012.
[6] J. Henriques, R. Caseiro, P. Martins, and J. Batista,” Exploiting the circulant structure of tracking-by-detection with kernels”, In ECCV, 2012.
[7] M. Godec, P. M. Roth, and H. Bischof, “Hough-based Tracking of Non-Rigid Objects”, In Proc. ICCV, 2011.
[8] Y. Wu and M.-H. Yang, “Online Object Tracking: A Benchmark”, In Proc. CVPR, 2013.
[9] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista,“High Speed Tracking with krnelized Correlation Filters”, PAMI, 2015.
[10] V. Belagiannis, F. Schubert, N. Navab, and S. Ilic, “Segmentation Based Particle Filtering for Real-Time 2D Object Tracking”, In Proc. ECCV, 2012.
[11] L. Sevilla-Lara and E. Learned-Miller, “Distribution fields for tracking”, In CVPR, 2012.
[12] G. Hager, G. D. Hager, M. Dewan, and C. V. Stewart, “Multiple kernel tracking with SSD”, In CVPR, 2004.
[13] A. P. Leung and S. Gong, “Mean shift tracking with random sampling”, In BMVC, 2006.
[14] M.Rutharesh, R.Naveenkumar, and S. Krishnakumar, “Object Tracking Using Partial Least Squares Analysis”, In IJARCSSE, vol. 4, no.2, pp.235-239, Feb 2014.
[15] H. Grabner, C. Leistner, and H. Bischof, “Semi-supervise On-line Boosting for Robust Tracking”, In Proc. ECCV, 2008.
[16] B. Berlin and P. Kay,” Basic Color Terms: Their Universality and Evolution”, UC Press, Berkeley, CA, 1969.
[17] J. van de Weijer, C. Schmid, J. J. Verbeek, and D. Larlus, “Learning color names for real-world applications”, TIP, 18(7):1512–1524, 2009.
[18] M. Danelljan, F. S. Khan, M.Felsberg, and J. van de Weijer, “Adaptive Color attributes for Real-Time Visual Tracking”, In Proc. CVPR, 2014.
[19] H. Possegger, T.Manuthner, and H. Bischof, “In Defense of Color-based Model-free Tracking”, In CVPR, 2015.
[20] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, “Visual Tracking: an Experimenta Survey”, PAMI, 36(7):1442–1468, 2014.
[21] M. Kristan, R. Pflugfelder, A. Leonardis, J. Matas, L.Cehovin, G. Nebehay, G. Fernandez, et al, “The Visual Object Tracking VOT2014 challenge results”, In Proc. VOT (ECCV Workshop), 2014.

Table 1: Performance evaluation of the four tracking algorithms using VOT14 dataset

	CSK			KCF			ACT			DAT		
sequence	**CLE**	**DP**	**FPS**	**CLE**	**DP**	**FPS**	**CLE**	**DP**	**FPS**	**CLE**	**DP**	**FPS**
ball	15.9	0.66	**117.39**	10	0.93	61.63	13.7	0.82	58.7	**8.16**	**0.97**	34.77
basketball	**6.55**	**1**	127.3	8.35	0.92	**154**	9.45	0.99	67.4	8.8	0.97	25.5
bolt	4.46	**1**	171.6	6.63	0.99	**211.47**	**4.13**	**1**	86.5	7.7	0.98	25.6
bicycle	**4.22**	**1**	205.6	5.5	**1**	**234**	4.65	**1**	84.4	9.33	0.96	50.2
car	29.7	0.68	262.98	**12.7**	**0.83**	**390.7**	27.4	0.71	103	21.6	0.75	51.2
david	12	0.86	40.76	**7.17**	**1**	**51.84**	26.5	0.41	31	25.7	0.49	42.04
drunk	37.8	0.38	18.37	**24.6**	0.4	**45.69**	28.8	**0.61**	31.9	43.6	0.37	29.68
Fish1	111	0.16	**199.15**	103	0.2	176.22	20.2	0.68	149	**9.47**	**0.9**	59.16
Fish2	307	0.23	**111.07**	142	0.28	96.05	80.9	**0.58**	58.3	**76.5**	0.44	26.31
Hand1	**57.1**	0.31	142.6	74	0.21	**183.58**	62.7	**0.43**	130	72.9	0.39	8.74
Hand2	69	0.2	442.3	49.5	0.23	**487.89**	61.6	0.26	152	**15.8**	**0.85**	31.68
polarbear	15.4	0.68	60.23	**11.2**	**0.96**	**115.57**	19.3	0.65	80.2	24.2	0.26	28.12
skating	**14.8**	**0.8**	70.78	16.2	0.73	**145.63**	15.9	0.73	68.7	139	0.34	28.6
sphere	8.9	**1**	**91.75**	9.47	0.99	87.29	**6.6**	**1**	33.4	13.3	0.84	40.09
sunshade	**3.3**	**1**	92.31	4.3	**1**	**93.79**	**3.3**	**1**	95.3	11.1	0.99	40.71
surfing	**1.67**	**1**	**271.84**	2.29	**1**	158.5	2.03	**1**	141	2.46	**1**	57.35
torus	52.6	0.3	**196.09**	**3.7**	1	140	3.94	**1**	80.2	6.62	**1**	40.55
trellis	11.7	0.99	**83.34**	**10.1**	**0.99**	50.43	33.8	0.77	17.5	16.4	0.73	57.96
tunnel	10.8	0.99	**92.9**	**6.5**	**1**	77.2	10.2	**1**	45.3	27.2	0.28	27.5
woman	11.4	0.94	**231.22**	**7.67**	**0.985**	75.8	8.34	**0.94**	87.5	14.9	0.76	39.68
mean	39.15	0.7	151.46	25.74	**0.78**	**151.85**	**22.17**	**0.78**	80.1	27.74	0.71	38.65