© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2020.3006520, IEEE Transactions on Fuzzy Systems

Fuzzy Detection aided Real-time and Robust Visual Tracking under Complex Environments

Shuai Liu[1,2,3], *Shuai Wang*[1,2,3], *Xinyu Liu*[1,2,3], *Chin – Teng Lin*[4,*], *Fellow, IEEE*, *Zhihan Lv*[5], *Senior Member, IEEE*

***Abstract*—Today, new generation of artificial intelligence has brought several new research domains such as computer vision (CV). Thus, target tracking, the base of CV, has been a hotspot research domain. Correlation filter (CF) based algorithm has been the base of real-time tracking algorithms because of the high tracking efficiency. However, CF based algorithm are usually failed to track objects under complex environments. Therefore, this paper proposes a fuzzy detection strategy to pre-judge the tracking result. If the pre-judge process determines that the tracking result is not good enough in the current frame, the stored target template is used for following tracking to avoid the template pollution. Testing on the OTB100 dataset, the experimental results show that the proposed auxiliary detection strategy improves the tracking robustness under complex environment by ensuring the tracking speed.**

***Index Terms*—Fuzzy detection; Visual tracking; Real time; Complex Environment; OTB100**

I. INTRODUCTION

Visual object tracking is one important research topics in the field of computer vision, mainly to complete the corresponding tasks by simulating the visual function of living things [1]. The entire tracking process is expressed as follows. First, the image sequence of the target is captured, and the tracked target is determined in the first frame in the initial step. Then, the tracking algorithm is used to predict the position of the target in each subsequent frame to achieve the real-time tracking of the target automatically in the tracking process step. In recent years, due to the continuous improvement of target tracking technology, it has been widely used in many important fields, such as video surveillance, virtual reality, and intelligent transportation [2–5]. However, this technology still faces many challenges in the actual scenarios due to the influence of various factors such as morphological changes or illumination changes of the tracking target. Existing vision tracking algorithms can solve specific target tracking problems under certain interference conditions to a certain extent.

The Classical tracking algorithms mainly include: subspace-based target tracking algorithms, detection-based target tracking algorithms, sparse expression-based target tracking algorithms, deep neural network-based target tracking algorithms, and correlation filtering-based target tracking algorithms. Black and Jepson [6] first added subspace learning methods in the object tracking process and then used the trained subspace feature library based on different angles to model the appearance of the target. This method has a good effect on the problem of target appearance changes caused by lighting changes. In order to solve the problem of target drift in the tracking process, Grabner et al. [7] proposed an online lifting algorithm, which applied a semi-supervised method in the Online Adaboost framework to judge and identified the target by combining the given target prior information and online classification results. Mei et al. [8, 9] firstly adopted a sparse representation model and proposed an object tracking algorithm based on L1 norm minimization. This tracking algorithm linearly represents each candidate target by means of super-complete dictionary, calculates the reconstruction error between the candidate samples and the target template, and selects the candidate sample with the smallest reconstruction error as the tracking result to achieve a better tracking effect. Wang et al. [10] took the lead in applying Deep Neural Network (DNN) to target tracking. This method can enable neural networks to reproduce specific targets by combining online training and offline training.

In recent years, target tracking algorithms based on correlation filter (CF) have attracted more researches. Bolme et al. [11] applied the correlation filtering theory to online target tracking, and obtained a large number of samples through affine transformation and Gaussian distribution to construct an adaptive CF. They calculated the least square error between the actual correlation output and the expected correlation output to predict the position of the target. The algorithm is robust and computationally efficient in many complex scenarios. Henriques et al. [12] added ridge regression and cyclic matrices in the CF framework, used cyclic displacement to collect samples, and performed discrete Fourier transform diagonalization. This algorithm is more robust to complex scenes. In another work, their team [13] proposed a Kernel CF (KCF) tracking algorithm, which introduced the square gradient histogram feature (HoG). The tracking performance of the algorithm was significantly improved. Danelljan et

This work was supported by the Natural Science Foundation of Hunan Province with No.2020JJ4434, Key Scientific Research Projects of Department of Education of Hunan Province with No.19A312; Hunan Provincial Science & Technology Project Foundation (2018TP1018, 2018RS3065), Key Research and Development Plan - Major Scientific and Technological Innovation Projects of Shandong Province with No.2019JZZY020101, National Natural Science Foundation of China with No. 61502254 and 61902203, Open Project Program of the State Key Lab of CAD&CG under Grant A1926, Zhejiang University.

Prof Shuai Liu, Xinyu Liu, and Shuai Wang are with the Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, College of Information Science and Engineering, Hunan Normal University, as well as Hunan Xiangjiang Artificial Intelligence Academy, Changsha, 410000, China.

Prof. Chin-Teng Lin is with the Centre for Artificial Intelligence, FEIT, University of Technology Sydney, NSW 2007, Australia. (e-mail: Chin-Teng.Lin@uts.edu.au).

Prof. Zhihan Lv is with School of Data Science and Software Engineering, Qingdao University, Qingdao, 266000, China.

1063-6706 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Carleton University. Downloaded on July 28,2020 at 01:17:33 UTC from IEEE Xplore. Restrictions apply.

al. [14] added the color name feature [15] to the relevant CF framework, which has the characteristics of insensitivity to light and fast calculation speed. For the scale estimation problem, Li et al. [16] proposed a method combining color name features and gradient histogram features. In order to solve the problem of target occlusion, Liu et al. [17] proposed a local-based CF algorithm. In the algorithm, the target is divided into multiple regions and the corresponding CF is trained for each region. In order to cope with changes in target appearance, Hong et al. [18] applied a theory of perceptual psychology and combined a long-term and short-term memory module to propose a new object tracking algorithm.

The tracking process of the traditional CF based tracking algorithm is summarized as follows. First, cyclic shift around the target position of the previous frame is used to generate a large number of samples to train the filter. Then the trained filter for rapid detection is used to determine the position of the target by the response value of the filter in the subsequent frames. The position corresponding to the maximum response value is determined as the current position of the target (best matching position). The essence of CF based object tracking method is to reduce the non-target response value and increase the target response value to accurately position the target. The tracking speed has been greatly improved by applying Fast Fourier Transform. However, it will cause tracking failure in subsequent frames if the selected best matching position is incorrect, thereby affecting the performance of the tracking algorithm.

Therefore, this paper proposes a fuzzy aided strategy to add a pre-judge process in the CF based algorithms. Firstly, selected positive and negative samples are trained to generate a fuzzy detection system. Then, the fuzzy detection module is added to the tracking process to pre-judge whether the current tracking is good. It is to say that the fuzzy detection module is used to judge whether the best matching position is the actual position of the current target. If it is not a good matching, the human visual mechanism is adopted to track the lost target. The visual mechanism continuously learns and stores the correct target from continuous frames. When the tracking failure is detected by the proposed fuzzy pre-judge module, the possible position is tracked by matching the stored target template to achieve an effective relocation. The proposed algorithm provides a solution for the template pollution problem of the CF based algorithm.

The main work arrangement of this paper is given as follows. The second section mainly introduces the theoretical basis of the strategy proposed in this paper. The third section focuses on the strategies proposed in this article. In the fourth section, the compared experiments between classical algorithms and the proposed algorithm are proposed and analyzed on OTB100 dataset. Finally, the main contents of this paper are summarized and the outlook of the future work is prospected in the fifth section.

II. RELATED WORK

A. Fuzzy decision system

The fuzzy system is a generalization of the deterministic system, which includes the input, output and state variables defined on the fuzzy set. Sometimes it is difficult to establish an accurate mathematical model because the control object has characteristics such as uncertainty or nonlinearity. In such cases, the fuzzy system can make full use of fuzzy information knowledge from a macro perspective to convert the difficult nonlinear system into multiple linear models, and then use the membership function to connect each local subsystem to form a system model which is easy to deal with. Fuzzy system can better solve non-linear problems, and it has wide application and research prospects [19]. For example, Han [20] took the fuzzy control approach to deal with stability analysis and control design for a class of time delay-dependent nonlinear systems with input saturation. Munoz-Salinas et al. [21] tackled the tuning of the fuzzy membership functions of a fuzzy visual system for autonomous robots. Jong et al. [22] used the concept of fuzzy algorithm to analyze the fuzzy ratio captured by humans and proposed a new SIRMs-FIS model with fuzzy weights and partial to overall fuzzy ratio. Li et al. [23] proposed a fuzzy reasoning modeling method based on T-S fuzzy system. A mathematical model of a time-varying second-order free motion system is established to deal with the problem by constructing a new input-output model and a new state space model. Hajek et al. [24] proposed a new type of Takagi-Sugeno-Kang interval-valued intuitionistic fuzzy inference system (IVIFIS). Chibani et al. [25] proposed a filter to solve the Takagi-Sugeno fuzzy form of nonlinear discrete-time systems with faults and unknown inputs.

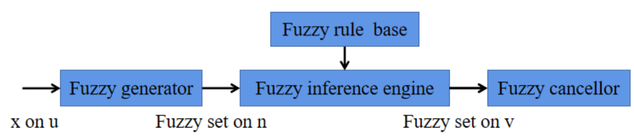


Fig.1. Schematic diagram of the composition of a Mamdani-type fuzzy decision system

The types of fuzzy systems are mainly divided into pure fuzzy logic systems, Takagi-Sugeno-type fuzzy logic systems, and Mamdani-type fuzzy systems. The most used Mamdani fuzzy system consists of three main components, which are fuzzy generator, fuzzy inference machine, and fuzzy eliminator (anti-fuzzy). It essentially adds a fuzzy generator and fuzzy canceller to the input and output parts of a pure fuzzy logic system. The structure is shown in Figure 1. First, data is converted into a fuzzy set by a fuzzy generator. Then, in the fuzzy inference machine, IF-THEN fuzzy rules are combined to map the fuzzy set from the input universe U to the output universe V according to the principle of fuzzy logic. Finally, the fuzzy set is transformed into a truth value variable by the defuzzifier.

The logical basis of the fuzzy system mainly includes fuzzy sets and membership functions. The fuzzy concept is described by fuzzy sets, which are defined as follows:

Given a domain U, any mapping μ_A from U to a closed interval of [0, 1], $\mu_A : U \rightarrow [0,1], u \rightarrow \mu_A(u)$ is called a fuzzy subset A of U, μ_A is called a membership function of the fuzzy subset. Moreover, $\mu_A(u)$ called u′s membership of A, which reflects the degree of u′s membership to the fuzzy subset A, and the fuzzy subset is also called fuzzy set.

Common membership functions include trigonometric functions, ladder functions, Gaussian functions, generalized bell functions, Sigmoid functions and Z functions. Among them,

1063-6706 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Carleton University. Downloaded on July 28,2020 at 01:17:33 UTC from IEEE Xplore. Restrictions apply.

the triangular membership function and the ladder membership function are piecewise linear functions, which are relatively simple to use and calculate; the Gaussian membership function and the bell membership function are relatively smooth curves, there is no zero point in the graph and it has relatively clear physical meaning. It is the most commonly used membership function. The Sigmoid membership function curve also has good smoothness, and it is different from the Gaussian membership function. The Sigmoid membership function is suitable for representing asymmetry. The Z-membership function is based on spline interpolation. Figure 2 shows a schematic diagram of triangular membership function and Gaussian membership function.

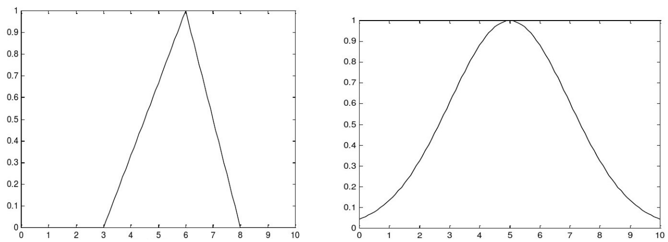


Fig.2. On the left is the triangular membership function, on the right is the Gaussian membership function

The determination of triangle membership function and Gaussian membership function can be expressed as Eq.1 and Eq.2 respectively.

$$f(x,a,b,c)=\begin{cases}0, & x\leq a\\ \frac{x-a}{b-a}, & a\leq x\leq b\\ \frac{c-x}{c-b}, & b\leq x\leq c\\ 0, & x\geq c\end{cases} \tag{1}$$

$$f(x,\sigma,d)=e^{-\frac{(x-d)^2}{2\sigma^2}} \tag{2}$$

The parameters a and c in Eq.1 determine the boundary of the triangle and the parameter b determines the “vertex” of the triangle. The parameter d is used to determine the center of the curve in Eq.2, and all parameters in Eq.2 are positive.

A membership function is mainly determined through experience and related experiments. A usual method is to roughly determine the membership function first, and then adjust the function through continuous learning from samples. There are several ways to determine the membership function.

(1) Fuzzy statistics method
The survey and statistics were carried out by proposing the fuzzy concept. The degree corresponding to different elements belong to a fuzzy set A is calculated through statistical experiments. The membership ψ of u′ to fuzzy set A can be expressed as Eq.3.

$$\psi=\frac{N'}{N} \tag{3}$$

Where N′ represents the number of $u' \in A$, and N represents the total number of experiments.

(2) Subjective experience method
When the domain is discrete, the membership is given based on subjective knowledge and experimental experience. The subjective experience method is widely used to determine the membership function.

(3) Neural network method
The membership function is automatically generated by a neural network. The value of the membership function is automatically adjusted through continuous learning of the network.

B. Human visual mechanism

Vision is the most important perceptual function of humans and higher animals. Most of the information comes from vision since the human society enters the information age. When the information enters vision, the human visual system filters the information quickly and efficiently. A small amount of information attracts the attention of the human cerebral cortex, and the rest of the information is ignored [26]. Lin et al. [27] proposed an improved neural network for automatic modulation classification in vehicular system, which divided human memory into three forms: perceptual, short-term and long-term memory. After the visual system collects information, it firstly forms sensory memory. Since sensory memory is the shortest type, a large amount of information is only stored in the human brain for a short time. Then, limited information is stored by short-term memory, which refers to the ability of the human brain to store in a short period of time. Finally, the important information is stored in the long-term memory, which has long memory time and unlimited memory content. Huang et al. [28] added visual memory to the intelligent transportation system to realize a short-term prediction of traffic flow. In response to the problem of occlusion of the target during tracking, Yun et al. [29] proposed a dynamic spectrum interaction of UAV flight formation with priority by using deep reinforcement learning approach. Yulita et al. [30] proposed a method that combines convolutional neural networks and bidirectional long-term and short-term memory to classify sleep stages. The schematic diagram of human visual memory mechanism is shown in Figure 3.

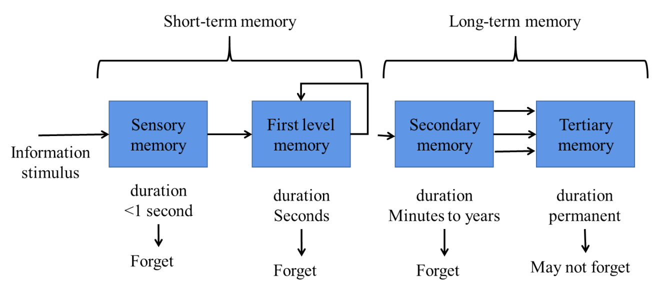


Fig.3. Schematic diagram of long short-term memory

The main idea of introducing human visual system into target tracking is as follows. The system screens and memorizes the appearance features of the current target, and stores relevant information in the feature information database during the tracking process. Then, the database is improved by constantly learning new features. Eventually, the long-term memory of the characteristics of the target is realized to further deepen the cognitive ability of the target. When the target tracking fails, the location of the target is analyzed and judged according to the characteristics of the target stored in the information database to realize the target relocation.

1063-6706 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Carleton University. Downloaded on July 28,2020 at 01:17:33 UTC from IEEE Xplore. Restrictions apply.

C. Correlation filter tracking algorithm

CF were first used in the field of signal processing. Correlation is used to measure the degree of similarity between two signals h and f by performing a correlation operation. Through Fourier transform, the operation of judging the similarity of two signals can be expressed as a dot product operation in the frequency domain.

Based on the application of the correlation filter in the field of signal processing, it is now applied to the field of object tracking. The steps of the correlation filtering object tracking algorithm are as follows. At first, the tracking target is manually marked in the first frame. Then, feature extraction is performed on the target candidate region. In next step, a cosine window is added into the feature for smoothing. Finally, a well-trained CF is constructed to detect candidate regions in the frequency domain. The position with the highest response value is tracked as the current position of the target. The essence of the algorithm is to build a target template, determine the position of the target by calculating the similarity of target template, and achieve continuous tracking by updating the template with the position of highest similarity in subsequent frames. The general flow of the algorithm is shown in Figure 4.

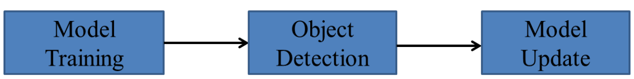


Fig.4. Flow chart of classic correlation filtering algorithm

(1) Model training

In the algorithm, the training sample matrix can form a cyclic matrix. A vector $v = [v_0, v_1, v_2, \cdots, v_(n-1)]^T$ is used as the base positive sample, and the offset transformation samples generated by the base positive sample is used as the negative sample. The positive and negative samples are used to train the classifier. The cyclic shift matrix is shown in Eq.4. The vector is moved by one element to get the offset vector in Eq.5, and then a cyclic shift is performed to obtain a cyclic matrix C(v) in Eq.6.

$$P = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{4}$$

$$v = \begin{bmatrix} v_{n-1}, v_0, \cdots, v_{n-3}, v_{n-2} \end{bmatrix}^T \tag{5}$$

$$V = C(v) = \begin{bmatrix} v_0 & v_1 & v_2 & \cdots & v_{n-1} \\ v_{n-1} & v_0 & v_1 & \cdots & v_2 \\ v_{n-2} & v_{n-1} & v_0 & \cdots & v_1 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ v_1 & v_2 & v_3 & \cdots & v_0 \end{bmatrix} \tag{6}$$

The n-row elements in the cyclic matrix represent n training samples. The training process of the samples is actually considered as a ridge regression problem, which has a closed-form solution. Since the ridge regression has a corresponding closed optimal solution for any input in Eq.7, the training , which is used to find an objective function $f(v_i) = \omega^T v_i$ to minimize the error function between the training sample and the response value, only needs a small calculations. Then, it is reformed as the matrix equation by Eq.7:

$$min_\omega \sum_{i=1}^{m} (V\omega - y)^2 + \lambda \|\omega\|^2 \tag{7}$$

By deriving the error function, a closed optimal solution can be obtained as Eq.8.

$$\omega = \left(V^T V + \lambda I\right)^{-1} V^T y \tag{8}$$

where, the cyclic matrix V is composed of all the training samples v, and y is the response value corresponding to all the training samples.

When the solutions are non-linear, it is a difficult problem for directly solving the equation. In order to transform a linear problem into a nonlinear problem, the feature space of the input training samples is usually mapped to a high-dimensional space in Eq.9, which is expressed as a nonlinear combination of v.

$$\omega = \sum_i \alpha_i \varphi(v_i) \tag{9}$$

where, $\varphi(v_i)$ represents a non-linear mapping function, α_i represents a coefficient of a linear combination obtained after all samples are mapped. Then, the kernel function is defined as Eq.10.

$$k(v, v') = \langle \varphi(v), \varphi(v') \rangle \tag{10}$$

Where, $\langle \bullet, \bullet \rangle$ represents the inner product between any two functions.

(2) Object detection and model update

In the next frame, according to the position and proportion of the target x determined in the current frame, a large number of candidate samples are generated using cyclic shift to determine the target candidate area. For each candidate sample y, the response value is calculated according to Eq.11.

$$\hat{y} = \hat{k}^{xy} \odot \hat{\alpha} \tag{11}$$

where, the response value $\hat{y}$ is shown after Fourier transform, and the position with the largest response value is selected as the position where the real target center is located.

In the actual object tracking scenario, the target may be greatly change due to various kinds of reasons. In order to prevent target drift, which eventually causes object tracking failure, the filter is updated during the tracking in real time. Therefore, the whole template update steps are performed at the new target position after the true position of the current frame target is determined. Meantime, the extracted features are used to update the target appearance model and classifier coefficients.

III. A FUZZY DETECTION AIDED TARGET TRACKING ALGORITHM

Since the problem of target template pollution exists in traditional CF based tracking algorithm, a fuzzy detection system is proposed in CF based target tracking, in order to improve the deficiencies of traditional target tracking algorithms.

1063-6706 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Carleton University. Downloaded on July 28,2020 at 01:17:33 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2020.3006520, IEEE Transactions on Fuzzy Systems

A. Extraction of positive and negative samples

In order to train a fuzzy detection system in target tracking, the first step is to define positive and negative samples. In this paper, the definition rules of positive and negative samples of the proposed algorithm are as follows. In the traditional CF based target tracking algorithm, if both a tracked target in the previous frame and current frame is correct, the tracked target in current frame is determined as a positive sample; otherwise, if a tracked target of the previous frame is correct, and is incorrect in the current frame, the tracked target in current frame is determined as a negative sample.

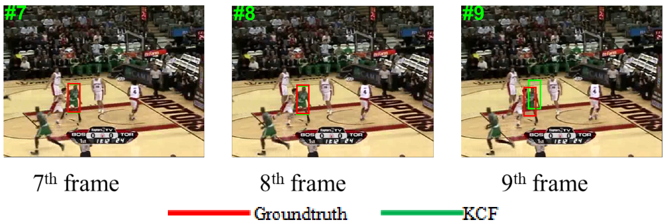


Fig.5. The 7th to 9th frame tracking images in video “basketball”

Figure 5 shows tracking result of KCF in continuous three frames with a video “basketball”. In Figure 5, the manually labeled position in the 7^{th} frame and the target position are actually tracked by the KCF basically coincide, which indicates that it is a good tracking. The same goes for the 8^{th} frame. However, the manually labeled position in the 9^{th} frame has a large deviation from the actual KCF tracking position, which indicates that it is a bad tracking. Both the 7^{th} frame and the 8^{th} frame are tracked correctly, therefore, the 8^{th} frame is defined as a positive sample. Then, since the 9^{th} frame is not tracked correctly, it is defined as a negative sample.

The positive and negative samples determined by the tracking situation of the previous and the current frames are of great help in training the fuzzy system. From the “correct tracking” to “correct tracking” makes the proposed fuzzy system learns the feature of good track. Moreover, from the “good tracking” to “bad tracking” makes the fuzzy system learns the failed tracking result exactly.

The definitions of positive and negative samples are expressed by following formulas.

$$\sqrt{(x_t - gt(x_t))^2 + (y_t - gt(y_t))^2} < \varepsilon \tag{12}$$

$$\sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2} > \sqrt{(h/\theta)^2 + (l/\theta)^2} \tag{13}$$

where, $(gt(x_t), gt(y_t))$ represents the actual location of the target in the current frame (the position of the ground-truth). (x_t, y_t) represents the tracking position of the target in current frame. (x_{t-1}, y_{t-1}) represents the tracking position of the target in the previous frame. l and h represent the width and height of the target frame respectively. θ and ε are two parameters to measure “good tracking” and “bad tracking”.

In the proposed method, a positive and negative sample ratio of 1: 1 is adopted according to the rules of sample selection. This is because the number of positive samples is much greater than the number of negative samples. Then, a certain number of positive and negative samples are selected for training the proposed fuzzy system in combination with the video tracking map. The response diagrams of some positive and negative samples selected from the video sequence “blurowl”, “couple”, “deer” are shown in Figure 6. The left side represents the response matrices of positive samples, and the right side shows the negative response matrices of the same sequence corresponding to the positive response graphs.

B. Selection of eigenvalues

The proposed fuzzy system recognizes positive and negative samples by using their response matrices. According to the differences between the positive and negative samples, the definition and selection of each feature, as well as the determination of parameters in the proposed fuzzy system are selected, which include the maximum response value, the larger response value ratio, the gradient values in x and y directions, the ratio of the maximum response value to the center of the target frame, and the variance.

(1) Number of maximum response value N_{mr}

During target tracking in the current frame, each candidate block in the given area needs to be compared with the tracked target template in the previous frame, and all response values r_{ij} between each sample and the target template are calculated to combine a response matrix R in Eq.14.

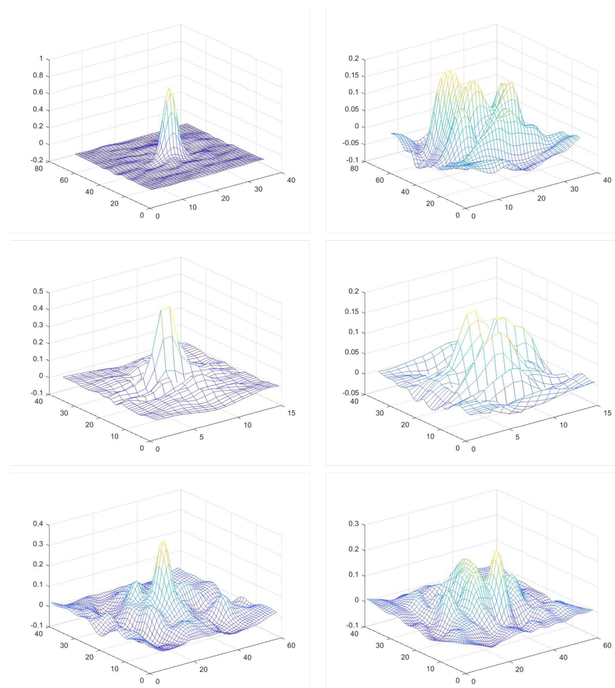


Fig.6. Response matrices of positive and negative samples in “blurowl” (upper), “couple” (medium) and “deer” (bottom) sequences, left are for positive samples, right are for negative samples

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} \tag{14}$$

The response matrix R’ is finally generated by decentralizing the boundary effect of the response matrix. In general, a response matrix may have multiple maximum response values.

In Figure 7, the left side is a single-peak response matrix and the right side is a multi-peak response matrix. Therefore,

1063-6706 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Carleton University. Downloaded on July 28,2020 at 01:17:33 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2020.3006520, IEEE Transactions on Fuzzy Systems

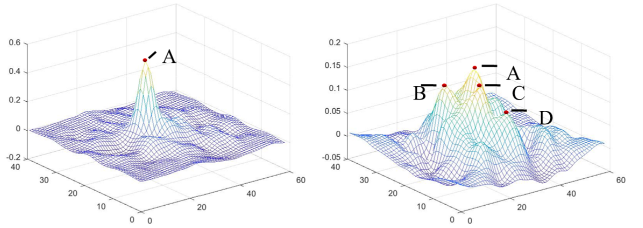


Fig.7. Different response matrix, the single-peak response matrix is on the left and the multi-peak response matrix is on the right

because the existence of multiple extreme points is likely to interfere with the determination of the true position of the target, the regional maximum is used as a key feature on the decentralized matrix.

For the response matrix with multi maximum peaks, a Boolean matrix is used to reform the response matrix. The maximum value in the matrix R' is defined as 1 and the other elements are defined as 0, thereby obtaining a Boolean matrix R" in Eq.15.

$$R'' = \begin{cases} 1, & R'(i,j) > RES' \\ 0, & else \end{cases} \tag{15}$$

where RES$'$ resprents the adjacent response value.

Then, a Boolean matrix R1 is calculated according to Eq.16, and the maximum response value $resmax_1 = max(R_1)$ is got. In this way, the ith largest response value $resmax_i = max(R_i)$ is obtained by matrix R_i which is generated by assigning the maximum value to 0 in R_{i-1}.

$$R_1 = R'.*R'' \tag{16}$$

The determination of the target position has a great relationship with the peak value according to the CF based algorithms, therefore, the number of maximum response value N_{mr} is calculated in Eq.18 as a feature of the target in the fuzzy system.

$$flag(i) = \begin{cases} 1, & i > 0 \\ 0, & i = 0 \end{cases} \tag{17}$$

$$N_{mr} = \sum_i flag(resmax_i) \tag{18}$$

(2) Large response value ratio R_L

The number of maximum values in the response matrix is not unique. The response matrix has a high probability leading to a "bad tracking" if most maximum values are small. Therefore, it is necessary to set a threshold τ to measure the ratio of large response value. Since the maximum value greater than τ is counted as the large response value N_L, the large response value ratio is calculated by Eq.19.

$$R_L = \frac{N_L}{N_{mr}} \tag{19}$$

The reason for using the proportion of larger response value as another input feature in fuzzy detection systems is that the relatively large response values greatly help distinguishing positive and negative samples. (3) Gradient values in *x* and *y* directions G_x, G_y

The gradient values in the *x* and *y* directions can usually be used to reflect the information difference between the position of the maximum response value and the surrounding positions. The gradient value represents the steepness of the maximum response value. The larger the gradient value is, the steeper the maximum response position is. Gradient values in the *x* and *y*-direction are defined by Eqs.20-21.

$$G_x = \frac{\partial R}{\partial x} = \begin{cases} \frac{r_{x,y}-r_{x-1,y}}{\Delta x} \\ \frac{r_{x,y}-r_{x+1,y}}{\Delta x} \end{cases} \tag{20}$$

$$G_y = \frac{\partial R}{\partial y} = \begin{cases} \frac{r_{x,y}-r_{x,y-1}}{\Delta y} \\ \frac{r_{x,y}-r_{x,y+1}}{\Delta y} \end{cases} \tag{21}$$

(4) The distance ratio between the maximum response value to the target in the previous frame D_r

A large number of samples are generated after cyclic shifting around the real position of the target in the previous frame. These samples constitute of a target search area, and the distance between each sample to the center of the frame is different. It is illogical if the candidate with the highest similarity is simply used as the best position ignoring the distance between the targets in continuous frames. Therefore, the distance ratio between the maximum response value to the target in the previous frame is considered in Eq.22 when constructing the fuzzy system, where (C_x, C_y) is the center of tracked target in the current frame, (C'_x, C'_y) is the center of tracked target in the previous frame, ω is the width of the frame, and *h* is the height of the frame.

$$D_r = \frac{|C - C'|}{|\omega + h|} \tag{22}$$

(5) Variance γ

Variance represents the mean square error value, which is the difference between the maximum values which are larger than a threshold. The difference between each maximum value is represented by the sum of the squares of the differences between the response values greater than a given threshold *val*. The variance γ is also used as an important input feature in the proposed fuzzy system by Eq.23.

$$\gamma = \sum_{i=1}^{number_2} (R'max - R(i))^2 \tag{23}$$

where, $number_2$ indicates the number of maximum responses that are greater than the threshold *val*, R'max indicates the maximum response value, *R*(i) indicates the i^{th} maximum response value.

Finally, the normalized features extracted from the positive and negative samples are put into the proposed fuzzy detection system for continuously training and testing. The features weight graph is shown in Figure 8.

C. Construction of the proposed fuzzy system

The construction of fuzzy system mainly includes the following parts:

(1) Determine the basic attributes

The basic properties of fuzzy systems mainly include fuzzy set composition operation, implication calculation, output composition operation, and inverse fuzzification calculation.

1063-6706 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Carleton University. Downloaded on July 28,2020 at 01:17:33 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2020.3006520, IEEE Transactions on Fuzzy Systems

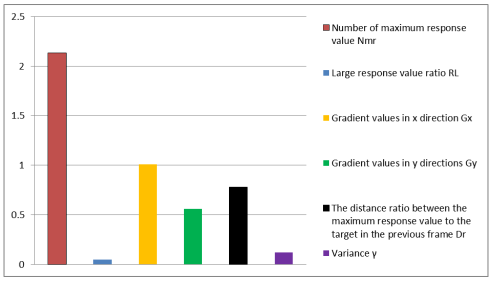


Fig.8. Weight of imput eigenvalue in the proposed fuzzy system

- Fuzzy set composition operations (computation of conjunctions) include:

$$and\left\{\begin{array}{c} min \\ prod \end{array}\right. \quad or\left\{\begin{array}{c} max \\ prober \end{array}\right. \tag{24}$$

- Application includes min method and prod score method;
- Aggregation of Outputs (Method for Fuzzy Rule Synthesis)

$$Aggregation\left\{\begin{array}{c} max \\ sum \\ prober \end{array}\right. \tag{25}$$

- The inverse fuzzification operation includes following Eq.26.

$$Defuzzification\left\{\begin{array}{c} centroid \\ bisector \\ mom \\ lom \\ som \end{array}\right. \tag{26}$$

In the set synthesis operation of the Mamdani-type fuzzy system constructed in this paper, the basic attribute of the "and" conjunction is the "min" method and the basic attribute of the "or" conjunction is the "max" method. The "min" attribute is used for the application calculation, and the composite operation of the output (Aggregation) uses "sum" method. Defuzzification uses centroid gravity center method.

(2) Determine the membership function of input and output variables

In fuzzy inference systems, membership functions are added to existing linguistic variable values, which mainly include triangular membership functions (trimf), trapezoid membership functions (trapmf), Gaussian membership functions (gaussmf) and so on. The data are discrete by observing the normalized feature values. In the proposed fuzzy detection system, it is determined that the membership function used is a triangular membership function.

(3) Establish fuzzy rules to constrain the output variables.

Fuzzy control rules are actually a collection of fuzzy conditional statements that are summarized by the operator's control experience. The principle of determining fuzzy control rules is to ensure that the output of the controller ensure the best dynamic and static characteristics of the system's output. The forms of fuzzy control rules is divided into the following two types:

- The main fuzzy control rules for state assessment are as follows.
  Ri: if x_1 **is** A_{i1} **and** x_2 **is** A_{i2} **... and x**$_n$ **is** A_{in} **then** y **is** C_i

  Among them, x_1, x_2, ..., x_n and y are linguistic variables or fuzzy variables, which represent the state variables and control variables of the system. And A_{i1}, A_{i2}, ... , A_{in} and C_i are linguistic values, which represent fuzzy sets in the domain.
- The objective evaluation fuzzy control rules is as follows.
  Ri: if (*U* is $C_i \rightarrow$ **(*x* is** A_1 **and *y* is** B_1**)) then *U* is** C_i

The general control rules are shown in Table 1. This paper establishes fuzzy rules according on the fuzzy rule table.

TABLE 1
TABLE OF FUZZY RULES

E				***U***			
DE	*NB*	*NM*	*NS*	*ZO*	*PS*	*PM*	*PB*
NB	*PB*	*PB*	*PB*	*PB*	*PM*	*ZO*	*ZO*
NM	*PB*	*PB*	*PB*	*PB*	*PM*	*ZO*	*ZO*
NS	*PM*	*PM*	*PM*	*PM*	*ZO*	NS	*NS*
ZO	*PM*	*PM*	*PS*	*ZO*	*NS*	*NM*	*NM*
PS	*PS*	*PS*	*ZO*	*NM*	*NM*	*NM*	*NM*
PM	*ZO*	*ZO*	*ZM*	*NB*	*NB*	*NB*	*NB*
PB	*ZO*	*ZO*	*ZM*	*NB*	*NB*	*NB*	*NB*

D. Strategy for optimizing recognition rate of fuzzy detection system

In the fuzzy detection system in this paper, the feature information of the normalized response matrix is continuously trained in the fuzzy detection system. With continuous training and testing, the accuracy of the positive and negative sample recognition reaches the highest rate 95.5% when the input of the controller is N_{mr}, G_x, and D_r. Therefore, the input of the fuzzy controller is set to these three eigenvalues. The overall framework of the fuzzy system generated on the Matlab R2017 platform is shown in Figure 9.

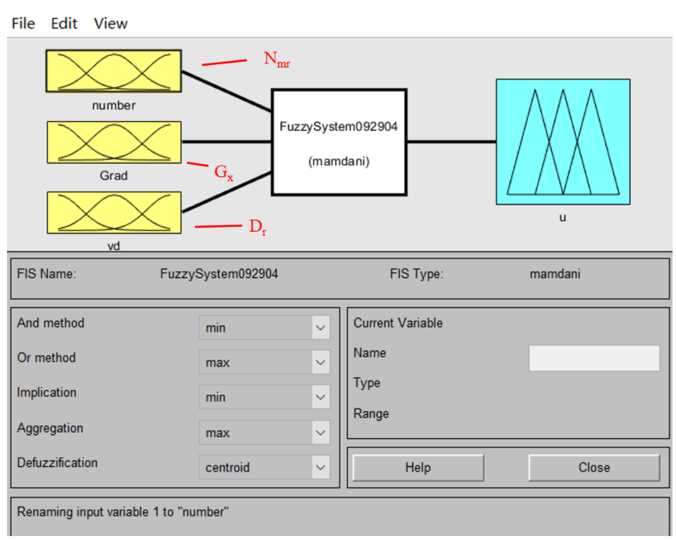


Fig.9. Structure of the proposed fuzzy detection system

By observing the distribution of the input variables, N_{mr} is basically concentrated between 0-1, the difference between G_x and the resmax is not large, the value of D_r is concentrated between 0-0.6. Therefore, each input variable is divided into

1063-6706 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Carleton University. Downloaded on July 28,2020 at 01:17:33 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2020.3006520, IEEE Transactions on Fuzzy Systems

five fuzzy sets: NS, NB, O, PS and PB. The value range of the fuzzy set is in [0, 1], and the membership range for each fuzzy set is assigned. Since the number of input variables is 3, we have 125 fuzzy rules which are shown in Figure 10.

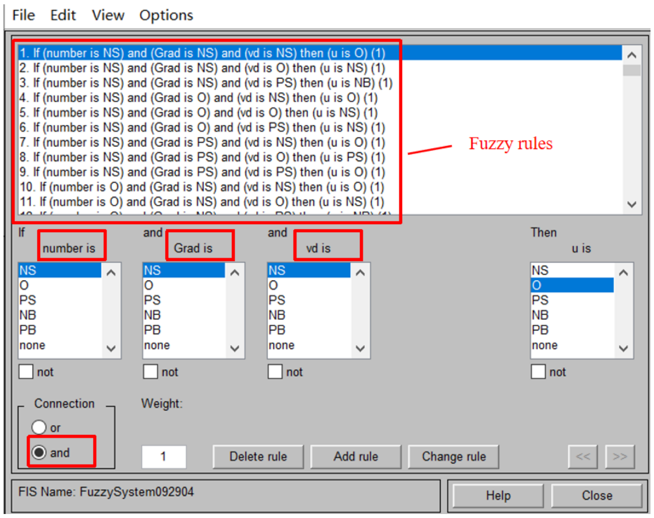


Fig.10. Fuzzy rule graph generated on MatlabR2017 platform

E. Improve CF based algorithm aided by the proposed fuzzy detection system

The traditional CF based tracking determines the position and scale of the target by the first frame of the video, and cyclically samples the target of the first frame to obtain training samples. For each subsequent frame, the position of the target is determined by calculating the response value between the candidate frame of the given area and the target template determined in the previous frame. Finally, the position with the largest response value is directly regarded as the position of the target of this frame.

The proposed method relies on vision to describe and memorize different types of features in the target, such as color, HOG, edge features and so on. Afterwards, the target model is reconstructed by statistical learning technology. After introducing the human visual mechanism, the apparent characteristics of the target are continuously learned and memorized for a long time. In the tracking process, a fuzzy detection system is added to detect if the current tracking is “good” by using the response matrix. If the target is blocked by an obstruction or exceeds the field of view at the t^{th} frame, the tracked target at this time is most likely incorrect. In order to track the target more accurately, it is necessary to find the accurate template memorized in the previous frame (t-1 frame) by using the storage mechanism, and then use it as a backup template for subsequent tracking. Especially when the target appears again, the backup target template is used to match the real template, and the target is relocated correctly. The tracking algorithm has the same capability as the human brain by introducing the human visual mechanism. That is the ability of visually remembering target and then reproducing the target. The diagram of the proposed object tracking algorithm (CF_based_Fuzzy) is shown in Figure 11.

The steps of the CF_based_Fuzzy algorithm are given as following steps.

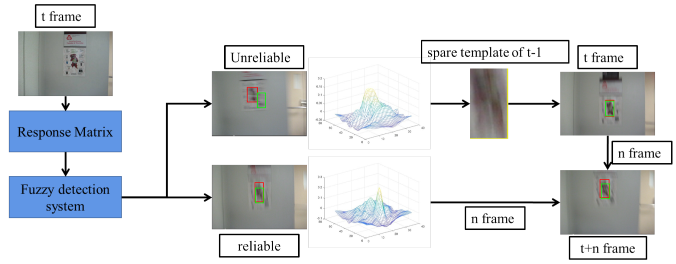


Fig.11. Framework of CF_based_Fuzzy algorithm

Step 1. In the fast detection phase, use the trained CF model to calculate the response matrix of target in each frame under the current target template;

Step 2. Put the response matrix into the fuzzy detection system to judge the current tracking result based on the obtained reliable information;

Step 3. According to the results of Step 2, adopt different tracking strategies to achieve accurate tracking of the target.

The specific flow is given in **Algorithm 1.**

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimental environment and evaluation criteria

In order to verify the effectiveness and real-time performance of the proposed algorithm (CF_based_Fuzzy), a one-time test (OPE) was used on the OTB platform to evaluate the algorithm and the traditional CF based algorithms. OTB databases are used in this experiment, which include OTB2013 (OTB 50) and OTB2015 (OTB 100). The video sequence on the OTB platform contains 11 different sub-challenges, which are Illumination Variation (IV), Scale Variation (SV), Occlusion (OCC), Deformation (DEF), Motion Blur (MB), Fast Motion (FM), In-Plane Rotation (IPR), Out-of-Plane Rotation (OPR), Out of View (OV), Background Clutter (BC) and Low Resolution (LR). Each challenge represents the target in video is under the corresponding complex scene. The software used for the evaluation algorithm is MATLAB R2017. The hardware environment is Inter® Core^{TM} i5-7500 CPU, 3.40GHz clock frequency and 64 GB of memory. The test uses accuracy and success charts as evaluation indicators:

(1) The precision refers to the percentage of frames where the distance between the position (a_i, b_i) of the target in the current frame and the actual position (a_t, b_t) of the manually marked target is less than a given threshold. The calculation method is as the following Eq.27 and Eq.28.

$$\beta = \sqrt{(a_i - a_t)^2 + (b_i - b_t)^2} \tag{27}$$

$$\xi = \frac{\sum_{i=1}^{n} M_i}{M_{total}} \quad M_i = \begin{cases} 1, & \beta < 20 \\ 0, & \beta \geq 20 \end{cases} \tag{28}$$

Among them, the threshold β is generally set to 20 pixels.

(2) The success rate is the area overlapping rate between the bounding box η_t of the real location and the bounding box η_k of the tracking result of the algorithm. The calculation method is as the following Eq.29:

$$S = \frac{|\eta_t \bigcap \eta_k|}{|\eta_t \bigcup \eta_k|} \tag{29}$$

1063-6706 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Carleton University. Downloaded on July 28,2020 at 01:17:33 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1 : CF_based_Fuzzy Algorithm

Input:

x: Training image patch, $m \times n \times c$
y: Back to goal, Gaussian, $m \times n$
z: Testing image patch, $m \times n \times c$

Output:

Response matrix: the center position (x_t , y_t) of the target in each frame.

```
Repect:
    Detect:
    step1:
        Construct the search window of the current frame;
        Collect samples in a circular shift around the sub-windows;
        Calculate the response values of all the samples and the target template;
        Find the maximum response value resmax of the current frame template.
    step2:
        Determine the reliability t1 of the maximum response value obtained in step1 by using the trained fuzzy detection system.
    if t1 > 0 then
        Use original template;
        Select the position corresponding to the maximum response value of the current frame as the true center position (x_t, y_t) of the target.
    else
        Select the target template of the previous frame memorized as the backup template;
        Compare with the target template of the current frame to get the maximum response value resmax;
        Compare with the target template of the previous frame to get the maximum response value resmax1.
        if resmax1 of the backup template > θ then
            Use backup template;
            Select the position corresponding to the maximum response value of the backup template as the true center position (x_t, y_t) of the target.
        else
            Use original template;
            Select the position corresponding to the maximum response value of the target template of the current frame as the true center position (x_t, y_t) of the target.
        endif
    endif
    step3:
        According to the maximum response value in step2, accurately track the target.
    Trained:
        Update the template in real time.
Until: the end of image sequence
```

B. Analysis of results

(1) Quantitative analysis

In order to objectively evaluate the performance of the algorithm, we analyze the overall performance and some sub-attributes of the algorithm. In this paper, both KCF and CSK combining fuzzy aided detection mechanisms are compared with the traditional KCF and CSK algorithms on the OTB platform.

First, KCF combining the proposed fuzzy detection system (KCF_based_Fuzzy) is compared with traditional KCF algorithms to evaluate the effectiveness of the proposed strategy. A comparison of the overall performance of the two algorithms is shown in Figure 12. In Figure 12, the accuracy of the proposed fusion algorithm on the OTB100 platform is 0.704, which is 1.00% higher than the KCF algorithm under the same situation. The success rate is 0.564, and the KCF algorithm success rate is 0.555. The contrast increased by 1.62%. Therefore, we know that the proposed fuzzy aided detection system improves the tracking efficiency of traditional algorithms to a certain extent.

Especially, the proposed KCF_based_Fuzzy shows great improvement under certain challenges, such as "Low Resolution", "Out of View", and "Motion Blur".

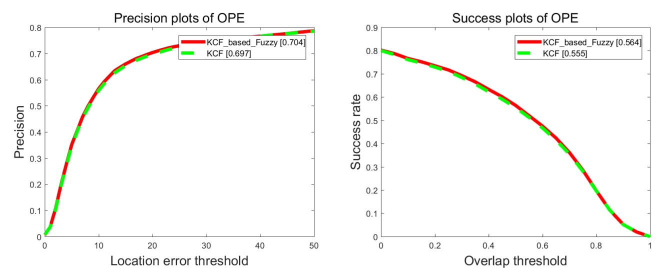


Fig.12. Comparison of accuracy and success rate between KCF_based_Fuzzy and KCF in the OTB100 dataset

For the challenge "Low Resolution", the improved algorithm KCF_based_Fuzzy introduces a blur detection system and storage mechanism to perform better tracking performance than the traditional KCF algorithm. In Figure 13, the accuracy increased from 0.560 to 0.575, an increment of 2.60%. The success rate increased from 0.295 to 0.304, an increment of 3.05%. The effectiveness of the proposed algorithm is proved.

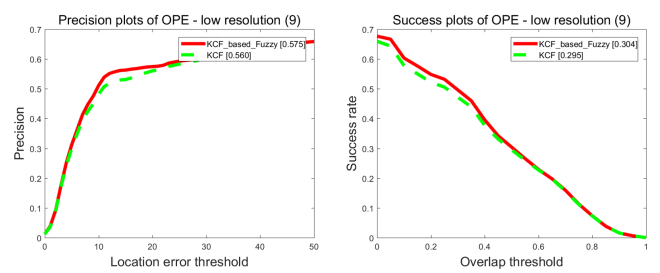


Fig.13. Comparison of accuracy and success rate between KCF_based_Fuzzy and KCF under challenge "Low Resolution"

For the challenge "Out of view", the comparison between the proposed KCF_based_Fuzzy and traditional KCF is used to predict the tracking result in Figure 14. The proposed tracking algorithm introduces the storage mechanism and the fuzzy detection mechanism to significantly improve the tracking performance. Compared with the traditional kernel correlation filtering algorithm, the accuracy is improved from 0.501 to 0.546, an increment of 8.98%. The success rate is increased from 0.457 to 0.496, an increment of 8.53%.

1063-6706 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Carleton University. Downloaded on July 28,2020 at 01:17:33 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2020.3006520, IEEE Transactions on Fuzzy Systems

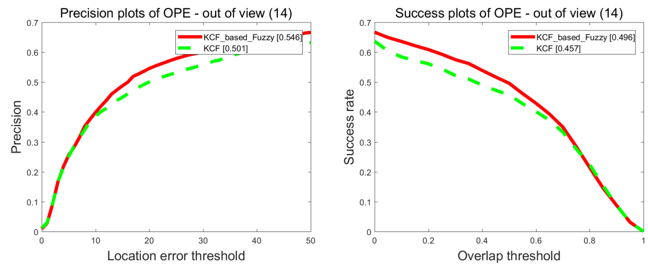


Fig.14. Comparison of accuracy and success rate between KCF_based_Fuzzy and KCF under challenge “Out-of-View”

For the challenge “Motion Blur”, the comparison between the proposed KCF_based_Fuzzy and traditional KCF is used to predict the tracking result in Figure 15. The proposed KCF_based_Fuzzy algorithm improves accuracy and success rate by 1.33% and 3.82% compared with the traditional KCF algorithm, it also validates that the proposed fuzzy aided strategy improves the tracking under challenge “Motion Blur” to a certain extent.

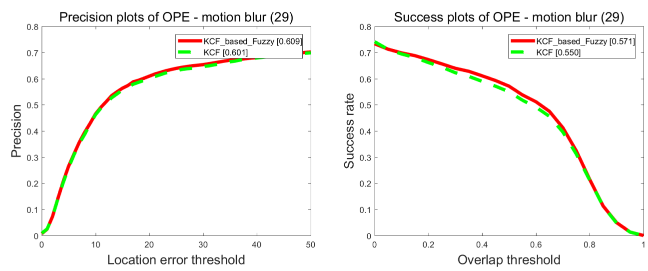


Fig.15. Comparison of accuracy and success rate between KCF_based_Fuzzy and KCF under challenge “Motion Blur”

In order to validate the robustness of the proposed strategy, the proposed CF_based_Fuzzy is further combined to the FDSST (Scale Filter) algorithm [31] to obtain an improved algorithm (FDSST_based_Fuzzy). At the same time, we also evaluated our KCF_based_Fuzzy and the latest algorithm KCF_reliable algorithm [32]. In this paper, both the four algorithms KCF_reliable, KCF_based_Fuzzy, FDSST, FDSST_based_Fuzzy are compared. Experimental evaluation is also performed on OTB100, and the result is shown in Figure 16 below:

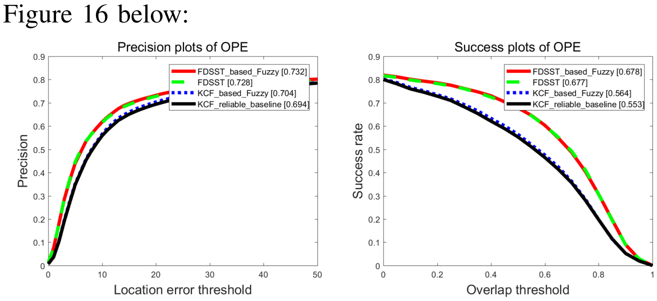


Fig.16. Comparison of accuracy and success rate between KCF_based_Fuzzy, KCF_reliable, FDSST_based_Fuzzy and FDSST in the OTB100 dataset

From the overall performance of the OTB100 evaluation set, the precision of FDSST_based_Fuzzy is 0.732, which is 0.549% higher than the accuracy of the original FDSST algorithm is 0.728. The success rate of FDSST_based_Fuzzy is 0.678, which is 0.148% higher than the traditional FDSST algorithm. The accuracy of the KCF_based_Fuzzy algorithm combined with the fuzzy detection system is 0.704, while the accuracy of the KCF_reliable algorithm is 0.694, which is an increase of 1.44%. The success rate of the KCF_based_Fuzzy algorithm is 0.564, while the success rate of the KCF_reliable algorithm is 0.553, which is an increase of 1.99%. Both the accuracy of the improved algorithms FDSST_based_Fuzzy and KCF_based_Fuzzy are higher than the traditional algorithms. By comparing the results of the two algorithms, it is verified again that the proposed method can improve the robustness of the algorithm.

(2) Qualitative analysis

The key frames in the video sequences ”blurowl”, ”box”, ”surfer” and ”twinnings” are selected from the OTB100 to perform a qualitative analysis on the performance of the proposed algorithm.

The challenges facing “blurowl” video sequences are “Motion Blur” and “Fast Motion”. In the tracking result shown in Figure 17, at 128^{th} frame and 165^{th} frame, both the two tracking algorithms track the target well. At 255^{th} frame, due to the rapid movement of the camera, the target appears to motion blur. At this time, the KCF algorithm still selects the position with the maximum response value as the best matching position of the target, resulting in tracking failure. Due to the failure of the previous tracking, it will continue to affect the accurate tracking of the subsequent 307^{th} frame. However, the KCF_based_Fuzzy algorithm uses fuzzy detection in the early stage to make accurate judgments in advance, and make timely corrections. Therefore, the proposed algorithm accurately tracks the target in subsequent frames. In summary, for the challenges Motion Blur and Fast Motion, the KCF_based_Fuzzy algorithm avoids the template pollution during the challenges, and provides the better tracking than the traditional algorithms.

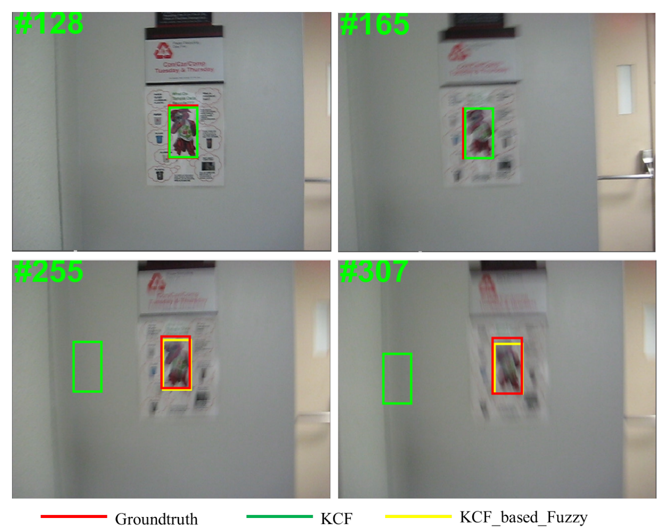


Fig.17. Comparison of tracking results between KCF_based_Fuzzy and traditional KCF algorithm on “blurowl” sequence

The challenges faced by the “box” video sequence include 5 challenges, which are Illumination Variation, Occlusion, Motion Blur, Out-of-View, and Low Resolution. It is clearly seen from Figure 18 that both the proposed KCF_based_Fuzzy algorithm and the original KCF algorithm have good tracking performance at 202^{th} frame. Then, at 487^{th} frame, when one of the Illumination Variation, Occlusion, Motion Blur, Out-of-View, and Low Resolution challenges occurs in the surrounding environment of the video, the KCF algorithm

1063-6706 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Carleton University. Downloaded on July 28,2020 at 01:17:33 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2020.3006520, IEEE Transactions on Fuzzy Systems

tracks a wrong target. However, without pre-judge ability, KCF still selects the maximum response value as the best matching position of the target, which leads to the failure of tracking. On the contract, the proposed KCF_based_Fuzzy algorithm makes a pre-judgment for timely corrections, which denotes that it accurately tracks the target in the subsequent 620^{th} and 1084^{th} frames.

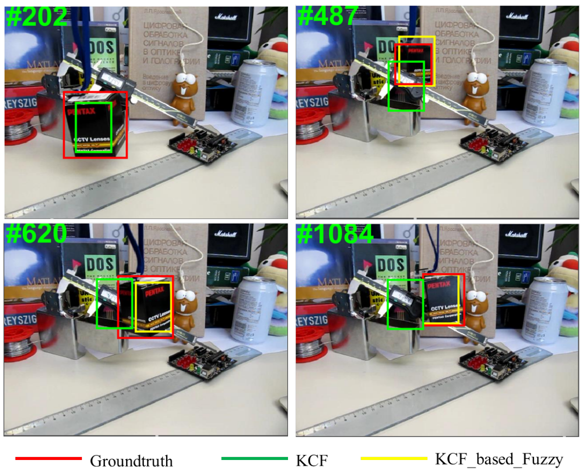


Fig.18. Comparison of tracking results between KCF_based_Fuzzy and traditional KCF algorithm on "box" sequence

The challenges of the "twinnings" sequence are Fast motion and Out-of-Plane Rotation. Figure 19 shows the comparison of the tracking between the two algorithms. When the target does not have Fast motion and Out-of-Plane Rotation problems at 400^{th} frame, both the two algorithms track the target accurately. However, at 425^{th} frame, when the target is under challenges Fast motion and Out-of-Plane Rotation, the fuzzy detection system pre-judges that the track is not so accurately at this time, and corrects the following tracking timely. Since the KCF algorithm cannot judge the tracking quality in advance, therefore, it fails to track accurately in the subsequent 440^{th} frame and 470^{th} frame. This also validates that the proposed KCF_based_Fuzzy algorithm shows better results than the traditional KCF algorithms.

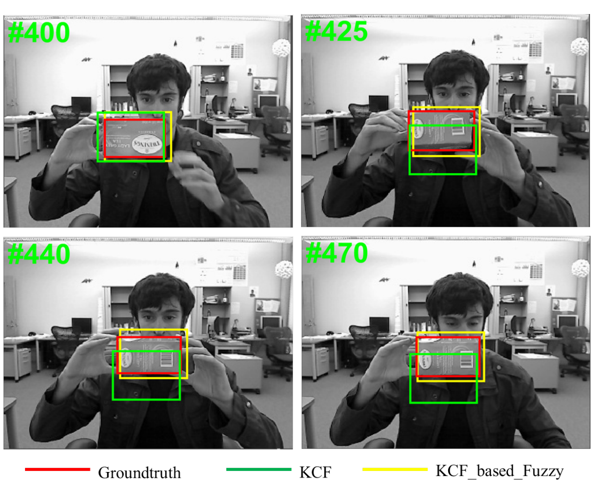


Fig.19. Comparison of tracking results between KCF_based_Fuzzy and traditional KCF algorithm on "twinnings" sequence

The challenges of the "surfer" sequences include Scale Variation, Fast Motion, In-Plane Rotation, Out-of-Plane Rotation, and Low Resolution. In Figure 20, when the target is under one of challenges Scale Variation, Fast Motion, In-Plane Rotation, Out-of-Plane Rotation, or Low Resolution in the 138^{th} frame, the KCF algorithm cannot track accurately in advance. In the subsequent 150^{th} and 157^{th} frames, because the surrounding environment of the target has been greatly affected, the tracking frame of the KCF algorithm is moving away from the target. However, the proposed KCF_based_Fuzzy algorithm uses the pre-judgment at 138^{th} frame, and accurately tracks the target in subsequent frames.

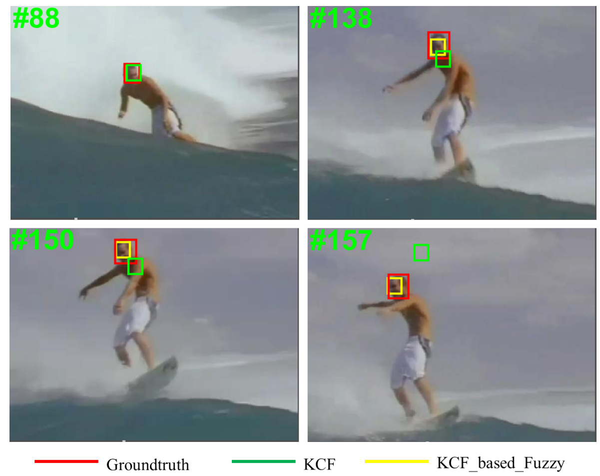


Fig.20. Comparison of tracking results between KCF_based_Fuzzy and traditional KCF algorithm on "surfer" sequence

V. CONCLUSION

Although the traditional CF based tracking algorithm is widely used in current target tracking algorithm, when the surrounding environment is under challenges, such as Out-of-View, Low Resolution, or Blurred Motion, the algorithm cannot achieve an accurately tracking. In order to solve the above problems, the paper constructed a fuzzy aided detection system and combined it into CF based algorithms. The proposed fuzzy aided system detected whether the target tracking result was accurate from the response matrix in advance, and determined whether the current frame used a visual memory mechanism. When the fuzzy system detected that the current frame was not accurately tracked, it relocated the target according to the stored target information in memory to guide following tracking. Experimental results showed that the proposed method had relatively good adaptability and robustness. By the comparisons of tracking performance between the traditional CF based algorithms, the tracking performance is significantly validated the effectiveness of the proposed strategy when dealing with complex visual challenges, such as Illumination Variation, Background Clutter, etc. Therefore, our future work will focus on the further improvement of these challenges to achieve more accurately tracking.

REFERENCES

[1] Z. Pan, S. Liu, and W. Fu, "A review of visual moving target tracking," *Multimedia Tools and Applications*, vol. 76, no. 16, pp. 16 989–17 018, 2017.

[2] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proceedings of the IEEE conference*

1063-6706 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Carleton University. Downloaded on July 28,2020 at 01:17:33 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2020.3006520, IEEE Transactions on Fuzzy Systems

on computer vision and pattern recognition, 2013, pp. 2411–2418.

[3] K. T. Hong, J. H. Shim, and Y. I. Cho, "Moving objects tracking method using spatial projection in intelligent video traffic surveillance system," *Journal of Korean Institute of Intelligent Systems*, vol. 25, no. 1, pp. 35–41, 2015.

[4] N. Kumar, J.-H. Lee, and J. J. Rodrigues, "Intelligent mobile video surveillance system as a bayesian coalition game in vehicular sensor networks: Learning automata approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 3, pp. 1148–1161, 2015.

[5] R. P. Loce, E. A. Bernal, W. Wu, and R. Bala, "Computer vision in roadway transportation systems: a survey," *Journal of Electronic Imaging*, vol. 22, no. 4, pp. 041 121–041 121, 2013.

[6] M. J. Black and A. D. Jepson, "Eigentracking: Robust matching and tracking of articulated objects using a view-based representation," *International Journal of Computer Vision*, vol. 26, no. 1, pp. 63–84, 1998.

[7] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," in *European conference on computer vision*, vol. 5302. Springer, 2008, pp. 234–247.

[8] X. Mei and H. Ling, "Robust visual tracking using l1 minimization," in *2009 IEEE 12th International conference on computer vision*. IEEE, 2009, pp. 1436–1443.

[9] ——, "Robust visual tracking and vehicle classification via sparse representation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 11, pp. 2259–2272, 2011.

[10] N. Wang and D.-Y. Yeung, "Learning a deep compact image representation for visual tracking," in *Advances in neural information processing systems*, 2013, pp. 809–817.

[11] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *2010 IEEE computer society conference on computer vision and pattern recognition*. IEEE, 2010, pp. 2544–2550.

[12] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *European conference on computer vision*. Springer, 2012, pp. 702–715.

[13] ——, "High-speed tracking with kernelized correlation filters," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 3, pp. 583–595, 2015.

[14] M. Danelljan, F. Shahbaz Khan, M. Felsberg, and J. Van de Weijer, "Adaptive color attributes for real-time visual tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1090–1097.

[15] J. Van De Weijer, C. Schmid, J. Verbeek, and D. Larlus, "Learning color names for real-world applications," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1512–1523, 2009.

[16] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *European conference on computer vision*. Springer, 2014, pp. 254–265.

[17] T. Liu, G. Wang, and Q. Yang, "Real-time part-based visual tracking via adaptive correlation filters," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4902–4912.

[18] Z. Hong, Z. Chen, C. Wang, X. Mei, D. Prokhorov, and D. Tao, "Multi-store tracker (muster): A cognitive psychology inspired approach to object tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4310–4318.

[19] J. Alcala-Fdez and J. M. Alonso, "A survey of fuzzy systems software: Taxonomy, current research trends, and prospects," *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 1, pp. 40–56, 2016.

[20] H. Han, "Theoretical analysis for a design of delay-dependent fuzzy control system with input saturation," in *2006 IEEE International Conference on Fuzzy Systems*. IEEE, 2006, pp. 2190–2194.

[21] R. Munoz-Salinas, E. Aguirre, O. Cordón, and M. Garcia-Silvente, "Automatic tuning of a fuzzy visual system using evolutionary algorithms: single-objective versus multiobjective approaches," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 2, pp. 485–501, 2008.

[22] C. H. Jong, K. M. Tay, and C. P. Lim, "A new fuzzy ratio and its application to the single input rule modules connected fuzzy inference system," in *2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2014, pp. 1586–1590.

[23] Li Xiaoshen, Yuan Xuehai, Jiang Mingzuo, and Zhang Chunling, "Fuzzy inference modeling method of time-varying system based on t-s fuzzy system," in *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2015, pp. 335–339.

[24] P. Hajek and V. Olej, "Interval-valued intuitionistic fuzzy inference system for supporting corporate financial decisions," in *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2018, pp. 1–7.

[25] A. Chibani, M. Chadli, P. Shi, and N. B. Braiek, "Fuzzy fault detection filter design for t–s fuzzy systems in the finite-frequency domain," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 5, pp. 1051–1061, 2017.

[26] S. Varija Raghu, D. F. Reiff, and A. Borst, "Neurons with cholinergic phenotype in the visual system of drosophila," *Journal of Comparative Neurology*, vol. 519, no. 1, pp. 162–176, 2011.

[27] Y. Lin, Y. Tu, and Z. Dou, "An improved neural network pruning technology for automatic modulation classification in edge devices," *IEEE Transactions on Vehicular Technology*, vol. 99, pp. 1–1, 2020.

[28] S. Huang and A. W. Sadek, "A novel forecasting approach inspired by human memory: The example of short-term traffic volume forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 5, pp. 510–525, 2009.

[29] Y. Lin, M. Wang, X. Zhou, G. Ding, and S. Mao, "Dynamic spectrum interaction of uav flight formation communication with priority: A deep reinforcement learning

1063-6706 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Carleton University. Downloaded on July 28,2020 at 01:17:33 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2020.3006520, IEEE Transactions on Fuzzy Systems

approach," *IEEE Transactions on Cognitive Communications and Networking*, pp. 1–1, 2020.

[30] I. N. Yulita, M. I. Fanany, and A. M. Arymurthy, "Sleep stage classification using convolutional neural networks and bidirectional long short-term memory," in *2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. IEEE, 2017, pp. 303–308.

[31] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Discriminative scale space tracking," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 8, pp. 1561–1575, 2016.

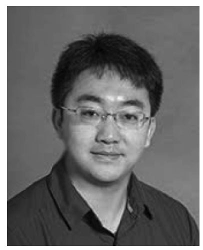
[32] S. Liu, C. Guo, F. Al-Turjman, K. Muhammad, and V. H. C. de Albuquerque, "Reliability of response region: A novel mechanism in visual tracking by edge computing for iiot environments," *Mechanical Systems and Signal Processing*, vol. 138, p. 106537, 2020.

Shuai Liu Dr. Shuai Liu received his BS, MS and PhD from Jilin University in 2002, 2004, and 2011. Now he acts as a full professor in College of Information Science and Technology, and Vice director of Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University. He is now a senior member of China Computer Federation. His research area contains fractals analysis, big data processing, computer vision, and information processing. He serves as an editor for many journals including Fractals, Mobile Network and Applications, Multimedia Tools and Applications, IEEE Access, etc.

Shuai Wang Mr. Shuai Wang is now act as a master under postgraduate in College of Information Science and Technology, Hunan Normal University.. His research interests include computer vision, target tracking, artificial intelligence application.

Xinyui Liu Miss. Xinyu Liu is now act as a master under postgraduate in College of Information Science and Technology, Hunan Normal University. Her research interests include computer vision, target tracking, artificial intelligence application.

Chin-Teng Lin Dr. Chin-Teng Lin (S'88–M'91–SM'99–F'05) received the M.S. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1989 and 1992, respectively. He is currently a Distinguished Professor with the Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW, Australia, He has authored over 200 journal papers in the areas of fuzzy systems and neural networks. Dr. Lin was elevated to be an International Fuzzy Systems Association Fellow in 2012.



Zhihan Lv Dr. Zhihan Lv (M'16, SM'19) received the Ph.D. degree from Paris Diderot University. He is currently an Associate Professor with Qingdao University. His research interests include Internet of Things, multimedia, augmented reality, virtual reality, 3-D visualization and graphics. He was a Marie Curie Fellow with the European Union's Seventh Framework Programme LANPERCEPT. Dr. Lv has been an Associate Editor of the PLOS One (since 2016), the IEEE Access (since 2016), the Neurocomputing (since 2016), and the IET Image Processing (since 2017).

1063-6706 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: Carleton University. Downloaded on July 28,2020 at 01:17:33 UTC from IEEE Xplore. Restrictions apply.