Human-Following of Mobile Robots Based on Object Tracking and Depth Vision

Dongyu Han
College of Electrical Engineering
Zhejiang University
Hangzhou, China
e-mail: dy1han@foxmail.com

Yonggang Peng*
College of Electrical Engineering
Zhejiang University
Hangzhou, China
e-mail: pengyg@zju.edu.cn

2020 3rd International Conference on Mechatronics, Robotics and Automation (ICMRA) | 978-1-7281-9032-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICMRA51221.2020.9398366

***Abstract*—Human-following is a promising ability for robots to serve and assist humankind. Correlation Filter (CF) with excellent performance and high running speed, is widely used in single object tracking, suitable for human following. In this paper, an improved correlation filter object tracking algorithm is adopted to achieve human following. For the lack of accuracy evaluation mechanism in most object trackers, fusion of Peak Side-lobe Rate (PSR) and depth information are introduced to evaluate the credibility of the tracking result. The sample and filter model are updated with a high credibility to avoid model drift. Face matching is used to re-track human when tracking fails. Based on a depth camera Kinect and a mobile robot, the experiments of following human show that proposed method has excellent abilities to deal with different scenarios and to follow target human stably for a long time.**

Keywords-human-following; mobile robot; correlation filter; tracking evaluation

I. INTRODUCTION

With the development of computer vision and robotics, object tracking algorithms have been applied in various fields, such as video surveillance and object tracking in unmanned aerial vehicle. Human following based on object tracking is a promising research filed. In the future, by following human, service robots will be able to provide lots of services and assistance. For instance, service robots follow the passengers to help carry the luggage in the airport and follow the waiters to serve customers in the restaurant.

The task to follow human mainly bases on human tracking and motion control of robot. The core part is object tracking. Object tracking methods using correlation filter (CF) have good robustness and obvious advantage on speed. The CF method aims to train a filter model to do correlation operations with object box and obtain the response map. The object location in the frame is predicted according to the position of maximum response. In 2010, David S. Bolme firstly applied correlation filter to object tracking in MOSSE [1]. Subsequently, CF object tracking algorithms developed rapidly. The CSK algorithm proposed in 2012 adopts cyclic shift to generate samples and trains the filter model by the extended ridge regression and kernel trick [2]. In 2014, the KCF algorithm extends multi-channel features based on CSK and applies Histogram of Oriented Gradient (HOG) to object tracking, which greatly improves tracking accuracy [3]. On the basis of KCF, the DSST [4] and fDSST solve the multi-scale problem of object tracking [5].

In 2016, Martin Danelljan applied deep feature to object tracking, and proposed learning continuous convolution operators for visual tracking (C-COT) [6]. In C-COT, the interpolation model is utilized to transform feature maps with different resolution into continuous spatial domain, which can integrate features with multi-resolution and improve the tracking effect. C-COT tracker has high tracking accuracy and low running speed because of the high computational complexity. On the basis of C-COT tracker, Martin Danelljan proposed efficient convolution operators for tracking (ECO) [7]. ECO tracker introduces a factorized convolution operator, a compact generative sample space model and a sparse model update strategy to solve the problem of high calculation complexity and overfitting in C-COT. The tracking accuracy and speed of ECO object tracker have been greatly improved.

In recent years, object tracking algorithms based on deep learning have developed rapidly. The deep learning object tracking algorithms extract deep features and train models in deep learning network. The deep feature can represent the target and effectively resist the target change better. Deep learning object tracking algorithms often have high tracking accuracy and relatively poor performance on running speed. Most of them cannot operate in real-time without GPU. Typical deep learning methods include series of Siam [8], GradNet [9], ATOM [10], etc.

Object tracking algorithms often train and test using public data sets. Therefore, most of them are appropriate for short-term video object tracking and lack the ability to detect failure. Compared with short-term video object tracking, human following is a long-term tracking, and it is essential to judge whether the tracking result is accurate. Furthermore, a re-recognition mechanism is required to re-track the target human when tracking fails.

In order to achieve the human following, the article deploys the ECO-HC algorithm on a mobile robot to follow human. 'HC' represents that the tracker uses improved Histogram of Oriented Gradient (fHOG) [11] and Color Name (CN) as features. Fusion of Peak Side-lobe Rate (PSR) and depth information is introduced to evaluate the credibility of the tracking result. The strategy of updating samples and models with a high credibility is adopted to avoid model drift. A face recognition algorithm is used to realize re-identification when following fails. Based on the improved ECO-HC tracker and depth camera Kinect, the mobile robot can accurately follow the target human.

II. ALGORITHM FRAMEWORK

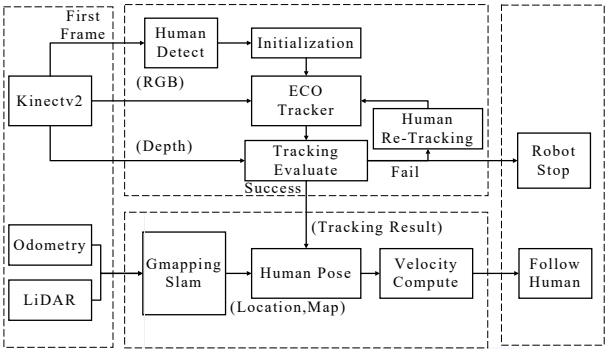


Figure 1. Framework of human-following algorithm.

The framework of proposed algorithm is shown in Figure 1. Firstly, the algorithm uses the depth camera Kinect v2 to obtain RGB and depth images. After face detection with RGB image, the depth image is used to segment the upper body. The face information is saved for re-identification. Then subsequent RGB images are input into the tracker for object tracking. *PSR* and depth information are combined to evaluate the tracking results and the filter model is updated with a high credibility according to the evaluation. Based on the human position on the RGB image and depth information obtained by the tracker, coordinate transformation is performed to obtain human pose relative to the robot. And then the mobile robot moves to follow the human. When tracking fails, the robot stops following human and activates the re-tracking mechanism. With the lidar and the odometry, simultaneous localization and mapping (SLAM) are realized. The algorithm will be described in detail in Chapter III.

III. HUMAN-FOLLOWING ALGORITHM

A. ECO-HC Tracker

Based on the principle of correlation filter, the workflow of ECO-HC object tracker is shown in Figure 2. 1) Firstly, the object image with region of interest (ROI) is inputted to initialize the tracker and train the filter model. 2) In the subsequent frames, according to the tracking results of the previous frame, predict the potential object box. 3) Hog and CN feature of the object box are extracted and processed through cosine window and Fourier transform. 4) Transform the multi-resolution feature into a continuous spatial domain with an interpolation model and get the feature matrix. 5) Perform the correlation operation between feature matrix and filter model to obtain the response map and predict the object position. 6) Update sample space and filter model every few frames to improve the tracking speed and avoid model drift. 7) Repeat the above steps for subsequent object tracking and train filter model.

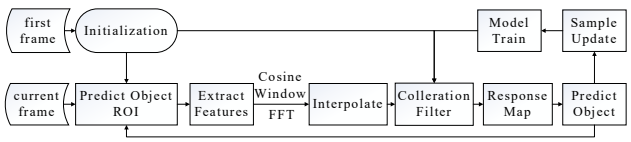
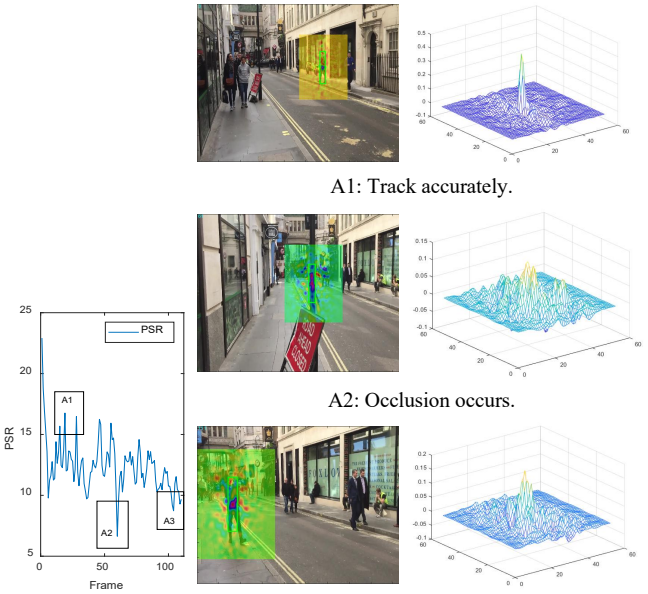


Figure 2. Flow chart of ECO-HC tracker.

B. Credibility Evaluation Mechanism



a. Value of *PSR* b. Frames and response maps

Figure 3. Changes of response map in different cases.

The peak to side-lobe ratio (*PSR*) measures the sharpness of the maximum peak in the response map, which represents the correlation between object box and the filter model. *PSR* is defined as

$$PSR = \frac{maxF_{x,y} - \mu}{\sigma} = \frac{F_{max} - F_{mean}}{\sqrt{\frac{1}{N}\left(\sum_{x,y}^{N}\left(F_{x,y} - F_{mean}\right)^2\right)}} \quad (1)$$

where:

$F_{x,y}$ —value of the point (x, y) in response map;

μ, σ — the mean and variance of the response map;

F_{max} — maximum value of the response map.

Figure 3 shows the changes of response maps in different cases. Part a represents the value of *PSR*. Part b shows the frames and response maps. When the tracking result is accurate, the response map shows a stable state with a single peak in part A1. When occlusion occurs, the response map fluctuates violently and has multi-peaks in part A2. The value of *PSR* decreases rapidly. However, the changes of object appearance will also cause the *PSR* to decrease, shown in part A3.

Comparing the *PSR* with a fixed threshold cannot accurately evaluate the tracking result. The article evaluates the credibility of tracking results by comparing pseudo average value PSR_{Aver} with *PSR*.

$$PSR_{Aver} = \begin{cases} (PSR_{Aver} \bullet n + PSR) / (n+1). & n \le 10 \\ PSR_{Aver} \bullet w_1 + PSR \bullet w_2. & PSR \ge PSR_{Aver} \bullet w_{min} \\ PSR_{Aver}. & PSR < PSR_{Aver} * w_{min} \end{cases} \quad (2)$$

As shown in (2), for the first n frames, the object tracking results are considered accurate, and the average of PSR is calculated. In subsequent frames, according to the value of the PSR, PSR_{Aver} is adaptively updated. When $PSR < PSR_{\text{Aver}} \cdot w_{min}$, the tracking result is inaccurate and PSR_{Aver} remains unchanged. When $PSR \geq PSR_{\text{Aver}} \cdot w_{min}$, PSR_{Aver} is updated with the fusion of PSR and PSR_{Aver}. w_1 and w_2 are the weights of PSR and PSR_{Aver} respectively. As shown in (3), w_1 and w_2 are updated adaptively with the linear model, ensuring that $w_1 \in [0.9,0.99]$ and $w_2 \in [0.01, 0.1]$. w_1 becomes larger as PSR decreases, which ensures that PSR_{Aver} varies slightly when credibility of tracking result is low. Adaptive update strategy of PSR_{Aver} allows the evaluation mechanism to adapt to changes of object and environment in the tracking process.

$$w_1 = 0.9 + \frac{0.09}{1 - w_{min}} \cdot \frac{|PSR_{Aver} - PSR|}{PSR_{Aver}}, w_2 = 1 - w_1. \quad (3)$$

Since the deformation of the object will lead to the change of PSR, it is unfaithful to determine that the tracking result is inaccurate when PSR decreases. The depth information is introduced to evaluate the tracking result with PSR. When the human is blocked or disappears, the depth value of target human will change sharply compared with the previous frame. In that case, it is considered that tracking object is not accurate.

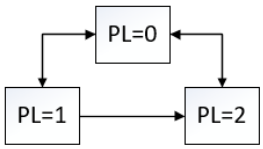
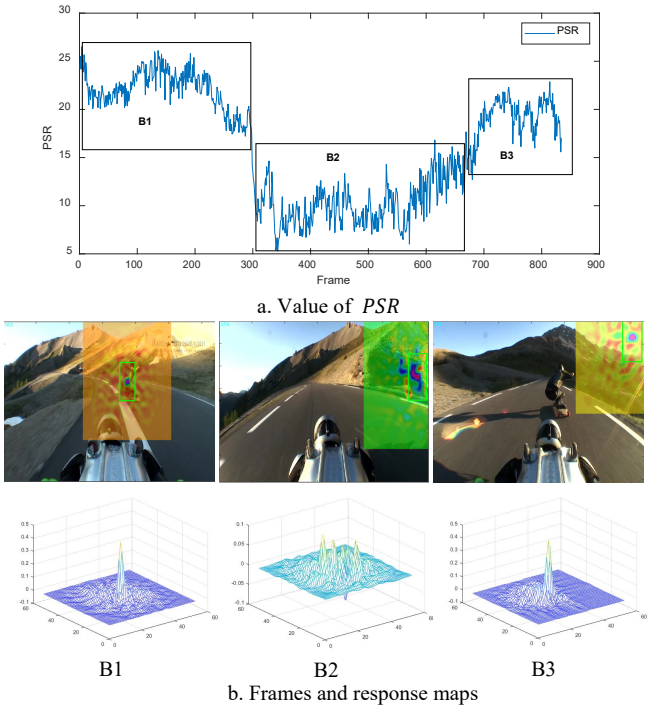


Figure 4. Credibility evaluation indicator PL with three states.

Combing with PSR and depth value of the human, the credibility indicator PL (Person is Lost) is set to evaluate the tracking result with three states, shown in the Figure 4. 1) PL=0. The tracking result is accurate. 2) PL=1. When $PSR < PSR_{\text{Aver}} \cdot w_{min}$ for a few frames and the human depth changes sharply compared with the previous frame, it is considered that the tracking result is not credible. 3) PL=2. When the human is lost for a long time, the re-tracking mechanism is activated to search for the target human. 4) From PL ≥ 1 to PL=0. When $PSR > PSR_{\text{Aver}} \cdot w_{max}$ for continuous multiple frames, it is believed that the target human is re-tracked successfully and PL returns to 0. $PSR_{\text{Aver}} * w_{max}$ is the threshold to judge whether the tracking result is accurate.

C. High-Confidence Update Strategy

The model of ECO-HC tracker is updated every Ns frame, which can increase the running speed and avoid model drift. However, when the object is out of FOV for multiple consecutive frames, the problem of model drift will still occur and the tracker may track a non-target object. Figure 5 demonstrates the process of model drift when object disappears for a long time. In the beginning, the tracker runs accurately and the response map maintains stable in part B1. Then the target disappears for a long time. The value of PSR decreases and response map fluctuates wildly as shown in B2. After a few of frames, the model drifts and PSR rises in B3. The response map returns stable and the wrong object is tracked.



a. Value of PSR

b. Frames and response maps

Figure 5. Model drifts when the object disappears for a long time.

The article adopts a high-confidence update strategy to prevent model drift. When PL$\geq$1, the filter model and samples stop updating to avoid being contaminated by the inaccurate samples. When tracking is successful and PL=0, the tracker updates the model and sample set every few frames.

D. Human Detection and Re-tracking Mechanism

The face recognition project Seetaface6 developed by Seetatech is used for face detection and recognition [12]. When the algorithm starts running, Seetaface6 is performed to obtain the face. Then the depth information is used to segment the upper body of the target human. After initializing the tracker, the robot starts to follow the human. When the tracking fails, the re-tracking mechanism is activated to obtain the target human through face matching with the preserved face picture.

IV. Experimental Results and Analysis

To validate the validity of the algorithm, a series of experiments were performed on an ROS robot. The robot is shown in Figure 6.

When the tracking result is accurate, the tracker obtains the two-dimensional coordinates on the RGB image and depth value of the human. The position of the human in pixel coordinate system is transformed to the three-dimensional

coordinate in the robot coordinate system. Then the angular velocity and linear velocity are calculated for the robot to follow the human.

Figure 6. ROS mobile robot.

A. Credibility Evaluation of the Tracking Result

Figure 7 shows the evaluation results of the human-following algorithm. Curve a means the pseudo-average value PSR_{Aver}, which remains relatively stable all the time. Curve b shows the changes of PSR. The value of PSR fluctuates with the deformation and disappearance of human. Curve c represents the depth of the human which remain stable when the tracking is successful. When the value of depth is out of the certain limit, it becomes -1 and the tracking result is considered unfaithful. Curve d is the credibility indicator PL which varies among 0,1 and 2.

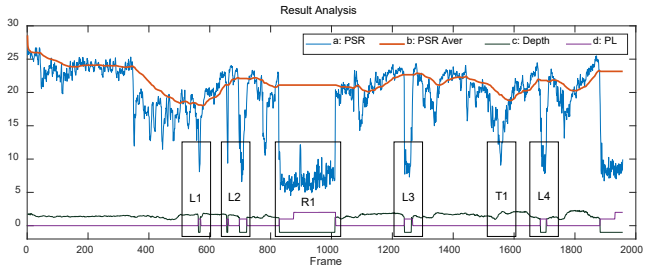


Figure 7. Indicators of the algorithm.

The part T1 in the Figure 7 indicates that the rapid deformation of human lead to a transient decrease in PSR. However, the depth value of human remained stable and it was considered that the target was not lost. In the Figure 8 which corresponds to part T1, the human moved quickly and turned around. The robot still followed the correct human. The box in the figure represents the predicted object position and size. The circle in the middle represents the center position and correlation. The more concentrated and darker the color is, the higher the credibility is.

Figure 8. Quick movement and turn of human.

The parts L1~L4 in Figure 7 indicate the short disappearance of the human. The process of short disappearance is shown in Figure 9. The PSR decreased rapidly and the depth of the human has a large jump. It was considered that the tracking result was not accurate. The robot stopped following human. The tracker model stopped updating. The re-tracking mechanism had not been activated. When the human returned to the camera's FOV, the tracker re-tracked the target human and the robot started to follow human again.

a. Blocked briefly by pedestrian.

b. Squatted briefly.

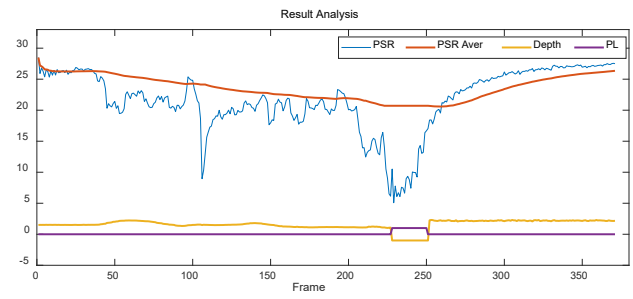
c. Disappeared briefly on the side.

Figure 9. The target human disappeared temporarily.

Figure 10. The target human disappeared for a long time and was re-tracked later.

The part R1 in Figure 7 indicates that when the human disappeared for a long time, the re-tracking mechanism was activated and the target human was re-followed successfully through face detection and matching. Figure 10 illustrates the process of human disappearing and being re-followed.

B. High-confidence Update Strategy



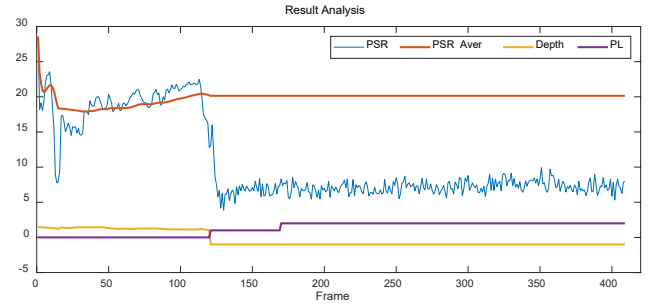
a. Indicators of the algorithm

b. Frames of human disappearance

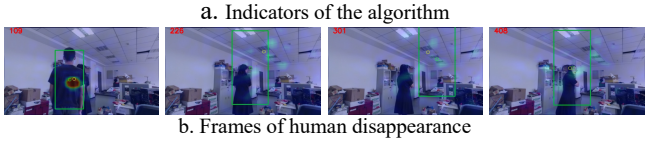
Figure 11. Human disappeared, model drifted and robot followed the non-target human.

Part a in Figure 11 and Figure 12 shows the changes of indicators when human disappeared for a long time. Part b shows the process with four frames. In Figure 11, where the tracker does not adopt the high-confidence update strategy, the PSR decreased firstly and rose again later. The model drifted and the robot started to follow the non-target human.

In Figure 12, a high-confidence update strategy is adopted to avoid model drift and the *PSR* fluctuated within a small range. The model did not drift and the tracking result was considered inaccurate.



a. Indicators of the algorithm



b. Frames of human disappearance

Figure 12. human disappeared, model did not drift, and robot did not follow the wrong object.

C. *Human Following in Complex Indoor Environments*

Human following experiments were carried out in some indoor environments. The 3D visualization tool of ROS (RVIZ) was used to record the routes of human and the robot on the map shown in the Figure 13. During the human following process, the human turned around, squatted, disappeared for a long time, etc. The robot could successfully distinguish different cases and follow the target human stably.

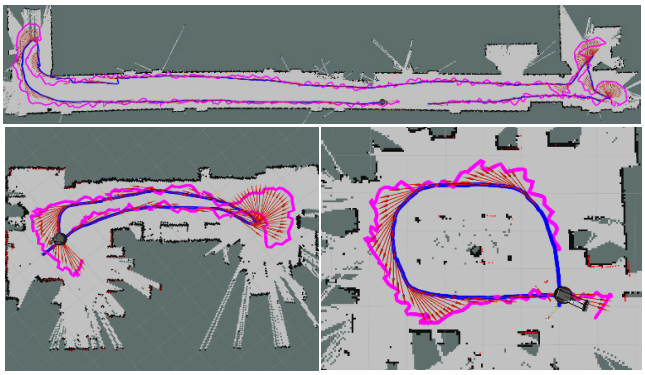


Figure 13. The route map. Red line: the predicted route of the target human; Blue line: the path of robot. The arrow: the orientation of the robot.

V. CONCLUSION

Aiming at the achievement of human following, the article applied the correlation filter object tracking algorithm ECO-HC on a robot. The Peak Side-lobe Rate and depth of the target human were used to evaluate the credibility of the tracking result. The strategy of updating the sample space and filter model with high confidence was adopted to avoid model drift. At the same time, face matching was used for re-tracking after tracking failure, which improves the stability and robustness of the target tracker. Based on the depth camera Kinect and a ROS robot, the experiments were performed to follow the human in different scenarios. The results show that the proposed method of following the human is feasible and has reliable abilities to detect tracking failure and re-follow the human. The robot can differentiate and deal with different cases, and stably follow human in a complex indoor environment based on the proposed method.

ACKNOWLEDGMENT

This work was supported by the Key R&D Program of Zhejiang Province, China under grant No.2017C01039.

REFERENCES

[1] D. S. Bolme, J. R. Beveridge, B. A. Draper and Y. M. Lui, "Visual object tracking using adaptive correlation filters," Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR), IEEE-CS Press, Aug. 2010, pp. 2544-2550, doi: 10.1109/CVPR.2010.5539960.

[2] J. F. Henriques, R. Caseiro, P. Martins and J. Batista, "Exploiting the Circulant Structure of Tracking-by-Detection with Kernels," Proc. the 12th European Conference on Computer Vision(ECCV) - Volume Part IV, Springer Press, Oct. 2012, pp.702-715, doi: 10.1007/978-3-642-33765-9_50.

[3] J. F. Henriques, R. Caseiro, P. Martins and J. Batista, "High-Speed Tracking with Kernelized Correlation Filters," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, 1 March 2015, no. 3, pp. 583-596, doi: 10.1109/TPAMI.2014.2345390.

[4] M. Danelljan, G. Häger, F. Khan, and M. Felsberg, "Accurate Scale Estimation for Robust Visual Tracking", Proc. the British Machine Vision Conference 2014, BMVA Press, 2014, doi: 10.5244/C.28.65

[5] M. Danelljan, G. Häger, F. S. Khan and M. Felsberg, "Discriminative Scale Space Tracking," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, 1 Aug. 2017, no. 8, pp. 1561-1575, doi: 10.1109/TPAMI.2016.2609928.

[6] M. Danelljan, A. Robinson, F. Shahbaz Khan, M. Felsberg, "Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking," Proc. the 14th European Conference on Computer Vision(ECCV), Springer Press, Sept. 2016, pp.472-488, doi: 10.1007/978-3-319-46454-1_29.

[7] M. Danelljan, G. Bhat, F. Khan and M. Felsberg, "ECO: Efficient Convolution Operators for Tracking," Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Press, Nov. 2017, pp. 6931-6939, doi: 10.1109/CVPR.2017.733.

[8] L. Bertinetto, J. Valmadre, J.F. Henriques, A. Vedaldi, P.H.S. Torr "Fully-Convolutional Siamese Networks for Object Tracking," Proc. the European Conference on Computer Vision Workshops (ECCVW), Springer Press, 2016, pp 850-865,doi: 10.1007/978-3-319-48881-3_56.

[9] P. Li, B. Chen, W. Ouyang, D. Wang, X. Yang and H. Lu, "GradNet: Gradient-Guided Network for Visual Object Tracking," Proc. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE Press, 2019, pp. 6161-6170, doi: 10.1109/ICCV.2019.0626.

[10] M. Danelljan, G. Bhat, F. S. Khan and M. Felsberg, "ATOM: Accurate Tracking by Overlap Maximization," Proc. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Press, 2019, pp. 4655-4664, doi: 10.1109/CVPR.2019.00479.

[11] Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester and Deva Ramanan, "Object Detection with Discriminatively Trained Part-Based Models," in Computer, vol. 47, no. 2, pp. 6-7, Feb. 2014, doi: 10.1109/MC.2014.42.

[12] Project Webpage. https://github.com/seetafaceengine/SeetaFace6, SeetaTech , "SeetaFace6".