Journal of Discrete Mathematical Sciences & Cryptography
ISSN 0972-0529 (Print), ISSN 2169-0065 (Online)
Vol. 27 (2024), No. 4, pp. 1455–1465
DOI : 10.47974/JDMSC-1999

Vigilance and surveillance reinforced using mathematical approaches in object tracking techniques

Kamal Upreti *
Department of Computer Science
CHRIST University
Delhi NCR Campus
Uttar Pradesh
India

Shreya Kapoor †
Department of Computer Science and Engineering
Dr. Akhilesh Das Gupta Institute of Professional Studies
Shashti Park
New Delhi
India

Soumi Ghosh §
Department of Information Technology
Maharaja Agrasen Institute of Technology
Rohini
Delhi
India

Dhyanendra Jain ‡
Department of Computer Science & Engineering
(Artificial Intelligence & Machine Learning)
ABES Engineering College
Ghaziabad
Uttar Pradesh
India

* *E-mail:* `kamalupreti1989@gmail.com (Corresponding Author)`
† *E-mail:* `kapoor.shreya2506@gmail.com`
§ *E-mail:* `ghoshdrsoumi@gmail.com`
‡ *E-mail:* `dhyanendra.jain@gmail.com`

Virendra Singh Kushwah [@]
School of Computer Science & Engineering
VIT Bhopal University
Bhopal-Indore Highway
Bhopal
Madhya Pradesh
India

Jyoti Parashar [#]
Department of Computer Science and Engineering
Dr. Akhilesh Das Gupta Institute of Professional Studies
Shashti Park
New Delhi
India

Rituraj Jain [$]
Tarannum Bloch [^]
Department of Information Technology
Marwadi University
Rajkot
Gujarat
India

Abstract

Visual tracking is crucial to the study of object recognition and has been utilized in a variety of realistic settings, such as robotics, traffic monitoring, self-driving automobiles, forensics, and more. This research concentrates on techniques for counting the total number of individuals entering or exiting a space under the watchful eye of a camera. The techniques described here can detect the number of persons in a scene, both for a single individual and for many passers in front of the camera. With the aid of surveillance that use the centroid concept, an effective solution has been devised for monitoring. Secondly, in this study, object tracking methods utilising deep learning are also reviewed and analysed. This study also compares the effectiveness of various algorithms on the LaSOT, VOT2015, VOT2016, VOT2017 and OTB2015 tests.

Subject Classification: *Primary 68R01, Secondary 97N70.*

Keywords: *Single object tracking, Multiple object tracking, Visual object tracking, Deep learning convolutional network, Bounding box.*

[@] *E-mail:* **virendra.kushwah@vitbhopal.ac.in**
[#] *E-mail:* **jyoti.parashar123@gmail.com**
[$] *E-mail:* **jainrituraj@yahoo.com**
[^] *E-mail:* **tarannum.bloch@marwadieducatuon.edu.in**

1. Introduction

Automated vehicles and robots are rapidly navigating using sensing devices such as camcorders. The most advanced technologies rely on computer image processing approaches along with hardware-based alternatives. In exceptional circumstances, such strategies have revealed to be inefficient. For instance, hardware-based alternatives do not operate efficiently in adverse climate such as fog, storm, hail, rain etc. Visual techniques are subject to changes like instability of camera and changes in optical conditions. In daily situation, we need a large amount of data about our surroundings. We must comprehend how well the items move in reference to the camera. It might also allow us to identify the interaction of items. For instance, in the event of a driverless car, understanding of pedestrian interactions will benefit in accurately assessing pedestrian movement. This prognosis will ultimately assist the driverless car in making informed decisions on a busy route [1].

Single object tracking focuses on monitoring a specified target throughout a video sequence, starting with its identification in the initial frame and tracking it across subsequent frames. It operates by detecting objects using various features such as manual or convolutional attributes, with deep learning being a key contributor due to its powerful feature extraction capabilities. Common phases include feature extraction from input images, identifying regions of interest through techniques like particle filter forecasting or sliding window methods [2], building the tracking model, which can be generative or discriminative, and continuously updating the model to adapt to changes in the environment and background.

2. Literature Review

Multi object audio-visual tracking is a tremendously rapidly developing field of study with potential in several diverse sectors. They include everything from automated categorization to adaptive digital systems to surveillance footage. A great monitoring system can act as a strong foundation for supporting additional approaches, like movement detection algorithms, facial or speech identifiers, posture identifiers, and environment analytical techniques, particularly during the final scenario. Several studies on the human tracking mechanism over the past few years have relied heavily on perception. A tracker that uses the kernelized correlation filter (KCF) to identify the goal across several levels was introduced by Pang et al. in [4]. Combining HOG, colour identification, and localized depth sequence elements went into its creation. A support bot built on the pedestrian detection technique has been suggested by Chi et al. [5] to carry out the task of human tracking. As for individual monitoring method, Iswanto et al. [6] used a mean-shift with particle-Kalman filter.

The Object Tracking Benchmark (OTB) serves as a standard for evaluating visual tracking techniques. It includes datasets like VOT2015, VOT2016, and VOT2017. These datasets assess tracking performance based on accuracy, robustness, and Expected Average Overlap (EAO). Notably, VOT2017 focuses on short-term causal trackers, where trackers are reset upon target loss and cannot use previous frames for current frame location determination [7]. Danelljan et al. [8] devised ECO trackers for object monitoring, employing efficient convolutions with factorization to improve computational efficiency.

Luca Bertinetto et al. [9] discovered that the prior approach of the model focuses on the tracking object's satellite data that is not resistant to modification. Furthermore, when distortion and blurriness are present, the application of color characteristics can identify and track the object successfully. The color attributes are not effectively displayed as the illumination changes.

3. Methodology

The single-object tracking algorithm comprises four essential components: the Characteristic Model, Movement Model, Perception Model, and Real-time Updating Process. These components synergize to fulfill distinct roles within the tracking process. The Characteristic Model utilizes image modification techniques to gather descriptive evidence of the target's appearance, forming the basis for the Perception Model. Meanwhile, the Movement Model offers potential locations where the object may appear based on contextual information. The Perception Model then predicts the object's location using inputs from the Characteristic and Movement Models. Finally, the Real-time Updating Process ensures that the model responds to changes in bounding boxes, preventing degradation of the estimated and observed models.

The object tracking procedure involves four phases: Target Setting Up, Appearance Modeling, Motion Analysis, and Target Placement. Firstly, Target Setting Up involves identifying the target of interest and creating a bounding box in the opening frame, followed by anticipating its coordinates in subsequent frames. Appearance Modeling captures changes in the target's appearance due to varying backgrounds, ensuring accurate tracking amidst such variations. Finally, Target Placement utilizes motion analysis to predict the most probable location of the target, facilitating precise tracking.

This segment has discussed tracking algorithms. These strategies differ based on whether we are tracking a single person or a group of people. To count the people, we used the centroid (c) of the location.

1. **Individual Tracking** - Only the first and last c coordinates are required. These dimensions are evaluated by comparing with the image›s dimensions to identify if an individual is entering or exiting as shown in Figure 1 and Table 1.

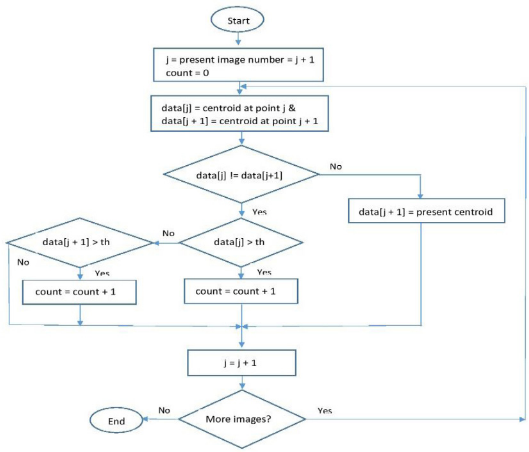

Figure 1

Flowchart for Single object tracking

Table 1

Situations for tracking a single person

Situations	Outcome
(v - Ci) > 0 & (v - Cf) < 0	Individual enters the frame
(v - Ci) < 0 & (v - Cf) > 0	Individual exits the frame
(v - Ci) < 0 & (x - Cf) < 0	No movement
(v - Ci) > 0 & (x - Cf) > 0	No movement

2. **Multiple Tracking** - The tracking method first establishes how many individuals will be in every frame, and then it tracks each individual's identification through one frame to another until it leaves the frame. Once movement is detected, the segmentation section is reached. The last bounding box of the blob is generated by integrating blob that correspond to the identical element. Bounding boxes are included, and their centroid is calculated if they are larger than a predetermined minimum dimension.

However, when people shift places in the frame, deciding which array to give the next centroid value to, becomes difficult. A decision is made based on the context of the data in each array, and the updated value is assigned accordingly as shown in Figure 2.

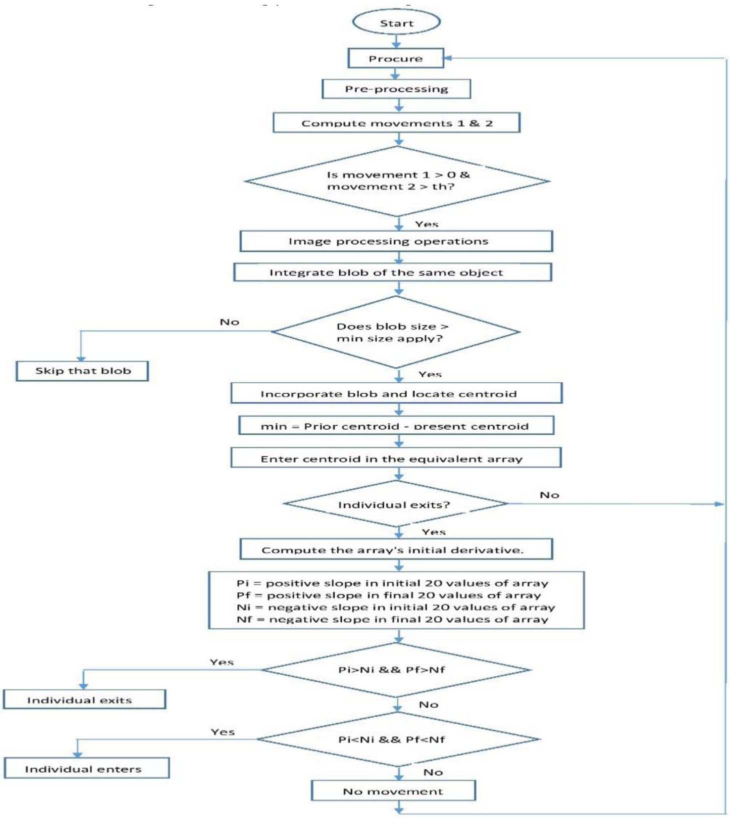

Figure 2

Flowchart for multiple object tracking

Eventually, we present the frames with the individuals and the area of focus outlined together with the number of individuals now entering the zone as shown in Table 2 and 3.

Table 2

Situations for tracking a 1^{st} individual in multiple object tracking

Situations	Outcome
$(i1')_p > (i1')_n \& (f1')_p > (f1')_n$	Individual enters the frame
$(i1')_p < (i1')_n \& (f1')_p < (f1')_n$	Individual exits the frame
$(i1')_p > (i1')_n \& (f1')_p < (f1')_n$	No movement
$(i1')_p < (i1')_n \& (f1')_p > (f1')_n$	No movement

Table 3

Situations for tracking 2^{ndt} individual in multiple object tracking

Situations	Outcome
$(i2')_p > (i2')_n \& (f2')_p > (f2')_n$	Individual enters the frame
$(i2')_p < (i2')_n \& (f2')_p < (f2')_n$	Individual exits the frame
$(i2')_p > (i2')_n \& (f2')_p < (f2')_n$	No movement
$(i2')_p < (i2')_n \& (f2')_p > (f2')_n$	No movement

3.2 *Computing the Bounding Box*

For every frame, segmentation mask was used to instantly calculate the definitive bounding box representing the true object location.

The criteria and limitations to ensure that the bounding box accurately encloses the target pixels while minimizing the inclusion of irrelevant background pixels are defined by the cost function:

$$\arg\_\min\left\{ CF(v_b) = \beta \bar{\sum_{x \notin B(v_b)}} \_[S(x) > 0] + \bar{\sum_{x \in B(v_b)}} \_[S(x) == 0] \right\},$$

$$\frac{1}{S_p} \bar{\sum_{x \notin B(v_b)}} [S(x) > 0] < \delta_p \ , \frac{1}{|B(v_b)|} \bar{\sum_{x \in B(v_b)}} [S(x) == 0] < \delta v_b \qquad (1)$$

Where,

CF denotes cost function, v_b - denotes the vector of the bounding box, $B(v_b)$ is the bounding box, S represents segmentation mask, S_p is the number of entity pixels, β is the tightness of bounding box. $|.|$ represents cardinality, When the condition in the [.] operator is satisfied, it returns 1% of omitted entity pixel resolution which included background pixels limited by δ_p and δv_b, respectively [13].

Figure 3

Single individual tracking for time t (a) to t+3 (d)

Figure 4

Multiple object tracking for time t (a) to t+10 (k)

4. Results

In Figure 3, we have captured the above image for individual tracking at different time intervals. For calculating error rates, we count on a sequence lasting 4 frames. For frames t to t+3, 1 object is visible which is being tracked, so there is 0% miss for this object. Figure 4 image was captured for multiple object tracking at different time intervals for calculating error rates, we assume that the sequences are of 11 frames. In the frames t to t+3, 6 objects are observable out of which only 2 are being tracked and 4 are not being tracked, so there is 100% miss for these 4 objects. For frames t to t+3, one object being tracked leaves the frame. For frames t+4 to t+9, second object is being tracked which leaves the frame after t+9.

5. Performance Comparison of Various Trackers

Area under curve (AUC): At the IOU threshold, the Precision - Recall curve's area under the curve is assessed. A significant region under the PR graph is characterised by good recall and high precision. Naturally, the PR results are plotted as a zig-zag pattern. This indicates that it isn't diminishing monotonically. Table 4 shows comparisons on various benchmarks

$$AUC = \int_0^1 \_\, p(r)dr \tag{2}$$

Table 4

Performance comparisons on various benchmarks

Trackers	Authors	OTB2015 AUC/ Success Rate	VOT2015 A R EAO	VOT2016 A R EAO	VOT2017 A R EAO	LASOT AUC/ Success Rate
C-COT	M. Danelljan et al. [14]	0.671	0.54 0.82 0.303	0.539 0.238 0.331	0.49 0.32 0.267	-
Staple	L. Bertinetto et al. [9]	0.578	0.57 1.39 0.300	0.544 0.378 0.295	0.52 0.69 0.169	0.243
MDNet	H. Nam et al. [15]	0.678	0.60 0.69 0.378	0.541 0.337 0.257	- - -	0.397
CFNet	J. Valmadre et al. [16]	0.568	- -	- -	- - -	0.275
ECO	M. Danelljan et al. [8]	0.691	- - -	0.550 0.200 0.375	0.48 0.27 0.280	0.324
SiamFC	L. Bertinetto et al. [17]	0.582	0.533 0.88 0.289	0.530 0.460 0.235	0.50 0.59 0.188	0.336
SiamRPN	B. Li et al. [18]	0.636	0.58 1.13 0.319	0.560 0.260 0.344	0.46 0.244	-

6. Conclusion

The research has demonstrated successful outcomes in both individual tracking and tracking for multiple objects. Despite limitations such as variations in illumination levels and backgrounds, all algorithms have been effectively designed to yield precise findings within specified constraints. Object Visual Tracking (OVT) analyses have revealed significant insights: MDNet achieved the highest AUC score of 0.397. Additionally, SiamRPN stands out for its remarkable processing speed of 200 frames per second while maintaining high performance levels. Across evaluations, SiamRPN consistently outperforms other trackers in success and precision metrics, showcasing its superiority in object tracking applications.

References

[1] Soleimanitaleb, Z., & Keyvanrad, M.A. Single Object Tracking: A Survey of Methods, Datasets, and Evaluation Metrics (2022). ArXiv, abs/2201.13066.

[2] Haider, Mohammad & Upreti, Kamal & Nasir, Mohammad & Alam, Mohammad & Sharma, A K. Addressing image and Poisson noise deconvolution problem using deep learning approaches. *Computational Intelligence Journal* (2022). 10.1111/coin.12510.

[3] Pang, L., et al.: A human-following approach using binocular camera. In: 2017 IEEE International Conference on Mechatronics and Automation (ICMA), pp. 1487–1492 (2017).

[4] Chi, W., et al.: A gait recognition method for human following in service robots. *IEEE Trans. Syst. Man, Cybern. Syst*. 48(9), 1429–1440, (2018).

[5] Iswanto, I.: Visual object tracking based on mean-shift and particle-Kalman filter. *Procedia Comput. Science A., Li, B*. 116, 587– 595 (2017).

[6] K. Upreti, S. Arora, A. K. Sharma, A. K. Pandey, K. K. Sharma and M. Dayal, "Wave Height Forecasting Over Ocean of Things Based on Machine Learning Techniques: An Application for Ocean Renewable Energy Generation," in *IEEE Journal of Oceanic Engineering*, doi: 10.1109/JOE.2023.3314090.

[7] Danelljan, Martin & Bhat, Goutam & Khan, Fahad & Felsberg, Michael. *ECO: Efficient Convolution Operators for Tracking* (2017). 6931-6939. 10.1109/CVPR.2017.733.

[8] Bertinetto, Luca & Valmadre, Jack & Golodetz, Stuart & Miksik, Ondrej & Torr, Philip. Staple: Complementary Learners for Real-Time Tracking (2016). 10.1109/CVPR.2016.156.

[9] Kristan, M. et al. The Visual Object Tracking VOT2016 Challenge Results. In: Hua, G., Jégou, H. (eds) Computer Vision – ECCV 2016 Workshops. ECCV 2016. Lecture Notes in Computer Science(), vol 9914. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48881-3_54

[10] M. Danelljan, A. Robinson, F.S. Khan, M. Felsberg, Beyond correlation filters: Learning continuous convolution operators for visual tracking, in: Proc. 2016 European Conf. Computer Vision, pp. 472–488 (2016).

[11] H. Nam and B. Han, Learning multi-domain convolutional neural networks for visual tracking, in: Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition, pp. 4293–4302 (2016).

[12] J. Valmadre, L. Bertinetto, J.F. Henriques, A. Vedaldi, and P.H.S. Torr, End-toend representation learning for correlation filter based tracking, in: Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition, pp. 5000– 5008 (2017).

[13] L. Bertinetto, J. Valmadre, J.F. Henriques, A. Vedaldi, P.H.S. Torr, Fully convolutional siamese networks for object tracking, in: Proc. 2016 European Conf. Computer Vision, pp. 850–865 (2016).

[14] Arora, R., Agrawal, A., Arora, R., Poonia, R. C., & Madaan, V. Prediction and forecasting of COVID-19 outbreak using regression and ARIMA models. *Journal of Interdisciplinary Mathematics*, 24(1), 227–243 (2021).