CSE586/EE554 Computer Vision II : Project 1 Report

Deep learning-based Video Tracking : Siamese FC with Particle Filter

Saniya Naphade, Vishakadatta J H, Yilin Liu, Shijia Zhang

Contents

1 Introduction

SiamFC has a straightforward system structure and can be pretrained disconnected on a huge dataset. It has no web based learning process by any stretch of the imagination. Subsequently, there are a whole lot of nothing answers for some intricate following situations, for example, occlusion and large target deformation. We propose a technique utilizing the Particle Filter strategy and combination with SiamFC[2]. The Particle Filter obtains the objective's direction data, which is utilized to process complex following scenes and to change the determination strategy for the search area. This additionally empowers our tracker to steadily follow quick moving targets.The presentation of the Particle Filter makes up for the inadequacies that SiamFC can just track disconnected. The combination of multiresolution highlights to get different reaction scores map causes the tracker to acquire vigorous highlights that can be adjusted to an assortment of following targets.

1.1 Motivation

Motivation behind any project in general would be either to improve it or build a new network altogether here we improve the SiamFC with particle filter approach.The main shortcomings of SiamFC is that the

- The occluded scenario is not taken care of very well without online learning
- The basic principle behind it finding the simple correlation operation which obtains the required target while tracking so, this means that a lot of dependency on cosine window. We know that this will cause a problem while the target in a fast-moving object.
- The correlation filter can introduce something called cyclic displacement if the online learning is used and this will cause boundary effects this will lead to accuracy problems in the tracking process.

these are the same data given by us in the project proposal [3] and this tells us that the overcoming the motion detection defect is to add the particle filter on to the SiamFC network.

1.2 Siamese Fully Connected network

Arbitrary item tracking has traditionally been tackled by studying a model of the object's look exclusively online, the usage of as sole training information the video itself. Despite the success of these methods, their online method inherently limits the richness of the version they could learn. Recently, numerous attempts have been made to make the most the expressive energy of deep convolution networks. However, while the object to track isn't recognized beforehand, it is necessary to perform Stochastic Gradient Descent online to conform the weights of the community, severely compromising the rate of the system. In this paper we equip a primary monitoring set of rules with a unique fully-convolution.But as mentioned earlier the downfall still persists in the SiamFC.

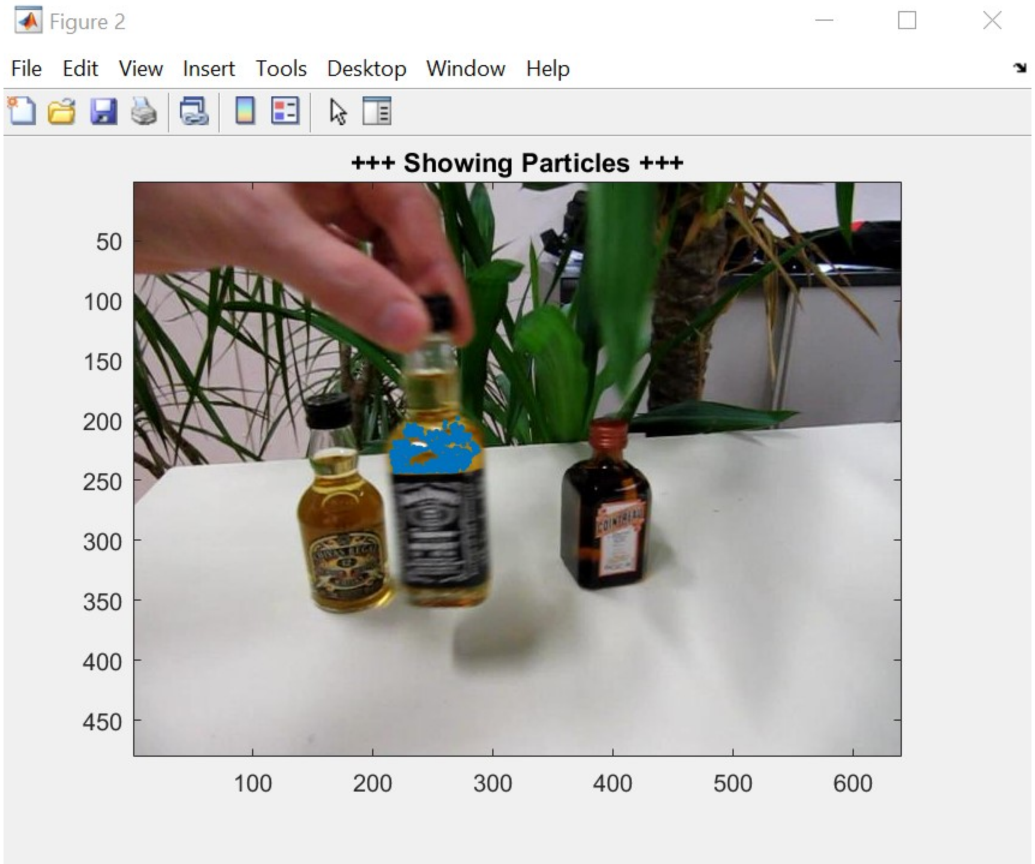


Figure 1: Particle Filter tracking the target

1.3 Related work

The Combined kalman and siamese network [6] speaks about the same connection of the siamfc and kalman but here the main difference is that the kalman does the motion modelling as well as the position determination and one input of the SiamFC tracker is the output of the filter and the other is the normal input target image.The input from the kalman is cropped before the SiamFC tracker uses it for tracking.The same way the work in [4]shows that the particle filter used is correlation filter similar to siamese tracker but again the cropping would method used would be the same as the SiamFC method.
We combined the 2 methodologies by eliminating the cropping part and using the addition of filter and SiamFC in a way where both the tracker and the filter can run the target image.

2 Particle Filter

Particle filtering[1] is a popular Monte Carlo (sampling) approach for appearing inference in state-space model where the state of a machine evolves in time and information approximately the state is obtained through noisy measurements made at on every occasion of the step. In a general discrete-time state-space model, the country of a system evolves in accordance to:

xt = ft(xt1, vt1)........ (1)

wherein xt is a vector representing the state of the device at time t, vt1 is the nation noise vector, ft is a probably non-linear and time-dependent feature describing the evolution of the state vector.

2.1 Particle filter implementation

We have implemented particle filter in such a way that it takes the RGB values of the pixels need to be tracked[5]. The particle filter we use has evolved from the idea that the RGB pixels of the image will not change over time and the main intention is to track it,of course there are a lot of downfalls to it as what if the color of the target object changes overtime or the object is a multi-colored target. The algorithm flow of our filter follows that the particles are created as per need we have done an approximated 1000 particle generation.The intention behind is to keep track of the pixels with the required color. Before the SiamFC was joined to the particle filter network we have done a demo on just implementation of the particle filter.The approach we took was that we enter manually the RGB color values of the pixel to be tracked and run the filter ,but this is not that user friendly . The new approach we came in was with that of using the already pre-trained network and the ground-truth of that image taken into consideration and then the particles would know the target object from this groundtruth so that it would start the tracking process.
The methodology followed after this is the update of particles this takes in argument such as F_update ,position,size of the particle and the image itself .
The F_update is motion modelling array which will have the prediction values or assist in getting the motion model of the particles.

F_update = [1 0 0 0; 0 1 0 0; 0 0 1 0; 0 0 0 1]........ (2)

here we use the velocity model to update the particles accordingly.
Once the particles are moving in accordance with the target using the motion model we use Log Likelihood to calculate the exact place where the target would be when in motion. Moving on the resample function is called so that we return to tracking of the target.
This is in general the algorithm on how the particle designed by us works.

The image shown here is the exact moment at which the occlusion has occurred and the target image has moved on ,the particles have not scattered and the target image is modelled to predict it's motion

3 Approach

The motivation behind using the network is to improve the SiamFC .The tracker we have designed is SiamFC + Particle Filter this is to overcome the limitation as mentioned earlier.When the SiamFC runs with the tracking process it gives a very satisfactory out-

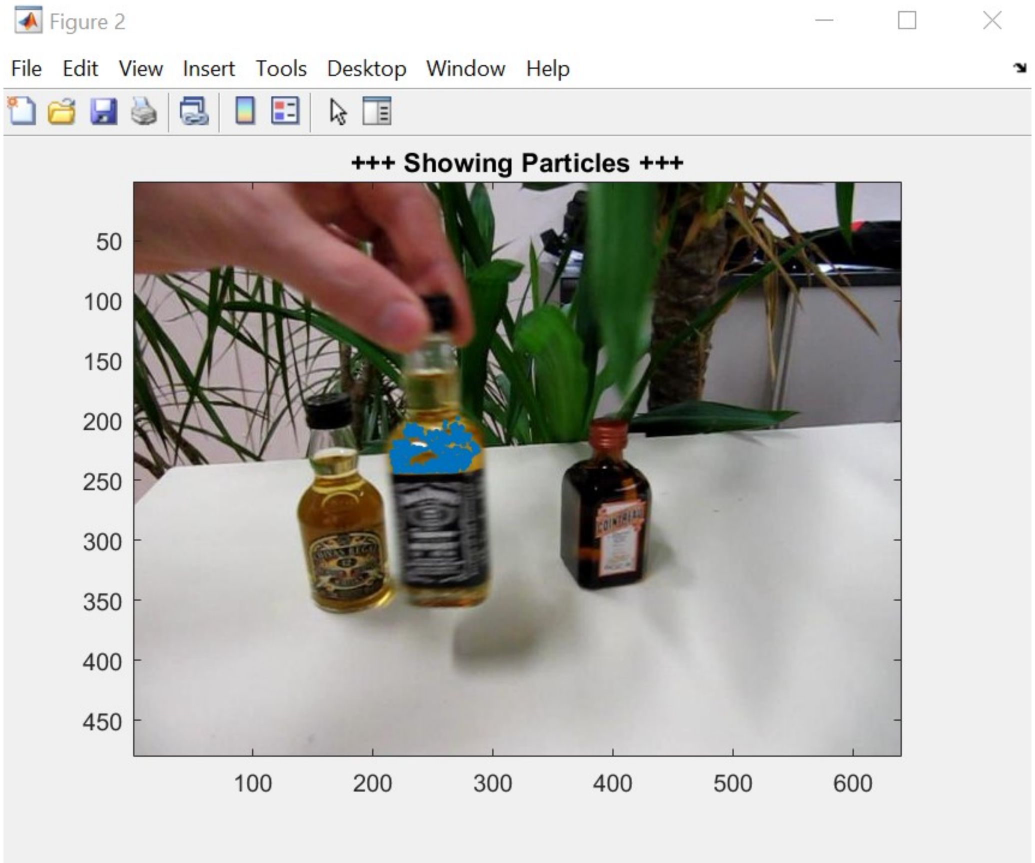


Figure 2: Particle Filter tracking the target

put.But at times when there is a occlusion it fails[6]. The reason being that the SiamFC detects the search region at the beginning and gradually moves on and when it needs to track it keeps the to be tracked image and convolves with the previous frame with a search region bigger then the previous.But,when this happens if the object is occluded then in this frame it will be nowhere to be found.I t also happens if the target image is fast moving by the time the SiamFC tracker detects it ,it would have moved on to a further away frame in the network[4].

3.1 Addition of the Particle filter

The main drawback as mentioned earlier is that the SiamFC detects the target very well but it so happens that the above mentioned scenarios hinder the performance of the SiamFC .Now to overcome this we will introduce the Particle filter in the network of SiamFC as shown in the figure below. In the figure you see that the target image is given to the SiamFC network and after that it is fed to the particle filter ,the reason is that the filter uses it's motion modelling to better predict the moving target.Section 2.1 shows explains the particle filter modelling case wherein the velocity is used for motion modelling.The algorithm in our code employs that the position detection of the target is done by the SiamFC and it gives us the variables target size and it's position .Now, these variables

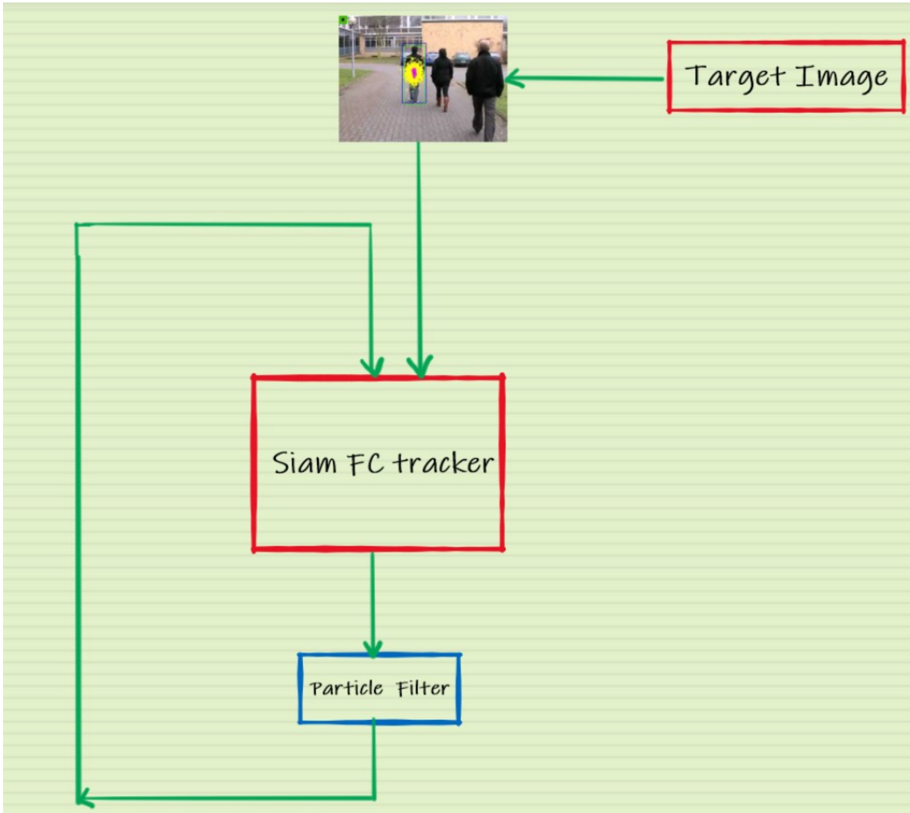


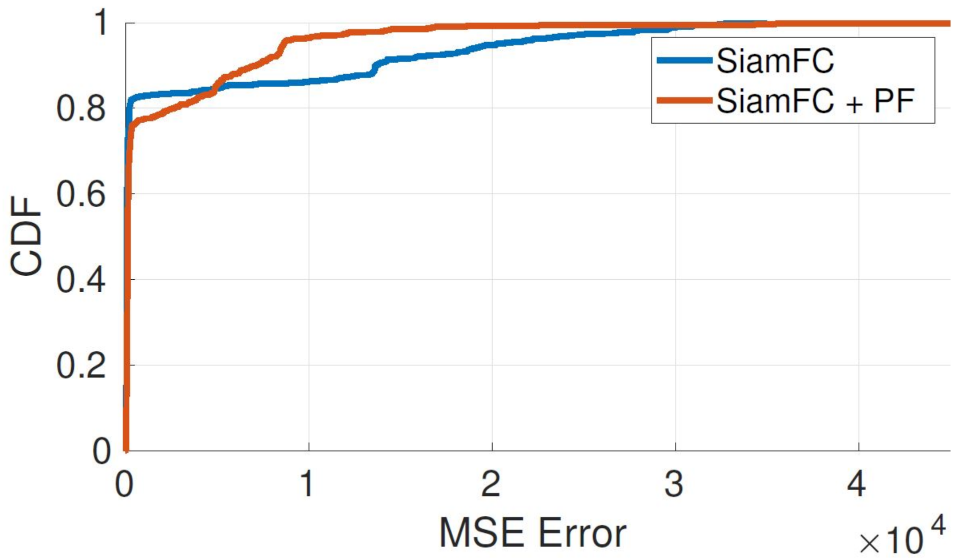
Figure 3: Block Diagram of SiamFC + Particle Filter

are used furthermore by the particle filter .As mentioned in the Section 2.1 we used the ground truth of the image to detect the target for particle filter but here we use the output of the SiamFC and then it is processed to get the proper motion prediction.This is mere equation of variables , the input of the filter network is the size,position of the target and the filter will model using the same algorithm as mentioned in Section 2.1.Now, the output obtained is better than the alone SiamFC tracker with better motion prediction.

4 Performance Evaluation

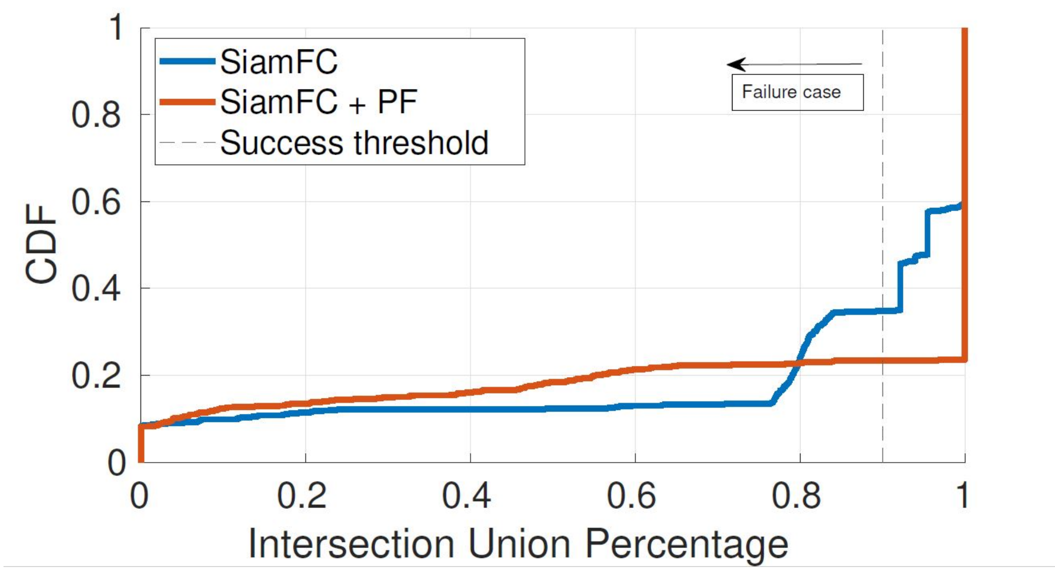
4.1 MSE Evaluation

For each bounding box, we can compute the geometry center location, so the error function can be the mean squared error (MSE) between prediction's box center and the ground truth's one. By calculating the MSE for all the frames, we could draw a cumulative distribution function (cdf) graph for all the cases in the video.
The graph shows the original SiamFC has a long tail in MSE error. That means this method performs not well in some cases in the video. While the SiamFC combined with particle filter could obviously shorten the tail. The SiamFC + particle filter have much better performance in these cases.



4.2 Overlapping Evaluation

For every frame, we could calculate the area of the overlapping part of prediction's box and ground truth's box. And then we use the overlapping percentage which is overlapping area divided by prediction's box area to analyze the performance of these two methods. And we draw cdf graph for all the frames in the video.
Using this to evaluate SiamFC + particle filter, we have about 80% successful cases that the prediction's box fully is evolved in the ground truth's one. However, only 40% cases' prediction's box approximately in SiamFC successfully and fully covers the ground truth's one.



4.3 Success Rate

Given a certain threshold, if the intersection-over-union (IoU) between prediction and ground truth is above it, means the tracker is successful in the current frame. Success Rate computes the percentage of frames which the tracker succeed.

Drawing a vertical line representing success threshold at 90% in the Intersection Union Percentage graph, the intersection union percentage under 90% are failure cases and the intersection union percentage above 90% are success cases. For defined success cases, our SiamFC + particle filter method's accuracy is about 78%. But the SiamFC's accuracy is only 65% approximately for defined success rate. The success rate increases clearly by combining the particle filter into SiamFC.

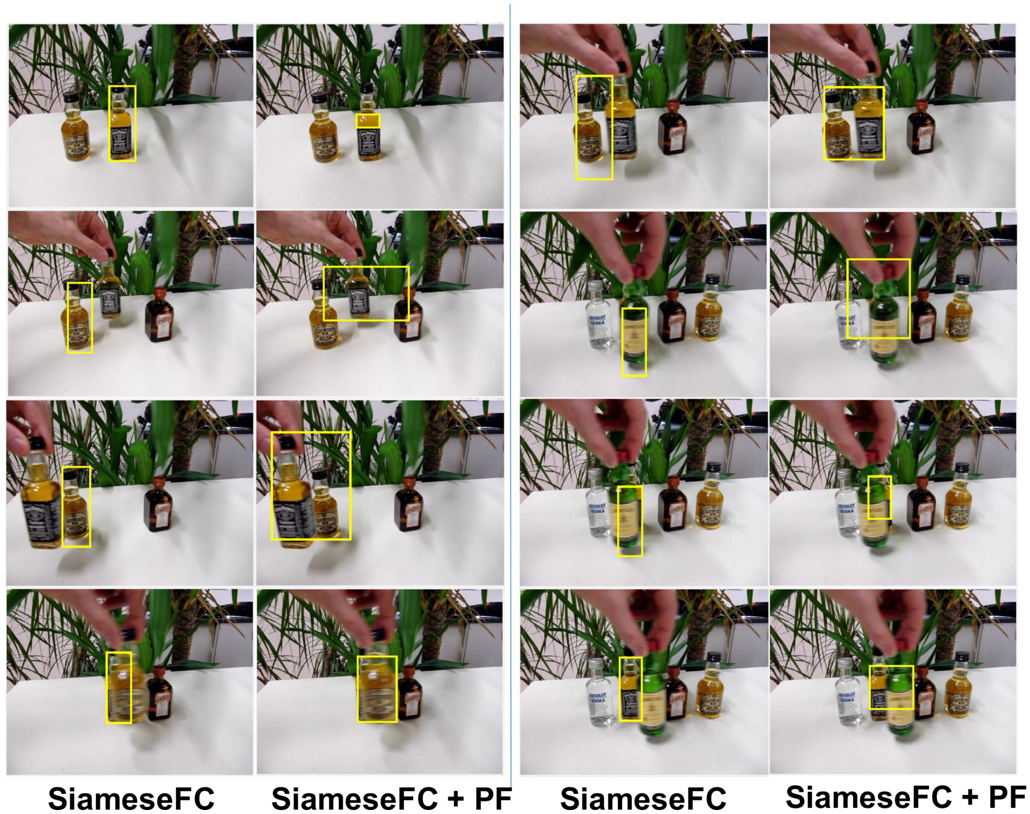
5 Results

We show eight comparison cases in the video as results. The left column is tracking by SiamFC and the right column is the corresponding case using our SiamFC + particle filter method.

The difficult and confusing cases in the video is the original item is passing another quite similar item. The advantage of our method is that SiamFC + particle filter can stick to the original item when passing by another similar one. The shape of prediction's box in SiamFC + particle filter changes but the prediction's box always cover the original item. However, the SiamFC is cheated by the confusing condition. The prediction's box jumps to the similar item directly and the following tracking totally fails because the prediction's box is always on the wrong item. The first seven comparison graphs clearly describe this situation.

However, our SiamFC + particle filter is a little bit tardy when the original item is leaving the similar item. As the last comparison graph shows, the prediction's box keeps large which covers both the original item and the similar item. As time and distance change, the prediction's box will shrink to the original item finally. But for the SiamFC, the prediction's box jumps to the original item from the similar item quickly.

The jumping process happens twice in SiamFC method, which show this method is not robust and stable when encountering similar items situation. Our SiamFC + particle filter can always track the original item because the prediction's box always covers the original item. Although the prediction's box is a little bit large and slow, we never lose the original item.



6 Conclusion

This project shows the stability and robustness of tracking items in videos using SiamFC + particle filter. We Compare our method with SiamFC in the video. A number of applications in video tracking, while the tracking items are confusing and similar with each other, we exploit particle filter which can stick to original item and make constraints prediction's box. The particle filter predicts the objective's direction data. Our method use the particle filter strategy and combination with SiamFC and change the determination strategy for the search area. Our method tracks items more continuously and robustly.

References

[1] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002.

[2] Luca Bertinetto, Jack Valmadre, João F Henriques, Andrea Vedaldi, and Philip HS Torr. Fully-convolutional siamese networks for object tracking. *arXiv preprint arXiv:1606.09549*, 2016.

[3] Vishakadatta J H, Yilin Liu, Shijia Zhang, and Saniya Naphade. Project one proposal. Feb 2020.

[4] Reza Jalil Mozhdehi, Yevgeniy Reznichenko, Abubakar Siddique, and Henry Medeiros.

Convolutional adaptive particle filter with multiple models for visual tracking. In *International Symposium on Visual Computing*, pages 474–486. Springer, 2018.

[5] Mai Thanh Nhat Truong and Sanghoon Kim. Object tracking method based on color particle filter. 2016.

[6] Lijun Zhou and Jianlin Zhang. Combined kalman filter and multifeature fusion siamese network for real-time visual tracking. *Sensors*, 19:2201, 05 2019.