

RESEARCH ARTICLE

SOTA: Sequential Optimal Transport Approximation for Visual Tracking in Wild Scenario

SEONGHAK LEE¹, JISOO PARK², RADU TIMOFTE^{1,3}, (Member, IEEE),
AND JUNSEOK KWON^{1,2}, (Member, IEEE)

¹School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, South Korea

²Graduate School of Artificial Intelligence, Chung-Ang University, Seoul 06974, South Korea

³Computer Vision Laboratory, CAIDAS & IFI, University of Würzburg, 97070 Würzburg, Germany

Corresponding author: Junseok Kwon (jskwon@cau.ac.kr)

This work was supported in part by the Chung-Ang University Research Scholarship Grants, in 2023; in part by the Institute of Information Communications Technology Planning Evaluation (IITP) Grant funded by Korean Government through the Ministry of Science and ICT (MSIT) [Artificial Intelligence Graduate School Program (Chung-Ang University)] under Grant 2021-0-01341; and in part by the Alexander von Humboldt Foundation.

ABSTRACT In this study, we introduce a probabilistic visual tracking method tailored for wild scenarios, where tracking environments experience abrupt changes over time. In probabilistic visual tracking, particularly when utilizing sequential Monte Carlo (MC) sampling, the careful choice of a proposal function is critical for attaining precise tracking results, where optimal transport techniques can assist in reducing the variance from important sampling and inducing a suitable proposal function. However, if a tracked object undergoes significant changes in appearance over time, it creates a large discrepancy between the proposal function and its target distribution. This situation presents the topological and representational challenges for conventional optimal transport techniques. To address this problem, the proposed visual tracker leverages the benefits of both MC sampling and optimal transport and presents the Sequential Optimal Transport Approximation (SOTA) visual tracker. For this, our method involves transforming the proposal function into its target distribution across successive temperature steps through sequential MC sampling. These MC steps can reduce the topological and representational burden on the optimal transport. Within each successive temperature step, one distribution is projected onto another one via optimal transport. By using optimal transport, the method can mitigate the variance from importance sampling. The experimental results demonstrate that the proposed method considerably outperforms other state-of-the-art methods across several benchmark dataset, particularly in tracking environments where abrupt changes occur.

INDEX TERMS Visual tracking, optimal transport, sequential Monte Carlo.

I. INTRODUCTION

Visual tracking is a fundamental technique that can be utilized to predict the trajectories of target objects and employed in diverse computer vision problems, such as autonomous driving [1], medical imaging [2], robotics [3], and surveillance [4]. In many studies, visual tracking has been formulated within the sequential Monte Carlo (MC) sampling framework [5], where the proposal function is employed

to approximate the posterior distribution of target object configurations. Recently, visual tracking has improved its performance by establishing the precise relationship between the proposal function and its target distribution through optimal transport techniques [6], [7].

A. MOTIVATION

1) MOTIVATION ON EXACT MAPPING

Optimal transport [6] learns a differentiable invertible mapping between the proposal distribution and the posterior distribution via a tractable Jacobian. Then, it can accurately

The associate editor coordinating the review of this manuscript and approving it for publication was Anandakumar Haldorai¹.

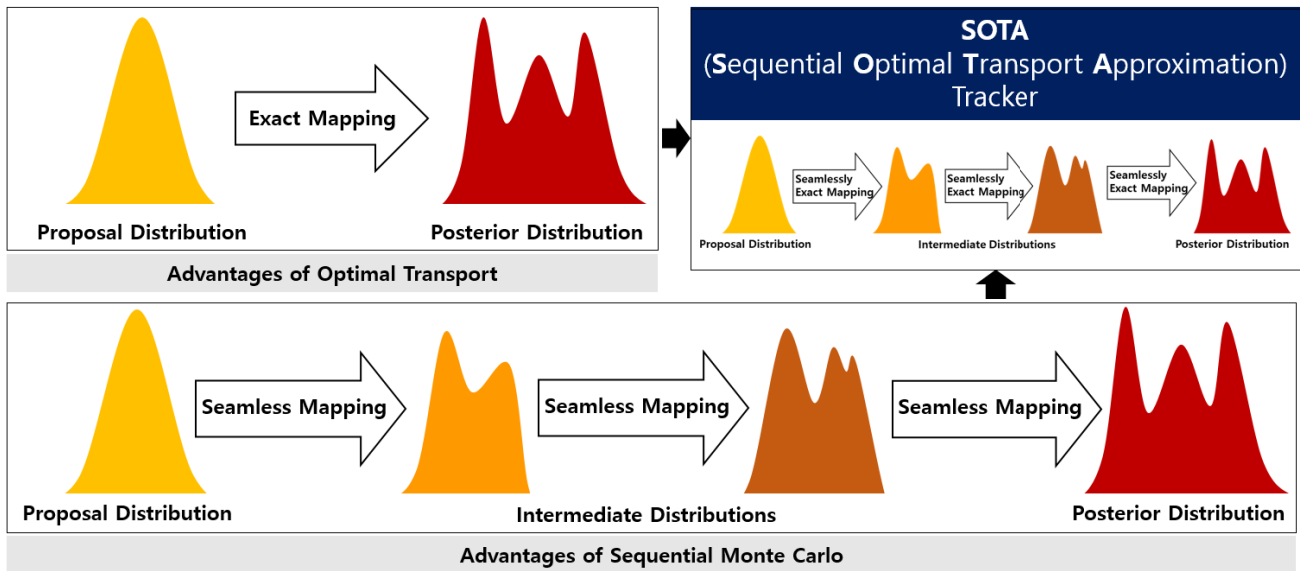


FIGURE 1. Basic idea of the proposed SOTA tracker. We combine optimal transport with the sequence MC sampling and leverage their complementary advantages (*i.e.*, exact mapping + seamless mapping) and present the Sequential Optimal Transport Approximation (SOTA) tracker that can robustly track the target objects in real-world environments where abrupt changes occur.

estimate the posterior distribution from the proposal distribution. However, if there is a substantial discrepancy between the two distributions and significant topological mismatches exist, finding the precise mapping becomes intractable using a naive optimal transport technique. *This problem can be addressed by the following sequential MC sampling.*

2) MOTIVATION ON SEAMLESS MAPPING

Sequential MC sampling [5] generates numerous adjacent and intermediate distributions between the proposal function and the posterior distribution through successive temperature steps. By doing this, it can mitigate the significant mismatch between two distributions during the process. However, sequential MC can easily lead to estimators with high error or require the use of many temperatures to approximate high-dimensional posterior distributions. *This problem can be addressed by the aforementioned optimal transport.*

B. DIFFERENCES

1) SEAMLESSLY EXACT MAPPING

In this study, inspired by [8], we combine optimal transport with the sequence MC sampling and leverage their complementary advantages. On one hand, optimal transport helps to decrease the variance resulting from importance sampling by establishing an exact mapping between the two distributions. On the other hand, sequential MC sampling can alleviate the topological and representational burdens on the mapping, particularly in cases of large discrepancies between the two distributions. While existing works typically use only one component—either optimal transport or sequential MC sampling—for visual tracking, we utilize both optimal transport and the sequence MC sampling, exploiting the

complementary advantages of each approach. Building on this finding, we introduce the Sequential Optimal Transport Approximation (SOTA) visual tracker. Figure 1 illustrates differences between the proposed method with existing approaches.

C. CONTRIBUTIONS

The main contributions of our method are as follows:

- The optimal transport module of SOTA learns an exact mapping between the proposal function and the posterior distribution through a tractable Jacobian, ensuring the estimated posterior distribution is unbiased. This leads to robust visual tracking.
- The sequential MC sampling module of SOTA learns a seamless mapping between the proposal function and the posterior distribution through numerous intermediate distributions, mitigating the significant mismatch between adjacent distributions. This leads to accurate visual tracking.
- By combining these two modules, the proposed SOTA tracker can track the target objects even under significant changes in the tracking environment.

II. RELATED WORK

A. VISUAL TRACKING UNDER REAL-WORLD ENVIRONMENTS

RTracker [9] conducted self-recovery under real-word tracking settings by utilizing a tree-structured memory to handle frequently missing trajectories. NetTrack [10] introduced a fine-grained network for dynamicity-aware association in highly dynamic tracking settings. PVT++ [11] proposed a latency-aware visual tracking method for online tracking



FIGURE 2. Pipeline of the Proposed SOTA Tracker. Given the target's configuration at the current frame t , the seamless mapping module of the proposed tracker approximates its configuration at the next frame $t + 1$ via sequential Monte Carlo sampling explained in Section III-B. Subsequently, the exact mapping module refines this approximated configuration by optimally adjusting it using the optimal transport technique described in Section III-A. This process iterates multiple times to accurately determine the target's configuration at the next frame $t + 1$, as illustrated in Section III-C.

settings by designing a lightweight network architecture. ROMTrack [12] combines inherent template and hybrid template features to robustly describe target object appearance in noisy visual tracking settings. CiteTracker [11] described the target object using a descriptive text that includes its class and attribute information for various tracking environments.

In contrast to these methods, which are built upon a novel architecture design in deep neural networks, our visual tracker SOTA attempts to overcome severe changes in real-world visual tracking environments in a mathematical manner.

B. VISUAL TRACKING BASED ON OPTIMAL TRANSPORT

For optimal transport from proposal to target distributions, the Hamiltonian proposal-based approach [13] balances exploitation and exploration sampling to locally and globally represent the target distribution. The Metropolis-adjusted Langevin algorithm [14] achieved a correct stationary distribution by reducing accumulation errors. A transport map accelerated proposal [15] explicitly measures the Wasserstein distance to minimize differences between initial and target distributions. However, the aforementioned proposal functions [13], [14] are suboptimal, and the transport map accelerated proposal [15] can easily fail if there is a substantial discrepancy between the two distributions. WTP [7] introduced an optimal transport proposal approach that combines an optimal proposal function with Markov chain Monte Carlo in the Wasserstein space.

In contrast to WTP, our method enables more accurate optimal transport within a sequential Monte Carlo framework, facilitating optimal transport across successive intermediate distributions and mitigating the large gap between adjacent distributions.

C. VISUAL TRACKING VIA MONTE CARLO SAMPLING

VAE-MC [16] has inferred tight evidence of lower bounds via variational autoencoding and accurately estimated target posterior distributions. EWLCFDP [17] augmented Wang-Landau Monte Carlo by using multi-scope marginal likelihood to increase the acceptance ratio of samples in promising regions, thereby obtaining a more reliable posterior distribution. A-MDPF [18] leveraged training data to accurately learn particle-based representations of posterior distributions for visual tracking.

In contrast to these methods, which struggle to estimate accurate posterior distributions when there is a large discrepancy between the proposal and target posterior distributions, we presented an optimal proposal function that accurately describes the target distribution, even amid significant changes in tracking environments.

III. PROPOSED METHOD

Section III-A details the optimal transport technique for exact mapping, while Section III-B explains the sequential Monte Carlo sampling algorithm for seamless mapping. Section III-C introduces the proposed SOTA tracker, which integrates exact mapping with seamless mapping. Fig. 2 summarizes the pipeline of the proposed SOTA tracker.

A. EXACT MAPPING: OPTIMAL TRANSPORT

We define a transport map $\mathcal{T}_\#$ that transforms the proposal distribution γ to the target posterior distribution π : $\pi = \mathcal{T}_\#(\gamma)$. Then, associated random variables can be sampled from these distributions, where $x_{\text{proposal}} \in \mathbb{R}^N \sim \gamma$ and $x_{\text{posterior}} \in \mathbb{R}^N \sim \pi$, where the relationship between these two random variable is defined using Φ :

$$x_{\text{posterior}} = \Phi(x_{\text{proposal}}). \quad (1)$$

The optimal mapping function Φ^* can be found as the function that minimizes the discrepancy between x_{proposal} and $x_{\text{posterior}}$, as follows.

$$\Phi^* = \arg \min_{\Phi} \int_{\mathbb{R}^N} c(x_{\text{proposal}}, \Phi(x_{\text{proposal}})) d\gamma(x_{\text{proposal}}), \quad (2)$$

where the discrepancy can be measured using a cost function $c(x_{\text{proposal}}, \Phi(x_{\text{proposal}}))$. This cost function can be simply implemented using the l_2 norm.

$$c(x_{\text{proposal}}, x_{\text{posterior}}) = |x_{\text{proposal}} - x_{\text{posterior}}|^2. \quad (3)$$

Then, we need to design the optimal mapping function Φ^* to satisfy the following three properties.

- Property 1: Φ^* minimizes the discrepancy between x_{proposal} and $x_{\text{posterior}}$ in (1).
- Property 2: Φ^* reduces the computational cost of the integral computation in (2).

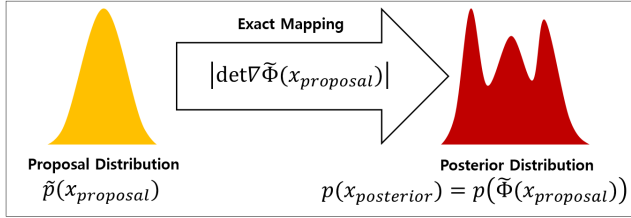


FIGURE 3. Exact Mapping via Optimal Transport. The proposal distribution is exactly related to the posterior distribution via the Jacobian determinant: $|\det \nabla \tilde{\Phi}(x_{propos})|$, i.e., $\tilde{p}(x_{propos}) = p(\tilde{\Phi}(x_{propos})) |\det \nabla \tilde{\Phi}(x_{propos})|$ in (4).

Thus, we search for the approximated mapping function $\tilde{\Phi}$ that fulfills all these properties via MCMC sampling [19].

For this purpose, we define the probability distribution p associated with $x_{posterior}$, i.e., $p(\tilde{\Phi}(x_{propos}))$ by (1). Then, the probability distribution p associated with x_{propos} , i.e., $\tilde{p}(x_{propos})$, can be defined as follows:

$$\tilde{p}(x_{propos}) = p(\tilde{\Phi}(x_{propos})) |\det \nabla \tilde{\Phi}(x_{propos})|, \quad (4)$$

where $\det \nabla \tilde{\Phi}(x_{propos})$ indicates the Jacobian determinant of $\tilde{\Phi}(x_{propos})$. In (4), $\tilde{p}(x_{propos})$ approaches to the optimal $p(x_{propos})$ as $x_{posterior}$ converges to $\tilde{\Phi}(x_{propos})$. Thus, we attempt to find $\tilde{\Phi}$ that minimize the discrepancy between the two distributions $\tilde{p}$ and p using the Kullback-Leibler divergence D_{KL} .

$$\begin{aligned} D_{KL}(p \parallel \tilde{p}) &= \mathbb{E}_p [\log p(x_{propos}) - \log \tilde{p}(x_{propos})] \\ &= \mathbb{E}_p \left[\log p(x_{propos}) - \log p(\tilde{\Phi}(x_{propos})) \right. \\ &\quad \times \left. |\det \nabla \tilde{\Phi}(x_{propos})| \right], \\ &= \mathbb{E}_p \left[\log p(x_{propos}) - \log p(\tilde{\Phi}(x_{propos})) \right. \\ &\quad \left. - \log |\det \nabla \tilde{\Phi}(x_{propos})| \right], \end{aligned} \quad (5)$$

where the second equality holds due to (4). Then, Property 1 can be satisfied by finding $\tilde{\Phi}$ that can minimize $D_{KL}(p \parallel \tilde{p})$ in (5). The objective function is described as follows.

$$\begin{aligned} \Phi^* &= \arg \min_{\tilde{\Phi}} \mathbb{E}_p \left[\log p(x_{propos}) - \log p(\tilde{\Phi}(x_{propos})) \right. \\ &\quad \left. - \log |\det \nabla \tilde{\Phi}(x_{propos})| \right] \\ &= \arg \min_{\tilde{\Phi}} \mathbb{E}_p \left[-\log p(\tilde{\Phi}(x_{propos})) \right] \\ &\quad - \mathbb{E}_p \left[\log |\det \nabla \tilde{\Phi}(x_{propos})| \right]. \end{aligned} \quad (6)$$

We can omit $\log p(x_{propos})$ in the second equality because of the independence between $\log p(x_{propos})$ and $\tilde{\Phi}$.

To find the solution in (6) while satisfying Property 2, we employ the MCMC sampling. For this, we represent the approximated transport map $\tilde{\Phi}(x_{propos})$ as $\tilde{p}(x_{propos})$ in (4)

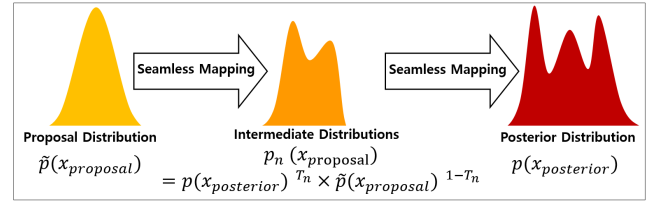


FIGURE 4. Seamless Mapping via Sequential Monte Carlo Sampling. The intermediate distributions are obtained via the successive temperature steps: $p_n(x_{propos}) = p(x_{posterior})^{T_n} \times \tilde{p}(x_{propos})^{1-T_n} = p_N(x_{propos})^{T_n} \times p_1(x_{propos})^{1-T_n}$ in (9).

and update $\tilde{\Phi}(x_{propos})$ via Monte Carlo sampling:

$$\begin{aligned} q(x'_{propos}; x_{propos}) &= f(\tilde{\Phi}(x'_{propos}); \tilde{\Phi}(x_{propos})) \\ &\quad \times |\det \nabla \tilde{\Phi}(x'_{propos})|, \end{aligned} \quad (7)$$

where a new sample x'_{propos} is proposed based on the current sample x_{propos} using the proposal function $q(x'_{propos}; x_{propos})$.¹ $f(\tilde{\Phi}(x'_{propos}); \tilde{\Phi}(x_{propos}))$ is assumed to follow a normal distribution with variance σ^2 , i.e., $\mathcal{N}_{\sigma^2}(x'_{propos}; x_{propos})$. Subsequently, we accept or reject the new sample x'_{propos} based on the following acceptance ratio:

$$\min \left[1, \frac{\tilde{p}(x'_{propos}) q(x_{propos}; x'_{propos})}{\tilde{p}(x_{propos}) q(x'_{propos}; x_{propos})} \right], \quad (8)$$

where the accepted K samples, $x_{propos}^{(1)}, \dots, x_{propos}^{(K)}$, approximately describe $\tilde{\Phi}(x_{propos})$.

B. SEAMLESS MAPPING: SEQUENTIAL MONTE CARLO SAMPLING

The proposal distribution is exactly related to the posterior distribution via the Jacobian determinant, i.e., $|\det \nabla \tilde{\Phi}(x_{propos})|$, as shown in Fig.3. However, if there is a large distributional gap between the proposal and posterior distributions, resulting in significant topological mismatches, computing the accurate Jacobian determinant becomes intractable. Thus, we need to additionally implement seamless mapping through sequential MC sampling [5].

To achieve this, we generate N adjacent distributions between the proposal and posterior distributions: $\{p_n(x_{propos})\}_{n=1}^{n=T}$, where $p_1(x_{propos}) = \tilde{p}(x_{propos})$ and $p_N(x_{propos}) = p(x_{posterior})$. The two adjacent distributions, $p_n(x_{propos})$ and $p_{n+1}(x_{propos})$ are similar. Thus, the distributional discrepancy between these distributions can be mitigated. We can find these intermediate distributions through successive temperature steps, as shown in Fig.4:

$$p_n(x_{propos}) = p_N(x_{propos})^{T_n} \times p_1(x_{propos})^{1-T_n}, \quad (9)$$

¹This can be implemented based on (4).

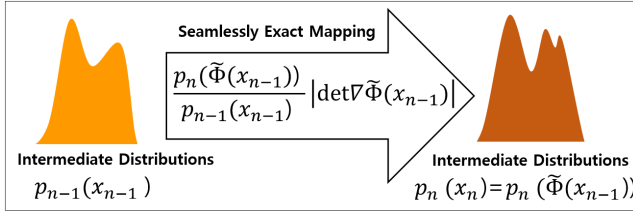


FIGURE 5. Seamlessly Exact Mapping via Sequential Optimal Transport Approximation. Two adjacent distributions are seamlessly and exactly mapped via the following weighting $\frac{p_n(\tilde{\Phi}(x_{n-1}^{1:M}))}{p_{n-1}(x_{n-1}^{1:M})} |\det \nabla \tilde{\Phi}(x_{n-1}^{1:M})|$ in (12).

where the proposal and posterior distributions are mixed based on the temperature T_j , resulting in the j -th intermediate distribution. We schedule the temperature T_j to smoothly transform the proposal distribution into the posterior distribution:

$$1 = T_0 > T_1 > \dots > T_{N-1} > T_N = 0. \quad (10)$$

To describe each intermediate distribution $p_n(x_{propos})$, we utilize M sample points, $x_n^1, \dots, x_n^M$, along with their corresponding weights $w_n^1, \dots, w_n^M$. Hereafter, we denote the m -th sample of x_{propos} for the n -th intermediate distribution as x_n^m . Then, the weight w_n aims to calibrate the discrepancy between two adjacent distributions, $p_{n-1}(x_{propos})$ and $p_n(x_{propos})$:

$$w_n(x_n^{1:M}) = \frac{p_n(x_n^{1:M})}{p_{n-1}(x_n^{1:M})}. \quad (11)$$

We denote the weight of x_n^m as w_n^m , where $\sum_{m=1}^M w_n^m = 1$.

However, this sequential MC sampling can easily lead to estimators with high sample variances due to its approximation process using sample points in (11). This problem can be mitigated by combining it with the aforementioned optimal transport.

C. SEAMLESSLY EXACT MAPPING: SEQUENTIAL OPTIMAL TRANSPORT APPROXIMATION

We combine optimal transport with the sequence MC sampling to leveraging their complementary advantages. Optimal transport helps to reduce the variance resulting from the use sample points in (11) by providing an exact mapping between the two distributions. This objective can be achieved by transforming (11) into the following computation, as shown in Fig.5:

$$w_n(x_n^{1:M}) = \frac{p_n(\tilde{\Phi}(x_{n-1}^{1:M}))}{p_{n-1}(x_{n-1}^{1:M})} |\det \nabla \tilde{\Phi}(x_{n-1}^{1:M})|, \quad (12)$$

where the sequential MC sampling can alleviate the topological and representational challenges of the exact mapping, specifically the term $|\det \nabla \tilde{\Phi}(x_{n-1}^{1:M})|$, by focusing only on the similar adjacent distributions $p_{n-1}(\cdot)$ and $p_n(\cdot)$.

Algorithm 1 Sequential Optimal Transport Approximation (SOTA) Visual Tracking

Input x : target state in the current frame

Output x_N^* : target state in the next frame

```

1: [Seamless Mapping]
2:  $N$ : the number of intermediate distributions
3: for  $n = 1$  to  $N$  do
4:    $\tilde{p}_n(x_n) = p_N(x_N)^{T_n} \times p_1(x_1)^{1-T_n}$  in (9)
5:   [Exact Mapping]
6:    $K$ : the number of sample points for approximating the
       optimal transport map
7:   for  $k = 1$  to  $K$  do
8:      $q(x'_n; x_n) = f(\tilde{\Phi}(x'_n); \tilde{\Phi}(x_n)) \times |\det \nabla \tilde{\Phi}(x'_n)|$ 
       in (7)
9:      $u \sim \mathcal{U}[0, 1]$ 
10:    if  $u \leq \min \left[ 1, \frac{\tilde{p}(x'_n)q(x_n; x'_n)}{\tilde{p}(x_n)q(x'_n; x_n)} \right]$  in (8) then
11:       $x_n^{(k)} = x'_n$ 
12:    else
13:       $x_n^{(k)} = x_n$ 
14:    end if
15:     $x_n = x_n^{(k)}$ 
16:    Find  $\Phi^*$  by  $\arg \min_{\tilde{\Phi}} \mathbb{E}_p \left[ -\log p(\tilde{\Phi}(x_n^{(k)})) \right] -$ 
        $\mathbb{E}_p \left[ \log |\det \nabla \tilde{\Phi}(x_n^{(k)})| \right]$  in (6)
17:  end for
18:  for  $m = 1$  to  $M$  do
19:    Estimate  $p_n(x_n^m)$  by  $w_n(x_n^m) \frac{p_{n-1}(x_{n-1}^m)}{|\det \nabla \tilde{\Phi}(x_{n-1}^m)|}$  in (12)
20:    Find  $x_n^*$  by  $\arg \min_{x_n^m} p_n(Y|x_n^m)$  in (14)
21:  end for
22: end for

```

In (12), we evaluate $p_n(x_n^{1:M})$ based on the observation Y , where $p_n(x_n^{1:M}|Y) \approx p_n(Y|x_n^{1:M})$ under the assumption of uniform prior probability on x_{n-1} . We calculate the likelihood $p_n(Y|x_n^{1:M})$ by measuring similarity between observed and template features. Here, the observed features $Y(x_n^{1:M})$ are derived from the image observation Y , described by the estimated state x_n , which includes the (x, y) coordinates of the target object and its scale. The template features $Y^{template}$ are derived from the image observation, which are manually determined by us in an initial frame.

In particular, we compute $p_n(Y|x_n^{1:M})$ as follows:

$$p_n(Y|x_n^m) = e^{-f(Y(x_n^m), Y^{template})}, \quad (13)$$

where the similarity function $f(\cdot)$ denotes the correlation operator described in [20]. We utilize the deep features extracted from the first and last layer before the correlation operator, apply equal weighting to both layers, and average the features to obtain $Y(x_n^m)$ and $Y^{template}$. Then, the optimal state x_n^* can be found as follow:

$$x_n^* = \arg \min_{x_n^m} p_n(Y|x_n^m). \quad (14)$$

TABLE 1. Performance according to the different number of sample points for approximating the transport map, K in (8), in terms of success AUC. The best results were indicated in boldface.

K in (8)	LaSOT [21]	UAV123 [22]	OTB100 [23]	NFS [24]
10	67.3	71.4	70.1	62.8
20	68.1	71.5	72.3	66.6
30	68.1	71.5	72.4	66.5
35	68.2	71.5	72.4	66.6
40	68.2	71.6	72.4	66.7

Algorithm 1 shows the whole process of the proposed sequential optimal transport approximation (SOTA) visual tracking.

IV. EXPERIMENTS

A. EXPERIMENTAL DETAILS

To justify the effectiveness of the proposed method, we compared the proposed visual tracker with recent state-of-the-art trackers: SiamRPN++ [25], PrDiMP [26], DiMP [27], SuperDiMP [26], [27], Stark [28], KeepTrack [29], CIA [20], TransT [30], TrSiam [31], TrDiMP [31], D3S [32], GFS-DCF [33], SBT [34], ToMP [35], Siam R-CNN [36], Alpha-Refine [37], STMTrack [38], DTT [39], Siamfc++ [40], CRACT [41], ATOM [42], UPDT [43], DPMT [44], TRAT [45], EWLCFDP [17], WTP [7], CiteTracker [11], and SiamAttN [46]. Please note that SuperDiMP is a tracking method that combines the bounding-box regressor of PrDiMP [26] with the DiMP classifier [27]. We selected these visual trackers because they are recent and demonstrate state-of-the-art performance, as mentioned by ToMP [35].

For comparisons, we used standard benchmark visual tracking datasets: OTB100 [23], LaSOT [21], UAV123 [22], VOT2020 [47], and NFS [24]. The OTB100 dataset includes 100 test videos with diverse attributes, such as motion blur, occlusions, scale variation, low resolution, illumination change, out-of-plane rotation, in-plane rotation, out-of-view, background clutter, deformations, and fast motion. The LaSOT dataset consists of 1,400 long-length videos, each typically including more than 2,512 frames. The UAV dataset contains 123 test videos captured from low-altitude UAVs and more than 110K frames. The NFS dataset includes 100 videos recorded at a high frame rate of 240 FPS, with a total of 380K frame. The VOT2020 dataset consists of 60 video sequences with segmentation masks.

As the evaluation metrics, we adopted seven methods: precision plot, normalized precision, success plot, area under the curve (AUC), accuracy, robustness, and EAO. The precision plot illustrates the number of frames where the distance between the predicted state of the target and the ground truth state is less than a predefined threshold. Normalized precision measures the precision normalized by the corresponding target scales. The success plot depicts the number of frames where the overlap between the predicted state of the target and ground truth state is greater than a predefined threshold. The AUC represents the total area under the success plot. Accuracy indicates the overlap ratio between

TABLE 2. Performance according to the different number of intermediate distributions, N in (9), in terms of success AUC.

N in (9)	LaSOT [21]	UAV123 [22]	OTB100 [23]	NFS [24]
1	67.7	70.9	72.2	66.2
5	68.1	71.5	72.3	66.6
10	68.0	71.6	72.1	66.7
15	68.0	71.5	72.2	66.7
20	68.2	71.6	72.3	66.7

the predicted state of the target and the ground truth state. Robustness counts the frames with nonzero overlap between the predicted state of the target and the ground truth state, where visual trackers are re-initialized if they lose track of the target to correct any biased robustness measurements. The EAO score is computed based on accuracy and robustness metrics.

All the experiments were conducted on a desktop computer with a 3.60GHz Intel CPU i7, GeForce Titan X. Our visual tracking method was implemented using PyTorch. The experimental results of the compared methods were obtained from the authors or the original papers.

B. HYPERPARAMETER SETTINGS

The hyperparameters K in (8) corresponds to the number of sample points used to approximate the transport map. As K increases, the transport map can be accurately approximated at the cost of high computational resources. Table 1 shows the visual tracking accuracy in terms of success AUC, as K increases. In all experiments, we fixed K to 20, because this setting was sufficient to achieve the best performance considering both accuracy and computational resources. As shown in the table, using more samples did not gave meaningful improvement in accuracy.

The hyperparameters N in (9) determines the number of the number of intermediate distributions. As N increases, our tracker incorporates more intermediate distributions between the proposal and posterior distributions, emphasizing the proposed seamless mapping to more efficiently mitigate a large gap between these two distributions during the exact mapping process. Table 2 includes the visual tracking results in terms of success AUC, as N increases. The setting $N = 1$ implies that we do not use any intermediate distributions and utilize only the exact mapping. As shown in the table, using five intermediate distributions was sufficient to decrease the gap between the proposal and posterior distributions, resulting in accurate tracking results. However, without the seamless mapping, *i.e.*, $N = 1$, our tracker could not achieve the best performance. In all experiments, we fixed N to 5.

The hyperparameters M in (12) indicates the number of sample points used to describe the posterior distribution. As M increases, our tracker can more accurately approximate the distribution. Table 3 presents visual tracking results in terms of success AUC, as M increases. Based on these results,

TABLE 3. Performance according to the different number of sample points for describing the posterior distribution, M in (12), in terms of success AUC.

M in (12)	LaSOT [21]	UAV123 [22]	OTB100 [23]	NFS [24]
5	68.0	70.3	72.3	66.1
15	68.1	71.5	72.3	66.6
20	68.1	71.6	72.3	66.6
25	68.1	71.6	72.4	66.7
30	68.2	71.6	72.4	66.8

TABLE 4. Comparison on the LaSOT [21] dataset. The best, second-best, and third best results were highlighted in blue, red, and green, respectively.

	Precision $\uparrow$	Normalized Precision $\uparrow$	AUC $\uparrow$
Stark [28]	72.2	76.9	67.1
TransT [30]	69.0	73.8	64.9
Siam R-CNN [36]	68.4	72.2	64.8
Alpha-Refine [37]	68.0	73.2	65.3
STMTrack [38]	63.3	69.3	60.6
DTT [39]	-	-	60.1
TrDiMP [31]	66.3	73.0	63.9
KeepTrack [29]	70.2	77.2	67.1
SuperDiMP [26] [27]	65.3	72.2	63.1
PrDiMP [26]	60.8	68.8	59.8
SAOT [48]	-	70.8	61.6
CIA [20]	71.5	-	67.6
WTP [7]	52.0	60.1	52.0
ToMP [35]	73.5	79.2	68.5
CiteTracker [11]	-	78.6	69.7
SOTA (Ours)	73.4	77.9	68.1

TABLE 5. Comparison on the UAV123 [22], OTB100 [23], and NFS [24] datasets in terms of success AUC. The best, second-best, and third best results were highlighted in blue, red, and green, respectively.

	UAV123 [22]	OTB100 [23]	NFS [24]
Stark [28]	68.2	68.1	66.2
TransT [30]	69.1	69.4	65.7
Siam R-CNN [36]	64.9	70.1	63.9
DiMP [27]	65.3	68.4	62.0
STMTrack [38]	64.7	71.9	-
SiamAttN [46]	65.0	71.2	-
TrDiMP [31]	67.5	71.1	66.2
KeepTrack [29]	69.7	70.9	66.4
SuperDiMP [26] [27]	67.7	70.1	64.8
PrDiMP [26]	68.0	69.6	63.5
CRACK [41]	66.4	72.6	62.5
TrSiam [31]	70.3	67.4	65.8
SiamRPN++ [25]	69.6	61.3	50.2
CIA [20]	71.3	68.9	66.7
ToMP [35]	69.0	70.1	66.9
EWLCFDP [17]	-	68.6	-
SOTA (Ours)	71.5	72.3	66.6

we fixed M to 15 by considering both tracking accuracy and computational cost across all experiments.

C. COMPARISONS

Table 4 shows the visual tracking results of state-of-the-art methods on the LaSOT dataset. For this experiment, we utilized three evaluation metrics: precision, normalized precision, and AUC, where higher values indicate more accurate tracking. ToMP and CiteTracker exhibited the best tracking results across all metrics, while the proposed method demonstrated the second-best performance. CIA,

TABLE 6. Comparison on the VOT2020 [47] datasets in terms of EAO score.

	Accuracy $\uparrow$	Robustness $\uparrow$	EAO $\uparrow$
Stark [28]	0.478	0.799	0.308
DiMP [27]	0.457	0.734	0.274
SuperDiMP [26] [27]	0.477	0.728	0.305
CIA [20]	0.481	0.782	0.309
ToMP [35]	0.453	0.814	0.309
Siamfc++ [40]	0.418	0.502	0.179
ATOM [42]	0.462	0.734	0.271
UPDT [43]	0.465	0.755	0.278
DPMT [44]	0.492	0.745	0.303
TRAT [45]	0.464	0.744	0.280
SOTA (Ours)	0.487	0.817	0.310

TABLE 7. Performance according to different combinations of proposed components in terms of success AUC. The best results were indicated in boldface.

	LaSOT [21]	UAV123 [22]	OTB [23]	NFS [24]
Seamless Mapping	63.5	64.0	68.1	64.7
Exact Mapping	67.7	70.9	72.2	66.2
Seamlessly Exact Mapping	68.1	71.5	72.3	66.6

Stark, KeepTrack, and SuperDiMP also produced accurate tracking results. As the LaSOT dataset typically contains long-term videos, there inevitably exist diverse changes in tracking environments over time and a significant discrepancy between initial and current circumstances. To address this, our method mitigated this discrepancy by employing optimal transport and sequential Monte Carlo.

Table 5 reports visual tracking results in terms of success AUC on the benchmark dataset: UAV123, OTB100, and NFS. For this experiment, we evaluated compared methods in terms of AUC. The proposed method exhibited state-of-the-art performance on the UAV123 dataset. As this dataset was captured from low-altitude UAVs, it contains many target objects undergoing significant changes in appearance over time. Despite this challenge, our method could robustly track the target objects because the proposed exact mapping accurately estimated the posterior distribution from a simple proposal distribution, and our seamless mapping mitigated a large mismatch between the two distributions. ToMP achieved the best tracking results on the NFS dataset, while CRACK showed the best tracking performance on the OTB100 dataset. In these datasets, our method consistently achieved accurate tracking results.

Table 6 contains quantitative tracking results on the VOT2020 dataset. For this experiment, we evaluated several visual tracking methods in terms of three metrics: accuracy, robustness, and EAO. The VOT2020 dataset contains 60 challenging video sequences, where target objects can easily lose their trajectories due to various obstacles in real-world scenarios. In particular, the robustness metric provides an effective way of verifying the effectiveness of tracking methods under these scenarios. As shown in the table, our method exhibited the best performance in terms of the robustness metric. It also produced the highest values of EAO.



FIGURE 6. Qualitative evaluation using the LaSOT dataset. The white boxes denote the tracking results of the proposed SOTA method.

ToMP, SuperDiMP, CIA, and DPMT also achieved state-of-the-art performance using this dataset.

Fig. 6 illustrates the qualitative visual tracking results of the proposed SOTA tracker. Despite challenges such as severe occlusions (e.g., Zebra and Tank sequences), fast motions (e.g., Umbrella and Volleyball sequences), background clutter (e.g., Truck and Train sequences), illumination variations (e.g., Yoyo sequences), and blurring (e.g., Swing sequences), the proposed method accurately tracked the target objects over time. However, when the target object and background had very similar appearances, as shown in Fig. 6(g), our tracker struggled to estimate the exact configuration of the target object.

D. ABLATION STUDY

The proposed method mainly consists of two components: exact mapping and seamless mapping. To analyze the contributions of these components in depth, we modified our visual tracker and created several variants by adopting different combinations of the proposed components, i.e., Variant 1: Seamless Mapping, Variant 2: Exact Mapping, and Variant 3: Seamless Exact Mapping (SOTA).

Table 7 presents the visual tracking results of the aforementioned variants of our method using different combinations of components. Although our exact mapping component (Variant 2) was used, the proposed variant already achieved state-of-the-art performance, which implies that the exact

TABLE 8. Performance according to different temperature scheduling strategies in terms of success AUC. The best results were indicated in boldface.

	LaSOT [21]	UAV123 [22]	OTB100 [23]	NFS [24]
Linear scheduling	67.7	71.5	72.1	66.1
Cosine scheduling	68.1	71.5	72.3	66.6

mapping, involving optimal transport between proposal and posterior distributions, is crucial for accurate tracking. The proposed seamless mapping can further enhance the accuracy of visual tracking, especially in cases where there is a large discrepancy between the aforementioned two distributions. Thus, our method achieves the highest performance by combining the exact mapping with the seamless mapping, as demonstrated by Variant 3.

We scheduled the temperature T_n in (10) using the following two different methods.

- Linear scheduling: $T_n = 1 - \frac{n}{N}$
- Cosine Linear scheduling: $T_n = \frac{1 + \cos(0.5\pi \frac{n}{N})}{2}$ in [49].

Table 8 includes tracking results in terms of success AUC for two temperature scheduling strategies. While the proposed visual tracker is not sensitive to the temperature scheduling strategy, the cosine scheduling approach produced slightly better performance than the linear scheduling approach. For this experiment, we fixed N in (9) to 5.

TABLE 9. Performance across various strategies for combining exact mapping and seamless mapping in terms of success AUC. The best results were indicated in boldface.

	LaSOT [21]	UAV123 [22]	OTB [23]	NFS [24]
2 Seamless mapping + 1 Exact Mapping	67.3	70.2	71.9	66.3
1 Seamless mapping + 2 Exact Mapping	68.0	71.5	72.3	66.3
1 Seamless mapping + 1 Exact Mapping	68.1	71.5	72.3	66.6

In Table 9, we adopted three different strategies for combining exact mapping and seamless mapping. The first strategy applied two seamless mappings followed by one exact mapping, iterating this process. The second strategy involved one seamless mapping followed by two exact mappings, with this process also being iterated. In the final strategy, we alternated between one seamless mapping and one exact mapping, which is equivalent to the proposed SOTA tracker. As shown in Table 9, the first strategy induced more errors in approximating high-dimensional posterior distributions, resulting in poorer tracking performance, whereas the second strategy exhibited performance similar to the proposed SOTA tracker. However, it required high computational cost due to the need for exact mapping, which involves derivative computations. Thus, we can conclude that alternating between one seamless mapping and one exact mapping, as used in the SOTA tracker, is the most efficient and accurate approach for visual tracking.

V. CONCLUSION AND DISCUSSION

In this paper, we introduce a probabilistic visual tracking method that leverages the benefits of both MC sampling and optimal transport. The MC steps can reduce the topological and representational burden on the optimal transport via successive temperature steps. Within each successive temperature step, one distribution is projected onto another one via optimal transport. By using optimal transport, the method can mitigate the variance from importance sampling. The experimental results demonstrate that the proposed method significantly outperforms other state-of-the-art methods across several benchmark dataset, particularly in tracking environments where abrupt changes occur.

In future work, the proposed tracker could be further improved by incorporating visual representation learning using SpectralGPT [50] in a self-supervised manner. We anticipate that this representation learning could reduce the topological and representational burden on optimal transport while also mitigating the variance introduced by importance sampling.

REFERENCES

- [1] P. Li and J. Jin, "Time3D: End-to-end joint monocular 3D object detection and tracking for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 3875–3884.
- [2] T. Fechter and D. Baltas, "One-shot learning for deformable medical image registration and periodic motion tracking," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2506–2517, Jul. 2020.
- [3] H. He, C.-L. Lu, Y. Wen, G. Saunders, P. Yang, J. Schoonover, J. Wason, A. Julius, and J. T. Wen, "High-speed high-accuracy spatial curve tracking using motion primitives in industrial robots," in *Proc. ICRA*, May 2023, pp. 12289–12295.
- [4] Y. Zeng, X. Fu, L. Gao, J. Zhu, H. Li, and Y. Li, "Robust multivehicle tracking with Wasserstein association metric in surveillance videos," *IEEE Access*, vol. 8, pp. 47863–47876, 2020.
- [5] P. Del Moral, A. Doucet, and A. Jasra, "Sequential Monte Carlo samplers," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 68, no. 3, pp. 411–436, Jun. 2006.
- [6] Q. C. Le, D. Conte, and M. Hidane, "Unbalanced optimal transport in multi-camera tracking applications," in *Proc. ICPR*, 2021, pp. 327–343.
- [7] J. Hong and J. Kwon, "Optimal visual tracking using Wasserstein transport proposals," *Expert Syst. Appl.*, vol. 209, Dec. 2022, Art. no. 118251.
- [8] A. Matthews, M. Arbel, D. J. Rezende, and A. Doucet, "Continual repeated annealed flow transport Monte Carlo," in *Proc. ICML*, 2022, pp. 15196–15219.
- [9] Y. Huang, X. Li, Z. Zhou, Y. Wang, Z. He, and M.-H. Yang, "RTracker: Recoverable tracking via PN tree structured memory," in *Proc. CVPR*, Jun. 2024, pp. 19038–19047.
- [10] G. Zheng, S. Lin, H. Zuo, C. Fu, and J. Pan, "NetTrack: Tracking highly dynamic objects with a net," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 19145–19155.
- [11] X. Li, Y. Huang, Z. He, Y. Wang, H. Lu, and M.-H. Yang, "CiteTracker: Correlating image and text for visual tracking," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 9940–9949.
- [12] Y. Cai, J. Liu, J. Tang, and G. Wu, "Robust object modeling for visual tracking," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 9555–9566.
- [13] N. Tripathaneni, M. Rowland, Z. Ghahramani, and R. Turner, "Magnetic Hamiltonian Monte Carlo," in *Proc. NIPS*, 2017, pp. 3453–3461.
- [14] O. Mangoubi and N. K. Vishnoi, "Nonconvex sampling with the metropolis-adjusted Langevin algorithm," 2019, *arXiv:1902.08452*.
- [15] M. Parno and Y. Marzouk, "Transport map accelerated Markov chain Monte Carlo," 2017, *arXiv:1412.5492*.
- [16] A. L. Caterini, A. Doucet, and D. Sejdinovic, "Hamiltonian variational auto-encoder," 2018, *arXiv:1805.11328*.
- [17] H. Zhang, G. Nie, J. Chen, J. Zhang, and G. Yang, "Uncertain motion tracking combined Markov chain Monte Carlo and correlation filters," *IEEE Access*, vol. 7, pp. 167076–167088, 2019.
- [18] A. Younis and E. Sudderth, "Differentiable and stable long-range tracking of multiple posterior modes," in *Proc. NeurIPS*, 2023, pp. 42799–42829.
- [19] M. D. Parno and Y. M. Marzouk, "Transport map accelerated Markov chain Monte Carlo," *SIAM/ASA J. Uncertainty Quantification*, vol. 6, no. 2, pp. 645–682, 2018.
- [20] Z. Pi, W. Wan, C. Sun, C. Gao, N. Sang, and C. Li, "Hierarchical feature embedding for visual tracking," in *Proc. ECCV*, 2022, pp. 428–445.
- [21] H. Fan, L. Lin, F. Yang, P. Chu, G. Deng, S. Yu, H. Bai, Y. Xu, C. Liao, and H. Ling, "LaSOT: A high-quality benchmark for large-scale single object tracking," 2018, *arXiv:1809.07845*.
- [22] M. Mueller, N. Smith, and B. Ghanem, "A benchmark and simulator for UAV tracking," in *Proc. ECCV*, 2016, pp. 445–461.
- [23] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 2411–2418.
- [24] H. K. Galoogahi, A. Fagg, C. Huang, D. Ramanan, and S. Lucey, "Need for speed: A benchmark for higher frame rate object tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1134–1143.
- [25] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "SiamRPN++: Evolution of Siamese visual tracking with very deep networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4277–4286.
- [26] M. Danelljan, L. Van Gool, and R. Timofte, "Probabilistic regression for visual tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7181–7190.
- [27] G. Bhat, M. Danelljan, L. Van Gool, and R. Timofte, "Learning discriminative model prediction for tracking," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6181–6190.

- [28] B. Yan, H. Peng, J. Fu, D. Wang, and H. Lu, "Learning spatio-temporal transformer for visual tracking," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10428–10437.
- [29] C. Mayer, M. Danelljan, D. P. Paudel, and L. Van Gool, "Learning target candidate association to keep track of what not to track," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 13424–13434.
- [30] X. Chen, B. Yan, J. Zhu, D. Wang, X. Yang, and H. Lu, "Transformer tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 8122–8131.
- [31] N. Wang, W. Zhou, J. Wang, and H. Li, "Transformer meets tracker: Exploiting temporal context for robust visual tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 1571–1580.
- [32] A. Lukežič, J. Matas, and M. Kristan, "D3S—A discriminative single shot segmentation tracker," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7131–7140.
- [33] T. Xu, Z.-H. Feng, X.-J. Wu, and J. Kittler, "Joint group feature selection and discriminative filter learning for robust visual object tracking," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7949–7959.
- [34] F. Xie, C. Wang, G. Wang, Y. Cao, W. Yang, and W. Zeng, "Correlation-aware deep tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 8741–8750.
- [35] C. Mayer, M. Danelljan, G. Bhat, M. Paul, D. P. Paudel, F. Yu, and L. Van Gool, "Transforming model prediction for tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 8721–8730.
- [36] P. Voigtlaender, J. Luiten, P. H. S. Torr, and B. Leibe, "Siam R-CNN: Visual tracking by re-detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6577–6587.
- [37] B. Yan, X. Zhang, D. Wang, H. Lu, and X. Yang, "Alpha-refine: Boosting tracking performance by precise bounding box estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 5285–5294.
- [38] Z. Fu, Q. Liu, Z. Fu, and Y. Wang, "STMTrack: Template-free visual tracking with space-time memory networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13769–13778.
- [39] B. Yu, M. Tang, L. Zheng, G. Zhu, J. Wang, H. Feng, X. Feng, and H. Lu, "High-performance discriminative tracking with transformers," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9836–9845.
- [40] Y. Xu, Z. Wang, Z. Li, Y. Yuan, and G. Yu, "SiamFC++: Towards robust and accurate visual tracking with target estimation guidelines," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 12549–12556.
- [41] H. Fan and H. Ling, "CRAFT: Cascaded regression-align-classification for robust tracking," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2021, pp. 7013–7020.
- [42] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg, "ATOM: Accurate tracking by overlap maximization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4655–4664.
- [43] G. Bhat, J. Johander, M. Danelljan, F. S. Khan, and M. Felsberg, "Unveiling the power of deep tracking," in *Proc. ECCV*, 2018, pp. 483–498.
- [44] F. Xie, N. Wang, Y. Yao, W. Yang, K. Zhang, and B. Liu, "Hierarchical representations with discriminative meta-filters in dual path network for tracking," in *Proc. PRCV*, Y. Peng, Q. Liu, H. Lu, Z. Sun, C. Liu, X. Chen, H. Zha, and J. Yang, Eds., 2020, pp. 303–315.
- [45] H. Saribas, H. Cevikalp, Ö. Köpüklü, and B. Uzun, "TRAT: Tracking by attention using spatio-temporal features," 2020, *arXiv:2011.09524*.
- [46] Y. Yu, Y. Xiong, W. Huang, and M. R. Scott, "Deformable Siamese attention networks for visual object tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6727–6736.
- [47] M. Kristan, J. Matas, A. Leonardis, T. Vojří, R. Pflugfelder, G. Fernández, G. Nebehay, F. Porikli, and L. Cehovin, "A novel performance evaluation methodology for single-target trackers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 11, pp. 2137–2155, Nov. 2016.
- [48] Z. Zhou, W. Pei, X. Li, H. Wang, F. Zheng, and Z. He, "Saliency-associated object tracking," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9866–9875.
- [49] A. Kukleva, M. Böhle, B. Schiele, H. Kuehne, and C. Rupprecht, "Temperature schedules for self-supervised contrastive methods on long-tail data," in *Proc. ICLR*, 2023, pp. 1–15.
- [50] D. Hong, B. Zhang, X. Li, Y. Li, C. Li, J. Yao, N. Yokoya, H. Li, P. Ghamisi, X. Jia, A. Plaza, P. Gamba, J. A. Benediktsson, and J. Chanussot, "SpectralGPT: Spectral remote sensing foundation model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 8, pp. 5227–5244, Aug. 2024.



SEONGHAK LEE received the B.S. degree in computer science and engineering from Chung-Ang University, Seoul, South Korea, in 2023, where he is currently pursuing the M.S. degree in computer science and engineering. His research interests include NeRF, talking head synthesis, and computer vision.



JISOO PARK received the B.S. degree in computer science and engineering from Chung-Ang University, Seoul, South Korea, in 2023, where she is currently pursuing the M.S. degree in artificial intelligence. Her research interests include speech separation, talking head synthesis, and computer vision.



RADU TIMOFTE (Member, IEEE) received the Dipl.Eng. degree from the Technical University of Iasi, Romania, in 2006, the M.Sc. degree from the University of Eastern Finland, in 2007, and the Ph.D. degree in electrical engineering from KU Leuven, Belgium, in 2013. From 2013 to 2016, he was a Postdoctoral Researcher with the Computer Vision Laboratory, ETH Zürich, Switzerland, where he was a Lecturer and a Research Group Leader, from 2016 to 2022. He is currently a Full Professor of computer science with the University of Würzburg, Germany. He is also the Co-Founder of Merantix and the Co-Organizer of NTIRE, CLIC, AIM, MAI, and AIS events. He is a member of CVF and ELLIS. He is a member of the editorial board for top journals, such as IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and NEUCOM. He was the 2022 Winner of the Humboldt Professorship for Artificial Intelligence Award. He regularly serves as the Area Chair for top conferences, such as CVPR, ECCV, ICCV, NeurIPS, and IJCAI. His team's work received multiple awards and his team won several challenges.



JUNSEOK KWON (Member, IEEE) received the B.Sc. and M.Sc. degrees in object tracking (supervised by Prof. Kyoung Mu Lee) and the Ph.D. degree in electrical engineering and computer science from Seoul National University, South Korea, in 2006, 2008, and 2013, respectively. He was a Postdoctoral Researcher with the Computer Vision Laboratory, ETH Zürich, from 2013 to 2014. He is currently an Associate Professor with the School of Computer Science and Engineering, Chung-Ang University, Seoul, South Korea. He is working in the field of object tracking to capture the dynamics of cities. His research interests include visual tracking, visual surveillance, and Monte Carlo sampling method and its variants.

...