Human Memory Update Strategy: A Multi-Layer Template Update Mechanism for Remote Visual Monitoring

Shuai Liu, Shuai Wang, Xinyu Liu, Amir H. Gandomi, *Senior Member, IEEE*, Mahmoud Daneshmand, *Senior Life Member, IEEE*, Khan Muhammad, *Member, IEEE*, and Victor Hugo C. De Albuquerque, *Senior Member, IEEE*

***Abstract*—In the era of rapid development of artificial intelligence, the integration of multimedia and human-artificial intelligence has become an important research hotspot. Especially in the multimedia environment, effective remote visual monitoring has become the exploration direction of many scholars. The use of traditional correlation filtering (CF) algorithm for real-time monitoring in the context of multimedia is a practical strategy. However, most existing filtering-based visual monitoring algorithms still have the problem of insufficient robustness and effectiveness. Therefore, by considering the strategy of updating human memory, this paper proposes a multi-layer template update mechanism to achieve effective monitoring in a multimedia environment. In this strategy, the weighted template of the high-confidence matching memory is used as the confidence memory, and the unweighted template of the low-confidence matching memory is used as the cognitive memory. Through the alternate use of confidence memory, matching memory, and cognitive memory, it is ensured that the target will not be lost during the monitoring process. Experimental results show that this strategy does not affect the speed (still real-time) and improves the robustness in the multimedia background.**

***Index Terms*—Multimedia environment, visual monitoring, filtering algorithm, human memory, template update.**

Manuscript received October 25, 2020; revised January 14, 2021; accepted February 4, 2021. Date of publication March 12, 2021; date of current version July 30, 2021. 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, Innovation and Entrepreneurship Training Program of Hunan Xiangjiang Artificial Intelligence Academy, Educational Reform Project of Hunan Xiangjiang Artificial Intelligence Academy, Open Project Program of the State Key Laboratory of CAD&CG under Grant A1926, Zhejiang University. The Associate Editor for this article was Prof. Huansheng Ning. *(Corresponding author: Khan Muhammad.)*

Shuai Liu, Shuai Wang, and Xinyu Liu are with the Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, College of Information Science and Engineering, as well as Hunan Xiangjiang Artificial Intelligence Academy, Hunan Normal University, Changsha 410000, China (e-mail: liushuai@hunnu.edu.cn; askshuai@gmail.com; askliu1995@gmail.com).

Amir H. Gandomi is with the Faculty of Engineering & Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia (e-mail: gandomi@uts.edu.au).

Mahmoud Daneshmand is with the Stevens Institute of Technology, Hoboken 07030, NJ USA (e-mail: mdaneshm@stevens.edu).

Khan Muhammad is with Visual Analytics for Knowledge Laboratory (VIS2KNOW Lab), Department of Software, Sejong University, Seoul 143-747, Republic of Korea (e-mail: khan.muhammad@ieee.org).

Victor Hugo C. De Albuquerque is with the Graduate Program on Teleinformatics Engineering, Federal University of Ceará, Fortaleza, Fortaleza/CE, Brazil; and the Graduate Program on Telecommunication Engineering, Federal Institute of Education, Science and Technology of Ceará, Fortaleza/CE, Brazil (e-mail: victor.albuquerque@ieee.org).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TMM.2021.3065580.

Digital Object Identifier 10.1109/TMM.2021.3065580

I. Introduction

In THE period of rapid development of artificial intelligence (AI) [1], the real-time fusion problem of many fields and AI has become more and more complicated. As one of the important fields of AI, computer vision is inevitably integrated into various other departments or fields. Computer vision is a discipline that studies how to make computers "see" like humans. In detail, it uses cameras and computers to replace the human eye, enabling the computer to achieve the functions of segmenting, classifying, identifying, tracking, and discriminating decisions like the human visual system [2], [3]. On the basis of image and signal processing technology, probability statistical analysis, computational geometry, and neural network technology, the computer and its related equipment carry out biological vision simulation [4], with the purpose of enabling it to recognize three-dimensional environmental information through two-dimensional images.

Target monitoring is one of the most important research fields of computer vision. It has wide application prospects in military reconnaissance, 3-D transmission, fire scene analysis, battlefield assessment, and security monitoring [5]–[7]. It has attracted many scholars to carry out research in this domain. For instance, Hare *et al.* [8] transformed the monitoring problem into a classification problem and proposed an adaptive visual target monitoring framework on the basis of structural output prediction. The intermediate classification link is avoided by explicitly introducing an output space to meet the monitoring function, and the monitoring results are directly generated. Henriques *et al.* [9] firstly used the cyclic matrix to collect positive and negative samples, then trained the target detector through ridge regression, and successfully used the diagonalization property of the cyclic matrix in Fourier space to convert the matrix operation into a vector product. The method greatly reduced the amount of calculation and improved the calculation speed. Danelljan *et al.* [10] regarded target monitoring as two independent problems of target center translation and target scale change. First, they used

DCF of HOG feature to train translation correlation filtering for detecting target center translation, and then used HOG feature MOSSE algorithm [11] to train another scale-dependent filter to detect changes in target scale. Bertinetto *et al.* [12] proposed a monitoring method that combines sensitive representations of two complementary factors. For template and pixel learning, an inherently robust model for color change and distortion was established, and the monitoring algorithm was improved. Lukezic *et al.* [13] first constructed the mask matrix using the color model of the foreground and background, then spatially restricted the filter, and constructed the weighting coefficients of different channels based on the response map information of different channels, which improved the accuracy of the video in complex color transformation rate. Wang *et al.* [14] proposed the concept of multi-peak detection, which improved the accuracy of monitoring by excluding disturbing similar backgrounds. Yang *et al.* [15] proposed a monitoring algorithm based on gray-level local binary pattern descriptors. This descriptor can not only describe local texture features, but also describe the distribution of brightness. Their algorithm can improve the robustness in the background of low-score variability. Wahlstrom *et al.* [16] used a Gaussian process to track extended objects or groups of objects. During each scan, the object generates multiple measurements to estimate the shape and kinematics of the object and learn the shape online through the Gaussian process. Li *et al.* [17] presented a new monitoring approach based on correlation filter, having a novel robust estimation of similarity transformation for large displacements. Liu *et al.* [18] proposed a new tracking method based on energy correlation filter (EACOFT), which sets the energy between foreground and background to be adaptively balanced, ensuring that the target always has higher energy than its background.

In recent years, although many countermeasures have been put forward for target monitoring, there are still many deficiencies in the existing monitoring algorithms in terms of the complex environmental characteristics under the current multimedia background, such as Illumination Variation, Occlusion, Fast Motion and so on. In order to better solve the above problems, this paper introduces a multi-layer template update strategy into the CF-based monitoring algorithm by considering the human memory update strategy to achieve effective monitoring in the multimedia environment. The main contributions of this article are summarized as follows:

1) A new visual memory update strategy with a multi-layer is proposed for effective monitoring in a multimedia environment. In this strategy, three memories are implemented: confident memory, matching memory, and cognitive memory. First, confident memory is cyclically used to store the weighted template of the high-confidence matching memory. Then, if the matching memory is low-confident, a real-time cognitive memory is used to store the unweighted template of the matching template. If the real-time cognitive memory is reliable, both the confident and matching memory are updated; otherwise, only the matching memory is updated. By using this new strategy, the monitoring speed is not greatly affected (still real time), and the target can be monitored more accurately.
2) The strategy of templates selection alternately is put forward. If the current frame is well tracked, the matching template is used for monitoring; otherwise, the template with higher reliability in confident and cognitive memories is selected for monitoring. In this way, since the template is used alternately in the visual monitoring process, the target is not lost if it cannot be tracked in a few frames.

The main structure of this paper is as follows: Section II explains the related work, including the visual memory mechanism, template matching update strategy, and the filter monitoring algorithm. Section III proposes a correlation filtering algorithm based on the visual memory multi-template update strategy. Section IV evaluates and compares the improved algorithm that integrates the multi-template update strategy of visual memory with the original algorithm. Finally, this work is summarized and the outlook for the future is presented in Section V.

II. Related Work

A. Visual Memory Mechanism

Visual memory is the process by which the human brain encodes, stores, and extracts information input from the external environment. The reason why human can remember things of interest is that the brain stores the information ingested by the eyeballs into neurons. After a long period of exploration, it was found that after the external information stimulates the human visual organs, the visual traces first form and enter the instantaneous memory. Instantaneous memory contains a large amount of information, and the retention time is very short, so it disappears quickly. For the stimulus information related to cognitive activities in transient memory, if human vision pays only a little attention, this related stimulus information is re-encoded by the visual system into a short-term memory system with limited capacity and a longer retention time. But without repeating it, this kind of memory can only save information for 15 seconds to 1 minute. When humans continue to repeat, the important information processed by short-term memory is sent to long-term memory storage, and eventually form long-term memory. Long-term memory is a real information library, which has a considerable storage capacity and a high ability to maintain information. The information stored in the long-term memory storage is long-lasting and presentable. Usually, long-term memory can be kept for several hours, or even years. However, it will be transferred to short-term memory when it is disturbed or weakened.

In recent years, the visual memory mechanism has been widely used in the field of target monitoring by many researchers. For instance, Frintrop *et al.* [19] detected the salient features of the target by using the visual attention mechanism in human visual intelligence. Based on the theory of the human visual system, Xu *et al.* [20] proposed a dynamic model that combines feature learning and feature association. Although this model had a significant effect on the target monitoring problem in the defined state, it could not solve this problem in the undefined state well. Liu *et al.* [21] studied the relationship between target monitoring and visual attention change, and calculated the saliency of objects by establishing a correlation model. When choosing the attention object in the early stage, the attention was usually focused on the salient objects, and in the later stage, the attention was transferred to the target. Guo *et al.* [22] quantified and analyzed the movement process of the object according to the dynamic and static characteristics of the object movement, so as to achieve the purpose of monitoring the target. Cheng *et al.* [23] proposed collaborative real-time streaming and storage sharing through monitoring and learning of popular social networks. In order to cope with the change of the posture of the target during the monitoring process, Kang *et al.* [24] optimized the template update process by proposing an update idea of the target model based on short-term memory. Montemayor *et al.* [25] proposed a particle filtering monitoring algorithm based on memory by combining the idea of particle filtering and visual memory theory to solve the problem of occlusion of objects during monitoring.

B. Template Matching

With the widespread application of target monitoring technology in various fields [26]–[28], its deficiencies have gradually emerged. In the actual monitoring process, the scene is often more complex and changeable. For example, the target is blocked by the interference during the movement of the target, the scene illumination changes, and the target size changes, which may cause the target appearance model to change significantly. It disappears and reappears in the monitoring field of view. At this time, the actual target template and the initial target template are already quite different. However, traditional target monitoring algorithms often use static templates, which can easily cause target monitoring failure. Most experimental verifications indicate that once the monitoring fails, it is difficult to reposition the target.

In recent years, many scholars have conducted in-depth investigations on these issues. Among them, the method based on template strategy is particularly remarkable. For example, Li *et al.* [29] proposed a real-time target monitoring algorithm based on improved template matching to solve the long-term monitoring problem in video streams. By using random forest to generate and train features, the similarity was evaluated and the real-time template was updated according to the improved normalized correlation coefficient. Bal *et al.* [30] proposed to use the intensity change function and template modeling to automatically monitor target monitoring in forward-looking infrared (FLIR) image sequences. This method used some characteristics of target intensity for modeling to make the monitor determine the true coordinates of one or more targets and thus monitor effectively. Lamberti *et al.* [31] proposed a template matching strategy to improve the robustness of the FLIR image of the target monitoring algorithm. This method used motion prediction metrics to identify the occurrence of false alarms and control the activation based on the template matching phase. Jia *et al.* [32] proposed a Bayesian fusion template-based matching algorithm to improve target monitoring. This method used two different template matchings combined with the Bayesian theory for weighting the monitoring results. Eventually, the updating and matching of the template were realized. Kumar *et al.* [33] proposed a target monitoring based on the dominant direction template and Kalman filtering. By formulating a small target search box at the Kalman estimated position of the target frame, the DOT template matching was only effective in this window. Williamson *et al.* [34] used a hybrid template/optical flow method to develop a method for monitoring targets in ultrasound image data in real time. Ross *et al.* [35] adopted the method of incremental principal component analysis to achieve the dynamic update of the target template. The proposed algorithm has certain robustness in solving changes in illumination and changes in size.

C. Correlation Filter Monitoring Algorithm

The correlation filtering algorithm is mainly divided into five parts: solving the matrix using least squares, solving using ridge regression, diagonalization of cyclic matrix in Fourier space, nonlinear regression filter solution, and quick test.

1) Solving the Matrix Using Least Squares: Set the training sample generated using cyclic shift be X. The unary linear representation is:

$$f\ (x_i) = w^T\ \vec{x}_i + b \tag{1}$$

Here w is the weight coefficient and b is the bias.

2) Solving Using Ridge Regression: Ridge regression is to add a regular term λI to $X^T X$, so that the matrix is non-singular, and then the formula "$X^T X + \lambda I$" can be inverted as shown in (2):

$$\tilde{w} = \left(X^T X + \lambda I\right)^{-1} X^T \vec{y} \tag{2}$$

3) Diagonalization of Cyclic Matrix in Fourier Space: For any cyclic matrix, the Fourier transform matrix can be diagonalized, as shown in the following (3):

$$X\ =\ C\ (x) =\ Fdiag\,(\hat{x})\,F^H \tag{3}$$

4) Nonlinear Regression Filter Solution: For solving the nonlinear regression filter, the first step is to find a nonlinear mapping $\varphi(x)$ to make the mapped samples linearly separable in the new space. Then, use ridge regression to find a classifier $f\ (x_i) = \omega^T\ \varphi(x_i)$ in the new space. Here $\varphi(x_i)$ represents the transformation of the sample x_i by a nonlinear mapping function φ. The solution ω of the linear filter is expressed by the linear combination of samples as shown in (4):

$$\omega\ = \sum_i \alpha_i\ \varphi\,(x_i) \tag{4}$$

5) Quick Test: In the process of detection, the response equation formed by the training samples and template samples of the filter algorithm in the high-dimensional space is expressed as given in (5):

$$\hat{f}\ (z) = \hat{k}^{tz} \odot \hat{\alpha} \tag{5}$$

Here k^{tz} represents the first row of the kernel matrix K, and $\hat{k}^{tz}$ represents the kernel correlation of the training sample t and the sample z to be tested in the Fourier domain.

First, the filter monitoring algorithm uses a cyclic matrix to cyclically sample the surrounding area of a given position in

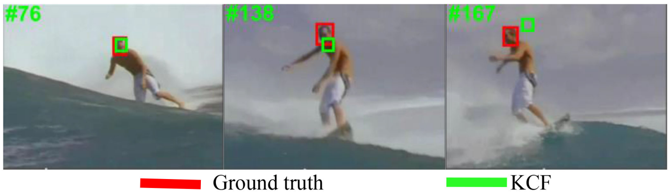


Fig. 1. Drift problem in target monitoring algorithm.

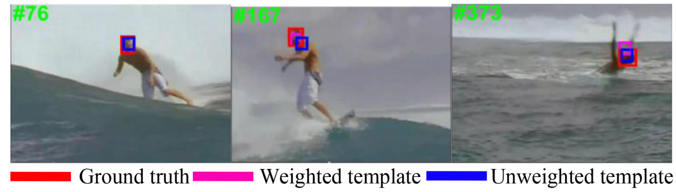


Fig. 2. The use of weighted and unweighted templates in the drift problem.

the previous frame to generate many samples. Then, it extracts the target feature information for each sample to generate a new filter. Finally, it calculates the response value of each sample to form a new response matrix by using the Fourier formula. Usually, the maximum response value is selected as the best matching position of the target.

III. Our Proposed Strategy

A. *Deficiencies of Existing Filtering Algorithms*

The traditional filtering algorithm achieves the update process of the target template by linearly combining the template information obtained from the previous frame and the template information of the currently trained filter. Then, the updated template is used as the target template for the current frame. For example, the template update equation of the KCF algorithm in the i frame is as follows:

$$\phi_i = (1 - \partial)\ \phi_{i-1} + \partial new\_\phi_i \tag{6}$$

$$\eta_i = (1 - \partial)\ \eta_{i-1} + \partial new\_\eta_i \tag{7}$$

Here ∂ is the adaptive rate of the monitor (also represents the memory factor of the updated template); ϕ_i and ϕ_{i-1} represent the best matching template of the current frame and the previous frame; new_ϕ_i is the target template of the current frame, new_η_i is the weight coefficient of the input training samples of the current frame; and η_i and η_{i-1} are the weight coefficients of the corresponding filters of the current frame and the previous frame.

Although the traditional correlation filtering algorithm uses a memory factor to update the template by saving the information of the previous frame and the current frame template, when the target encounters problems such as occlusion or drift, the template may cause pollution. If the templates in subsequent frames continue to be updated in this manner, the target positioning may fail. As shown in Fig. 1, the existing filtering algorithm can monitor the target well at 76th frame. However, when the target drifts at the 138th frame, the algorithm cannot fully extract the feature information of the target at this time, causing the template to be polluted, which affects the subsequent monitoring effect (monitoring failure at the 167th frame).

From the update process of the KCF algorithm, it can be known that the best matching template ϕ_{i-1} in $(i-1)^{th}$ frame and the target template new_ϕ_i in the current i^{th} frame can also be called weighted templates and unweighted templates. The updated best matching template ϕ_i is called the original template, which is very important for the prediction of the target at i^{th} frame.

In the case where the target is blocked or drifting, the use of weighted templates or unweighted templates to predict the best matching position of the target in the current frame is likely to cause storage errors, because the target model may change greatly at this time. As shown in Fig. 2, the traditional algorithm using both weighted and unweighted templates can monitor the target very accurately at 76th frame. However, the traditional algorithm cannot fully extract its feature information at this time due to the large-scale motion of the target at the 167th frame, resulting in the target template being polluted (as shown in the figure). In other words, the unweighted template has a deviation in predicting the target position. However, the weighted template at this time can accurately predict. It can be seen from the 373th frame that the monitoring algorithm using the weighted template can no longer accurately monitor, and the monitoring strategy using the unweighted template can also predict the location of the target very well.

B. *Visual Memory Process Establishment*

After humans ingest the moving object of interest through the eyeball, the visual system extracts the characteristic information of the object and stores it in the brain neurons, thus forming a short-term memory. When an object is in motion, the human visual system picks up new characteristic information according to the state of the object at every moment. Then the human visual system fuses and stores the feature information newly acquired and stored in the brain neurons, so as to perfect the information features in the target information storage library and realize accurate monitoring of the target. However, when moving objects are affected by factors such as occlusion or scale changes, the human visual system has already sensed the changes in the external environment of the object. Therefore, the human eyeballs cannot accurately capture the information characteristics of the current target. At this time, the human brain analyzes the information stored by the neurons and determines the location of the target at this time. When the target reappears, the target's appearance model may have changed significantly. It is not enough to rely on the target feature stored in the human brain to recognize the target. It is also necessary to re-remember the target's feature information at this time and improve the feature database. The process of human memory can be represented as Fig. 3.

According to the memory location process of human monitoring targets, this article divides human visual memory into three main stages:

1) Matching memory stage: First, the information features stored in human brain neurons are merged with the new information features currently picked up by the human

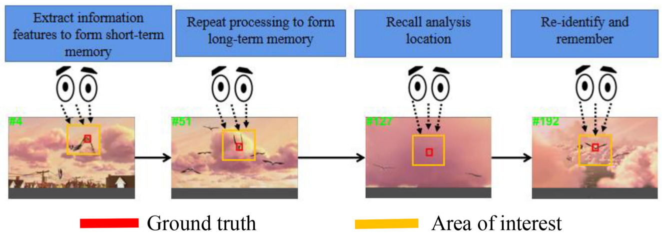


Fig. 3. Human memory process diagram.

eye to form a feature information database. Then the current target state is matched with the feature information database to judge the current monitoring effect.

2) Confidence memory stage: If the characteristics of the currently matched target are consistent with the characteristics of the moving target, it indicates that the target can be accurately monitored. At this time, the target feature information stored in the brain neurons is converted into confidence memory.
3) Cognitive memory stage: After drifting or occlusion, the appearance characteristics of the target may have changed significantly. At this time, storing target information features in the confidence memory is no longer applicable, so it is necessary to re-extract the current target apparent feature information as cognitive memory to improve the target feature information database.

C. Features of Multi-Layer Visual Memory Template

According to the above analysis, the use of weighted or unweighted templates alone does not provide good monitoring. Therefore, it is necessary to use the three stages of visual memory model, combining a weighted template and an unweighted template to timely repair the original template pollution problem.

1) When the target is moving normally for the ϑ^{th} frame, the motion state of the target is not affected by the surrounding environment. The monitor first extracts the target information features captured by the current camera. Then it uses the new target information features to improve the target information repository. Finally, the target monitor matches the status of the current target according to the perfect target information repository (original template ϕ_ϑ). If the current matching memory is of high confidence, the confidence memory is used to store the weighted template of the original template $\phi_{\vartheta-1}$ and the corresponding weight coefficient $\eta_{\vartheta-1}$ as the confidence memory.

$$\begin{cases} template_c = \phi_{\vartheta-1} \\ \sigma \ = \eta_{\vartheta-1} \end{cases} \tag{8}$$

Here $\phi_{\vartheta-1}$ is the original template information of the $(\vartheta-1)^{th}$ frame (weighted template) and $\eta_{\vartheta-1}$ is the weight coefficient of the filter in the $(\vartheta-1)^{th}$ frame.

2) If the target is blocked by the surrounding environment or changes in lighting at the κ^{th} frame, the current camera cannot capture the movement information of the target.

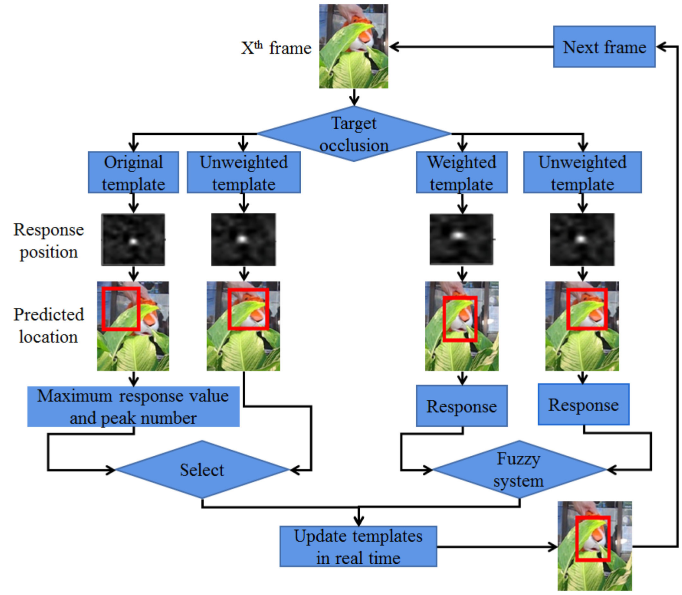


Fig. 4. Process diagram of multi-layer visual memory template for target monitoring.

In this case, the target monitor cannot improve the target information storage library, and the original template ϕ_κ stored in the current matching memory has low confidence. In this case, the monitor needs to ensure that the confidence memory is not being updated at this time. After that, the target information feature in the confidence memory is used to analyze and infer the target location. At the same time, cognitive storage is used to store the unweighted template of the original template new_ϕ_κ and the corresponding weight coefficient new_η_κ as cognitive storage.

$$\begin{cases} template_d = new\_\phi_\kappa \\ v \ = new\_\eta_\kappa \end{cases} \tag{9}$$

Here new_ϕ_κ is the target template information of the κ^{th} frame (unweighted template) and new_η_κ is the weight coefficient of the input training samples in the κ^{th} frame.

3) When the target resumes normal movement, although the current camera can capture the target motion information at this time, the appearance model of the current target may have changed significantly. Therefore, the matching memory at this time is of low confidence. The target monitor will only select a template with higher reliability $template_A$ from the weighted template $template_c$ in the confidence memory storage library and the unweighted template $template_d$ in the cognitive memory storage library for monitoring.

$$template_A = f\left(template_c, template_d\right) \tag{10}$$

Among them, f represents the degree of reliability. The flow chart for the multi-layer visual memory template is shown in Fig. 4.

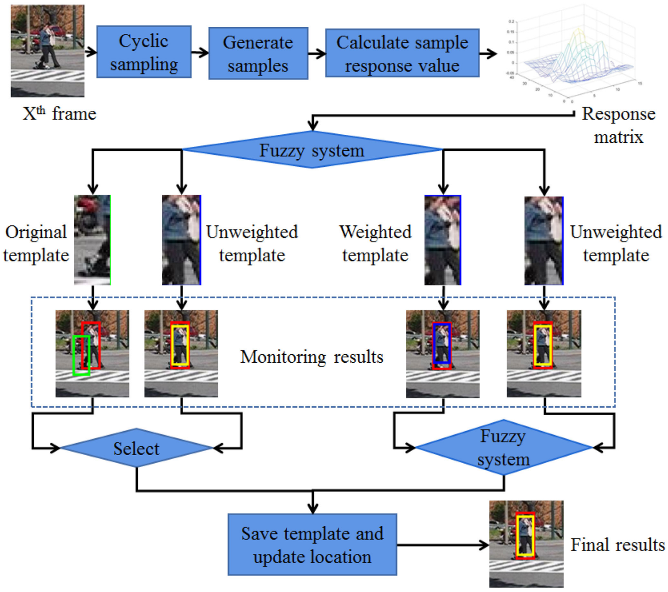


Fig. 5. Flow chart of filtering algorithm based on visual memory multi-template.

D. *Filtering Algorithm Based on Visual Memory Multi-Template*

For the deficiency of the existing filtering algorithm, it is necessary to use the visual memory multi-template update strategy. In this strategy, three memories are implemented: confidence memory, matching memory, and cognitive memory. First, the confidence of the template is judged by the fuzzy auxiliary system [36], and the confidence memory is used regularly to store the weighted template of the high-confidence matching memory. If the matching memory is of low confidence, real-time cognitive memory is used to store the unweighted template of the matching template. Then, the real-time cognitive memory is judged. If the current real-time cognitive memory is reliable, the confidence memory and the matching memory are updated at the same time; otherwise, only the matching memory is updated. If the target monitoring effect of the current frame is good, the original template of the matching memory storage is used for monitoring; otherwise, the more reliable weighted template of the confidence memory storage and the unweighted template of the cognitive memory storage is selected for monitoring. In this way, in the multimedia background, due to the alternating use of templates in the process of monitoring, the accuracy of target monitoring can be well guaranteed. The flow chart of filtering algorithm based on visual memory multi-template is shown in Fig. 5.

According to the description presented so far, the proposed visual memory multi-template is applied to the relevant filter monitoring algorithm. The specific process with necessary details of the proposed algorithm is given in Algorithm 1.

Algorithm 1: Filtering Algorithm Flow Based On Visual Memory Multi-Template.

Input: Target initial position (X_1, Y_1)
Output: The best match position (X_i, Y_i) predicted for each frame and the target monitoring frame $Targetbox$

```
Repeat
  Detect:
    Step 1: Generate a large number of samples at the target position (X_{i-1}, Y_{i-1}) area of the previous frame using cyclic shift, then use the calculator to find the response value of each sample separately to form the response matrix response_i of the original template.
    Step 2: Use the trained blur detection system to judge the reliability t of the response matrix response_i of the current frame.
    If t < 0 then
    Extract the weighted template information in the confidence memory repository and unweighted template information in the cognitive repository.
    Use the fuzzy system to judge the reliability t_1 of the response matrix of the weighted template and the reliability t_2 of the response matrix of the unweighted template.
    Calculate the maximum response value F'_max of the weighted template.
      If (t_1 > t_2 &amp;; &amp;; t_2 > 0)||F'_max ⟩ λ then
        Use the position of the maximum response value F'_max of the weighted template template_c as the best matching position (X_i, Y_i) of the current target and output the target monitoring frame Targetbox.
      Else
        Use the position of the maximum response value F''_max of the unweighted template template_d as the best matching position (X_i, Y_i) of the current target and output the target monitoring frame Targetbox.
    End
  Else
    Calculate the maximum response value F_max and peak number Num of the current frame response matrix response_i.
    If F_max > δ || Num < ∂ then
      Use the position of the maximum response value F_max of the original template φ_i of the current frame as the best matching position (X_i, Y_i) of the current target and output the target monitoring frame Targetbox.
      Else
      Extract unweighted template information from the cognitive repository.
      Use the position of the maximum response value F''_max of the unweighted template template_d as the best matching position (X_i, Y_i) of the current target and output the target monitoring frame Targetbox.
      End
    End
  Train: Update matching memory, confidence memory, and cognitive memory in real time.
  Until: End of video.
```

Note: λ, δ, and ∂ are the set threshold.

IV. Experimental Results and Analysis

A. Evaluation Criteria

As more and more scholars research target monitoring, many excellent algorithms have been proposed. In order to evaluate the accuracy of these algorithms in a more fair and objective manner, some benchmarks and data sets for evaluation algorithms have been proposed. For example OTB [37] data set, VOT [38] data set, LASOT [39] data set, among which OTB data set and VOT data set are the most widely used data sets. This experiment uses the OTB data set to evaluate the performance of our algorithm. The OTB data set includes OTB500, also known as OTB2013, which was proposed in 2013. It was later extended to OTB100, also known as OTB2015. The benchmarks for evaluation of the OTB platform are Precision and Success rate. The evaluation methods include one-time pass assessment (OPE), temporal robustness assessment (TRE), and spatial robustness assessment (SRE).

1) Accuracy (Precision): First calculate the Euclidean distance τ between the center position (a_i, b_i) of the target monitoring current frame and the manually marked true center position (a_t, b_t). The formula is given in (10).

$$\tau = \sqrt{(a_i - a_t)^2 + (b_i - b_t)^2} \tag{10}$$

Then, calculate the ratio P of the number of frames n in video sequence, whose Euclidean distances are less than the threshold to the number of frames in the entire sequence. The specific formula is given in (11).

$$P = \frac{\sum_{i=1}^{n} M_i}{n}, \; M_i = \begin{cases} 1, \tau \leq 20 \\ 0, \tau > 20 \end{cases} \tag{11}$$

2) Success rate: First calculate the intersection and union of the target's current monitoring frame K_i and the manually marked target frame G_t, and then calculate the intersection and union ratio S. The specific equation is (12).

$$S = \frac{|K_i \cap G_t|}{|K_i \cup G_t|} \tag{12}$$

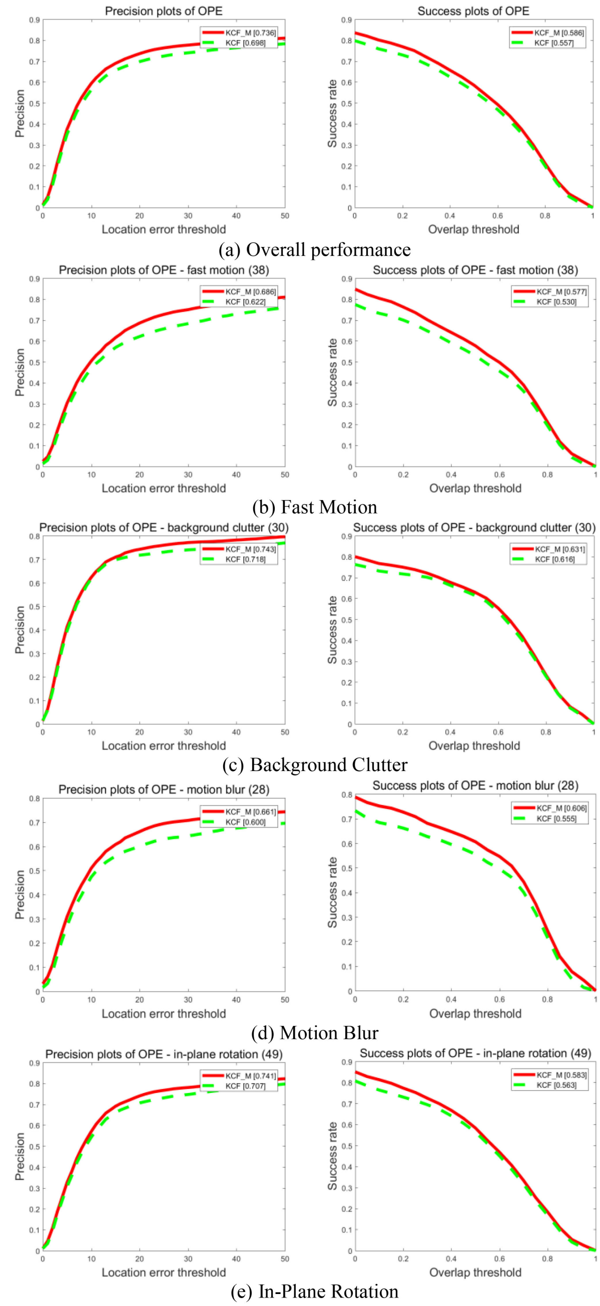


Fig. 6. Accuracy and success rate of the KCF and the KCF_M algorithm.

B. Quantitative Analysis

In this paper, a one-time pass evaluation (OPE) of the improved KCF algorithm combined with the visual memory multi-template (KCF_M) and the existing KCF algorithm is adopted. The 11 challenging attributes of the OTB platform 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). Accuracy (Precision) and Success Rate of the above attributes are shown in Fig. 6.

From the analysis of the overall performance of Fig. 6(a), the KCF_M algorithm based on the multi-template update of visual memory is significantly better than the original KCF algorithm. The accuracy of the improved KCF_M algorithm is 0.736, but the accuracy of the original KCF algorithm is 0.698, which is improved 5.44%. The success rate of the improved KCF_M algorithm is 0.586, but the success rate of the original KCF algorithm is 0.557, which is improved 5.21%. It can be seen that the multi-template update strategy based on visual memory plays a great role in the original algorithm.

From the perspective of Fast Motion in Fig. 6(b), the KCF_M algorithm based on multi-template updating of visual memory has a significant improvement in the fast movement properties compared to the original KCF algorithm. The accuracy rate of the improved KCF_M algorithm on Fast Motion is 0.686, which is 10.29% higher than the original KCF algorithm, and the accuracy rate of the original KCF algorithm is 0.622. The success rate of the improved KCF_M algorithm on Fast Motion is 0.577, and the success rate of the original KCF algorithm is 0.530, which is increased by 8.87%. The results show that the KCF_M algorithm based on the visual memory multi-template update strategy has a significant improvement in the Fast Motion attributes.

From the perspective of Background Clutter in Fig. 6(c), the KCF_M algorithm based on the multi-template update of visual memory has a significant improvement compared to the original KCF algorithm in the property of Background Clutter. The accuracy rate of the improved KCF_M algorithm is 0.743, the accuracy rate of the original KCF algorithm is 0.718, and the increase rate is 3.48%. The success rate of the improved KCF_M algorithm is 0.631, while the success rate of the original KCF algorithm is 0.616. The success rate of the improved KCF_M algorithm is increased by 2.44%. The results show that the KCF_M algorithm based on the visual memory multi-template update strategy has a significant improvement in the Background Clutter properties.

From the perspective of Motion Blur in Fig. 6(d), the KCF_M algorithm based on multi-template update of visual memory has a significant improvement compared to the original KCF algorithm in the attribute of Motion Blur. The accuracy of the improved KCF_M algorithm on Motion Blur is 0.661, the accuracy of the original KCF algorithm is 0.600, and the improvement rate is 10.17%. The success rate of the improved KCF_M algorithm on Motion Blur is 0.606, and the success rate of the original KCF algorithm 0.555, the increase rate is 9.19%. The results show that the KCF_M algorithm based on the visual memory multi-template update strategy has a significant improvement in the Motion Blur attribute.

From the perspective of the In-Plane Rotation in Fig. 6(e), the KCF_M algorithm based on visual memory multi-template update has a significant improvement in the In-Plane Rotation attribute compared to the original KCF algorithm. The improved KCF_M algorithm has an accuracy rate of 0.741 on the In-Plane Rotation, while the original algorithm KCF has an accuracy rate of 0.707. The accuracy rate of the improved KCF_M algorithm is increased by 4.81%. The improved KCF_M algorithm has a success rate of 0.583 on the In-Plane Rotation, while the original algorithm KCF has a success rate of 0.563. The accuracy rate of the improved KCF_M algorithm is increased by 3.55%. It is concluded that the KCF_M algorithm based on the visual memory multi-template update strategy has a significant improvement in the In-Plane Rotation attribute.

For other attributes on the OTB platform, the KCF_M algorithm based on visual memory multi-template update also has a very good improvement effect. The specific statistical results are shown in Table I and Table II.

To further prove the effectiveness of the mechanism, Fig. 7 is provided, showing the overall comparison between the FDSST_M and the KCF_M algorithms based on the visual memory template and other recent methods including FDSST_based_Fuzzy algorithm, FDSST algorithm [40], CSK algorithm, MRCT algorithm [41], and DSST algorithm.

It can be seen from Fig. 7 that the FDSST_M algorithm based on the visual memory template is the best. The accuracy of the FDSST_M algorithm is 0.742 higher than the 0.728 of the original FDSST algorithm. The success rate of the FDSST_M algorithm is 0.696 higher than the 0.679 of the original FDSST algorithm.

TABLE I
THE ACCURACY OF THE KCF ALGORITHM AND THE IMPROVED KCF_M ALGORITHM ON 11 CHALLENGE ATTRIBUTES OF OTB2015

	Algorithm		
Attributes	KCF	KCF_M	Increase ratio (%)
FM	0.622	0.686	10.29
BC	0.718	0.743	3.48
MB	0.600	0.661	10.17
DEF	0.621	0.649	4.51
OPR	0.678	0.714	5.31
OV	0.514	0.615	19.65
SV	0.637	0.679	6.59
LR	0.581	0.686	18.07
IPR	0.707	0.741	4.81
IV	0.728	0.735	0.96
OCC	0.632	0.660	4.43

TABLE II
SUCCESS RATE OF THE KCF ALGORITHM AND IMPROVED KCF_M ALGORITHM ON 11 CHALLENGE ATTRIBUTES OF OTB2015

	Algorithm		
Attributes	KCF	KCF_M	Increase ratio (%)
FM	0.530	0.577	8.87
BC	0.616	0.631	2.44
MB	0.555	0.606	9.19
DEF	0.511	0.533	4.31
OPR	0.529	0.558	5.48
OV	0.457	0.530	15.97
SV	0.419	0.449	7.16
LR	0.295	0.417	41.36
IPR	0.563	0.583	3.55
IV	0.554	0.567	2.35
OCC	0.515	0.540	4.85

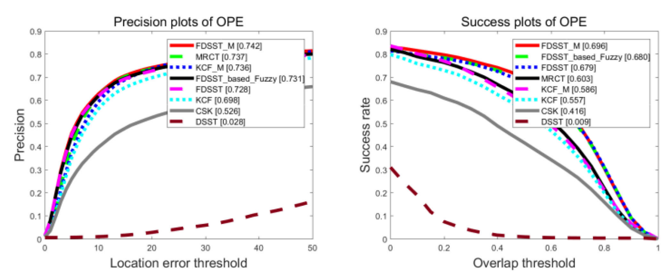


Fig. 7. The overall accuracy and success rate of FDSST_M, FDSST, FDSST_based_Fuzzy, MRCT, DSST, CSK, KCF, and KCF_M algorithms on the OTB100 platform.

C. Qualitative Analysis

The qualitative analysis of the improved algorithm is carried out by selecting key frames in part of the video sequences "bird1", "Couple", "Freeman1", and "Jumping". Fig. 8 shows the improvement effect of the KCF_M. In the figure, the red box represents the "ground truth" of the target, the green box represents the target box tracked by the KCF algorithm, and the yellow box represents the monitored target box by the improved KCF_M algorithm.

In the "bird1" sequence of Fig. 8(a), the target faces three difficulties: Fast Motion, Occlusion, and Out-of-View. The target is in a normal motion state at the 24^{th} frame. Both the KCF algorithm and the KCF_M algorithm can accurately monitor it. At the 61^{st} frame, the target started to move fast, which has exceeded the field of view of the camera, and the KCF algorithm can no

Fig. 8. Qualitative comparison of monitoring effects between the KCF_M algorithm and the KCF algorithm.

longer monitor it normally. However, the KCF_M algorithm that incorporates a multi-template update strategy for visual memory can still monitor it accurately. When the target moved to the 111st frame, the target began to be blocked by white clouds again. At this time, the KCF algorithm has been unable to accurately monitor it, but the KCF_M algorithm for multi-template update of fusion visual memory uses the weighted template in the confidence memory repository to achieve accurate monitoring, so that the target can be tracked in the subsequent 186th frame. This shows that the KCF_M algorithm using visual memory multi-template update can solve the difficult problems of Fast Motion, Occlusion, and Out-of-View very well.

In the "Couple" sequence of Fig. 8(b), the target faces two difficulties including Background Clutter and Low Resolution. When the target faces the challenge of Low Resolution at the 21st frame, both the KCF algorithm and the KCF_M algorithm can accurately monitor it. However, when the target is in the 48th frame, it is affected by the Background Clutter and Low Resolution, which causes the target template of the KCF algorithm to begin to be contaminated. However, the KCF_M algorithm that incorporates a multi-template update strategy for visual memory can use the weighted template in the confidence memory repository for accurate monitoring. In the subsequent 64th and 90th frames, the KCF_M algorithm continues to use the weighted template in the confidence memory repository and the unweighted template in the cognitive memory repository for location monitoring. This result also shows that the multi-template updating strategy of visual memory is an effective strategy to solve Background Clutter and Low Resolution.

In the "Freeman1" sequence of Fig. 8(c), the target faces three challenges: Scale Variation, In-Plane Rotation, and Out-of-Plane Rotation. At the 35th frame, the target is in a normal walking state, both the original KCF algorithm and the improved KCF_M algorithm can accurately monitor it. But at the 148th frame, the target began to change in scale, causing the template of the KCF algorithm to gradually become contaminated. The visual memory multi-template update strategy uses the weighted template and unweighted template stored in the confidence memory and cognitive memory to make corrections, which avoided the error of target position judgment caused by template contamination. In 174th frame, the target has undergone In-Plane Rotation and Out-of-Plane Rotation. Therefore, the multi-layer template update strategy again uses the weighted template in the confidence memory repository and the unweighted template in the cognitive memory storage library for correction, so that the target can still be accurately monitored at the subsequent 205th frame. It shows that the use of visual memory multi-template update strategy can solve the problems of Scale Variation, In-Plane Rotation, and Out-of-Plane Rotation very well.

In the "Jumping" sequence of Fig. 8(d), the target faces two challenges: Motion Blur and Fast Motion. At the 60th frame, the target is in an unobstructed environment, both the KCF algorithm and the KCF_M can accurately monitor it. When the target moves to the 98th frame, the KCF algorithm cannot monitor it accurately due to the rapid movement of the target, but the KCF_M algorithm using the visual memory multi-template update strategy do track it well. At the 138th frame, due to the rapid movement of the target, the rapid movement of the target resulted in the unclear target captured by the camera. At this time, the target template of the KCF algorithm has already been contaminated and is being monitored incorrectly. However, the KCF_M algorithm using the visual memory multi-template update strategy of visual memory corrects the target positioning process by combining the weighted template in the confidence memory repository and the unweighted template in the cognitive memory repository, so that the target can still be accurately located at the following frames. It shows that the KCF_M algorithm using the multi-template update strategy of visual memory is very effective in solving the problem of Motion Blur and Fast Motion.

V. Conclusion

Although the existing filter monitoring algorithm has achieved some good results in the multimedia background environment, once the monitoring fails, it loses the target's memory and cannot monitor it again. This paper introduced the visual memory multi-template update strategy into the CF-based monitoring algorithm, to achieve effective monitoring in the multimedia environment by considering the human memory update strategy. In this strategy, three memories were implemented: confidence memory, matching memory, and cognitive memory. By using these three memory and template interchange strategies, the improved algorithm achieved more accurate monitoring results without affecting the monitoring speed (still real-time). Through the experimental evaluation on the OTB platform, the results showed that this method is very effective in improving the robustness of the multimedia background environment. This template update mechanism infers that human cognitive computation/science such as human visual memory will bring newer ideas and methods in research of target tracking, which will obtain better effectiveness and accuracy from human cognition with complex environment and movement.

References

[1] L. Deng, "Artificial intelligence in the rising wave of deep learning: The historical path and future outlook [Perspectives]," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 180–177, Jan. 2018.

[2] S. Khattak, R. Hamzaoui, T. Maugey, S. Ahmad, and P. Frossard, "Bayesian early mode decision technique for view synthesis prediction-enhanced multiview video coding," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1126–1129, Nov. 2013.

[3] D. K. Shetty, U. Dinesh Acharya, N. Malarout, R. Gopakumar, and P. P. J., "A review of application of computer-vision for quality grading of food products," in *Proc. Int. Conf. Automat., Comput. Technol. Manage. (ICACTM)*, London, United Kingdom, pp. 297–303, 2019.

[4] B. G. Weinstein, "A computer vision for animal ecology." *J. Animal Ecol.*, vol. 87, no. 3, pp. 533–545, 2018.

[5] S. Ahmad, R. Hamzaoui, and M. Al-Akaidi, "Optimal packet loss protection of progressively compressed 3-D meshes," *IEEE Trans. Multimedia*, vol. 11, no. 7, pp. 1381–1387, Nov. 2009.

[6] M. Asad *et al.*, "A split target detection and tracking algorithm for ballistic missile tracking during the re-entry phase," *Defence Technol.*, vol. 16, no. 6, pp. 1142–1150, 2019.

[7] M. Roder, J. Cardinal, and R. Hamzaoui, "Efficient rate-distortion optimized media streaming for tree-structured packet dependencies," *IEEE Trans. Multimedia*, vol. 9, no. 6, pp. 1259–1272, Oct. 2007.

[8] K. Shojaei and M. Dolatshahi, "Line-of-sight target tracking control of underactuated autonomous underwater vehicles," *Ocean Eng.*, vol. 133, pp. 244–252, 2017.

[9] J. F. Henriques, C. Rui, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proc. 12th Eur. Conf. Comput. Vis.*, pp. 702–715, 2012 .

[10] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proc. Brit. Mach. Vis. Conf.*, pp. 1–11, Sep. 1-5 2014, doi: 10.5244/c.28.65.

[11] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2544–2550, 2010.

[12] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. S. Torr, "Staple: Complementary learners for real-time tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1401–1409, 2016.

[13] A. Lukežic, T. Vojír, L. C. Zajc, J. Matas, and M. Kristan, "Discriminative correlation filter with channel and spatial reliability," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4847–4856.

[14] M. Wang, Y. Liu, and Z. Huang, "Large margin object tracking with circulant feature maps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 4800–4808, 2017.

[15] Y. Yang, J. Yang, L. Liu, and N. Wu, "High-speed target tracking system based on a hierarchical parallel vision processor and gray-level LBP algorithm," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 47, no. 6, pp. 950–964, Jun. 2017.

[16] N. Wahlström and E. Özkan, "Extended target tracking using gaussian processes," *IEEE Trans. Signal Process.*, vol. 63, no. 16, pp. 4165–4178, 2015.

[17] Y. Li *et al.*, "Robust estimation of similarity transformation for visual object tracking," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, pp. 8666–8673, 2019.

[18] Q. Liu *et al.*, "EACOFT: An energy-aware correlation filter for visual tracking," *Pattern Recognit.*, vol. 112, 2021, doi: 10.1016/j.patcog.2020.107766.

[19] D. Li, J. Guo, and T. Xu, "The real-time target tracking algorithm based on improved template matching and its hardware implementation," in *Proc. 2015 Int. Conf. Commun., Signal Process. Syst.*, vol. 386, pp. 531–539, 2016.

[20] A. Bal and M. S. Alam, "Automatic target tracking in FLIR image sequences using intensity variation function and template modeling," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 5, pp. 1846–1852, Oct. 2005.

[21] F. Lamberti, A. Sanna, and G. Paravati, "Improving robustness of infrared target tracking algorithms based on template matching," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 2, pp. 1467–1480, Apr. 2011.

[22] Z. Jia, A. Balasuriya, and S. Challa, "Target tracking with Bayesian fusion based template matching," in *Proc. IEEE Int. Conf. Image Process.*, vol. 2, pp. II–826, 2005.

[23] X. Cheng, J. Liu, H. Wang, and C. Wang, "Coordinate live streaming and storage sharing for social media content distribution," *IEEE Trans. Multimedia*, vol. 14, no. 6, pp. 1558–1565, 2012.

[24] W. Tom, C. Wa, S. K. Roberts, and C. Sunita, "Ultrasound-based liver tracking utilizing a hybrid template/optical flow approach," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, pp. 1605–1615, 2018.

[25] D. A. Ross, J. Lim, R. S. Lin, and M. H. Yang, "Incremental learning for robust visual tracking," *Int. J. Comput. Vis.*, vol. 77, no. 1-3, pp. 125–141, 2008.

[26] L. Yang, J. Han, D. Zhang, N. Liu, and D. Zhang, "Segmentation in weakly labeled videos via a semantic ranking and optical warping network," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4025–4037, Aug. 2018.

[27] G. Ciaparrone *et al.*, "Deep learning in video multi-object tracking: A survey," *Neurocomputing*, vol. 381, pp. 61–88, 2020.

[28] J. Han, L. Yang, D. Zhang, X. Chang, and X. Liang, "Reinforcement cutting-agent learning for video object segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 9080–9089, 2018.

[29] S. Frintrop and M. Kessel, "Most salient region tracking," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2009, pp. 1869–18748.

[30] X. Xu, Z. Wang, and Z. Chen, "Visual tracking model based on feature-imagination and its application," in *Proc. Int. Conf. Multimedia Inf. Netw. Secur.*, Nanjing, China, pp. 370–374, 2010.

[31] H. Liu and Y. Shi, "Robust visual tracking based on selective attention shift," in *Proc. IEEE Control Appl., (CCA) Intell. Control*, pp. 1176–1179, 2009.

[32] W. Guo, C. Xu, S. Ma, and M. Xu, "Visual attention based motion object detection and trajectory tracking," in *Proc. Conf. Adv. Multimedia Inf. Process.*, vol. 6298, pp. 462–470, 2010.

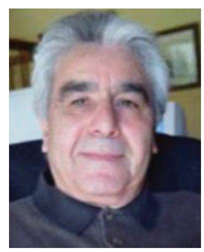
[33] L. Ma, C. Jian, L. Jing, J. Wang, and H. Lu, "Visual attention model based object tracking," in *Proc. Adv. Multimedia Inf. Process.-pcm -pacific Rim Conf. Multimedia*, vol. 6298, pp. 483–493, 2010.
[34] H. B. Kang and S. H. Cho, "Short-term memory-based object tracking," in *Proc. Image Anal. Recognit.: Int. Conf.*, vol. 3212, pp. 597–605, 2004.
[35] A. S. Montemayor, J. J. Pantrigo, and J. Hernández, "A memory-based particle filter for visual tracking through occlusions," in *Proc. Int. Work-Conf. Interplay Between Natural Artif. Computation*, Springer, vol. 5602, pp. 125–141, 2009.
[36] S. Liu, S. Wang, X. Liu, C.-T. Lin, and Z. Lv, "Fuzzy detection aided real-time and robust visual tracking under complex environments," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 1, pp. 90–102, Jan. 2021.
[37] Y. Wu, J. Lim and M. Yang, "Online object tracking: A benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, pp. 2411–2418, 2013.
[38] M. Kristan *et al.*, "The visual object tracking VOT2017 challenge results," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Venice, Italy, pp. 1949–1972, 2017.
[39] H. Fan *et al.*, "LaSOT: A high-quality benchmark for large-scale single object tracking," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Long Beach, CA, USA, pp. 5369–5378, 2019.
[40] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Discriminative scale space tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 8, pp. 1561–1575, Aug. 2017.
[41] H. Hu, B. Ma, J. Shen, and L. Shao, "Manifold regularized correlation object tracking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1786–1795, May 2018.

Shuai Liu received the B.S., M.S., and Ph.D. degrees from Jilin University, Changchun, China, in 2002, 2004, and 2011, respectively. He is currently a Full Professor with the College of Information Science and Technology and the Vice Director of the Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, Changsha, China. His research interests include computer vision and information processing. He is currently the Editor for many journals. He is currently a Senior Member of China Computer Federation.

Shuai Wang is currently working toward the master's degree under postgraduation with the College of Information Science and Technology, Hunan Normal University, Changsha, China. His research interests include computer vision, target tracking, and artificial intelligence.

Xinyu Liu is currently working toward the master's degree under postgraduation with the College of Information Science and Technology, Hunan Normal University, Changsha, China. Her research interests include computer vision, target tracking, and artificial intelligence.

Amir H. Gandomi (Senior Member, IEEE) is currently a Professor of data science and an ARC DECRA Fellow with the Faculty of Engineering and Information Technology, University of Technology Sydney (UTS), Ultimo, NSW, Australia. Prior to joining UTS, he was an Assistant Professor with the School of Business, Stevens Institute of Technology, Hoboken, NJ, USA and a Distinguished Research Fellow with BEACON center, Michigan State University, East Lansing, MI, USA. He has authored or coauthored more than 230 journal papers and seven books which collectively have been cited more than 20000 times (H-index = 66). His research interests include global optimization and (big) data analytics using machine learning and evolutionary computations. He has been named as one of the most influential scientific mind and Highly Cited Researcher (top 1% publications and 0.1% researchers) for four consecutive years from 2017 to 2020. He also ranked 18th in GP bibliography among more than 12,000 researchers. He was an Associate Editor, the editor, and the Guest Editor in several prestigious journals such as an Associate Editor for *Swarm and Evolutionary Computation*, the IEEE TRANSACTIONS ON BIG DATA, and IEEE INTERNET OF THINGS JOURNAL. He is active in delivering keynotes and invited talks..



Mahmoud Daneshmand (Senior Life Member, IEEE) received the B.S. and M.S. degrees in mathematics from the University of Tehran, Tehran, Iran and the M.S. and Ph.D. degrees in statistics from the University of California at Berkeley, Berkeley, CA, USA. He is the Co-Founder and currently a Professor with the Department of Business Intelligence and Analytics and a Professor with the Department of Computer Science, Stevens Institute of Technology, Hoboken, NJ, USA. He has more than 40 years of industry and university experience as a Professor, a Researcher, an Assistant Chief Scientist, the Executive Director, a Distinguished Member of Technical Staff, a Technology Leader, the Chairman of the Department, and the Dean of the School with Bell Laboratories, Murray Hill, NY, USA, AT&T Shannon Labs—Research, Florham Park, NJ, USA, the University of California at Berkeley, the University of Texas at Austin, Austin, TX, USA, the Sharif University of Technology, Tehran, Iran, the University of Tehran, New York University, New York, NY, USA, and the Stevens Institute of Technology, respectively.

Khan Muhammad (Member, IEEE) received the Ph.D. degree in digital contents from Sejong University, Seoul, South Korea. He is currently an Assistant Professor with the Department of Software, Director of Visual Analytics for Knowledge Laboratory (VIS2KNOW Lab), and a Lead Researcher of Intelligent Media Laboratory, Sejong University. He is currently a Professional Reviewer for more than 90 well-reputed journals and conferences. He has filed more than ten patents and authored or coauthored more than 150 articles in peer-reviewed international journals and conferences in his research areas. His research interests include medical image analysis (brain magnetic resonance imaging (MRI), diagnostic hysteroscopy, and wireless capsule endoscopy), information security (steganography, encryption, watermarking, and image hashing), video summarization, computer vision, fire or smoke scene analysis, and video surveillance. He is a Member of the Association for Computing Machinery.

Victor Hugo C. de Albuquerque (Senior Member, IEEE) graduated in mechatronics engineering from the Federal Center for Technological Education, Fortaleza, Brazil. He received the M.Sc. degree in tele informatics engineering from the Federal University of Ceará, Fortaleza, Brazil, in 2007 and the Ph.D. degree in mechanical engineering from the Federal University of Paraíba, João Pessoa, Brazil, in 2010. He is currently is a collaborator Professor and senior researcher at the Graduate Program on Teleinformatics Engineering at the Federal University of Ceará, Brazil, and at the Graduate Program on Telecommunication Engineering, Federal Institute of Education, Science and Technology of Ceará, Fortaleza/CE, Brazil. He is a specialist, mainly, in Image Data Science, IoT, Machine/Deep Learning, Pattern Recognition, Robotic.