REAL-TIME MOVING TARGET TRACKING ALGORITHM OF UAV/UGV HETEROGENEOUS COLLABORATIVE SYSTEM IN COMPLEX BACKGROUND

Xiao LIANG[1], Maoxian SHEN[2], Guangxun DU[3], Guodong CHEN[4]

Heterogeneous collaborative systems composed of UAV/UGV (Unmanned Aerial Vehicle and Unmanned Ground Vehicle) can cooperate to accomplish many complex tasks. The speed and accuracy of target tracking in UAV/UGV heterogeneous collaborative system are the basis of relative positioning and collaborative tasks. The correlation filter-based algorithm has achieved a satisfactory performance in many different simulation and experiments. However, the problems about occlusion, deformation and light changes limit its application. Based on context-aware correlation, an improved filter model is studied, and it can obtain different confidence maps by filtering different parts of the target. After the comparison of confidence maps from the proposed model and context-aware correlation filter, the maximum confidence is obtained. Simulation and experiment results show that compared with other correlation filter-based algorithms, the proposed method has a better performance in accuracy and success rate when dealing with occlusion, deformation and light changes.

Keywords: UAV/UGV collaborative system, Complex background, Moving target tracking, Correlation filter, Occlusion, deformation and light changes

1. Introduction

By the mutual cooperation and assist from UAV (Unmanned Aerial Vehicle) and UGV (Unmanned Ground Vehicle), UAV/UGV heterogeneous collaborative system can accomplish many complex and high-level tasks autonomously. The system uses the respective characteristics of UAV and UGV to complete more complex collaborative tasks. UAV flies in three-dimensional space and can detect a wide area. On the other hand, UGV moves in two-dimensional space but can conduct complex interactive activities. UAV/UGV heterogeneous collaborative system is a symbol that robots have higher autonomy and intelligence and it also brings many challenges to some new problems.

[1] Shenyang Aerospace University, China, E-mail: connyzone@126.com
[2] Shenyang Aerospace University, China
[3] Shenyang Aerospace University, China
[4] Shenyang Aerospace University, China

The speed and accuracy of target tracking in UAV/UGV heterogeneous collaborative system are the basis of relative positioning and collaborative tasks. The core function of collaborative positioning is target tracking which is an important bridge connecting UAV and UGV. The main purpose of target tracking is to determine the position of a selected target in a video. Based on the initial state of a target in the first frame, target tracking algorithm estimates the position of the target in subsequent frames. Many researchers have studied the problem of target tracking for years and carried out many solutions. Nowadays, the main challenge comes from illumination variations, occlusions, deformations, rotations and so on.

Generally, target tracking methods can be divided into two major categories: generative and discriminative method. Generative tracking learns a generative model from a given initial condition and then target tracking becomes a search problem which means it locates the target within the neighborhood of the target in previous frame. Discriminative methods are often taken as a binary classification problem [1]. The HOG [2] (Histogram of Oriented Gradient), haar features and regional covariance [3] are mainly used as the target representation. On the other hand, SVM (Support Vector Machine) [4], multi-instance learning [5], random forest [6] and Ada-boost [7] are usually taken as a visual classifier. The current mainstream algorithms are correlation filter tracking algorithm [8-14] and convolutional neural networks algorithms [15-20]. Specially, the correlation filter tracking algorithm has drawn much attention due to its better real-time performance and it is mainly based on the framework of ridge regression. About the framework of the correlation filter, Henriques transforms linear space to non-linear space by kernel function and uses the gradient histogram feature of multi-channel [2], the method improved the nonlinear tracking performance of correlation filter. Learning from the random fern classifier in TLD (Tracking-Learning-Detecting, TLD), Ma improves confidence filter by DSST (Accurate Scale Estimation for Robust Visual Tracking, DSST) [8]. Li proposed an adaptive scale kernel correlation filtering algorithm [10], which can solve the problem of target scale change, but it is slightly insufficient for other situations (such as occlusion and illumination). Based on correlation filter, Danelljan uses space penalty to handle boundary effect. But the method breaks the closed-form solution of ridge regression equation, so the equation can be only solved by Gauss-Seidel iteration which will slow down the calculation [12]. The method aims at long-term target tracking involving occlusion and it uses PSR (Peak SideLobe Ratio) to determine whether the target is occluded. After occlusion, the random fern classifier is used to reposition the target and the accuracy is greatly improved. Since the random fern classifier needs to be trained in each frame, the speed is too slow. In order to add more background information, Mueller proposes a new filter framework (context-aware correlation filtering) which has closed-loop solution

and better real-time capability [21], but the algorithm only eliminates background interference. By extracting the depth features of the target, convolutional neural network class target tracking has strong anti-interference ability. Therefore, target tracking with depth features is more accurate, but convolutional neural network depends on the performance of the hardware and it is insufficient in real time. In summary, the improvement of related filtering algorithms mostly increases accuracy by sacrificing real-time performance.

In this paper, a UAV/UGV heterogeneous collaborative system is proposed. In the system, UAV can track UGV automatically. So a real-time moving target tracking algorithm based on context-aware correlation filter is studied. In order to reduce the influence of occlusion, deformation and light changes, an improved filtering template is trained to filter two different components of target. Compared with the result of context-aware correlation filter, the maximum confidence can be obtained. By dividing the target into two parts, it can improve the performance of blocking deformation. The two components use different features and kernel functions to enhance the robustness. Simulation and experiment results show that the proposed method has a better performance in dealing with occlusion, deformation and light changes. Furthermore, the algorithm performances well both on accuracy and real-time.

The structure of this paper is as follows: the first chapter introduces the collaborative guidance system. The second chapter introduces the basic theory of the relevant filtering algorithm. The third chapter is the improvement. The fourth chapter is the simulation and experiment.

2. UAV/UGV Heterogeneous Collaborative System

The working process of the UAV/UGV heterogeneous collaborative system described in this article is as follows: UAV flies ahead of UGV. By the camera and other sensors, UAV can get a two-dimensional horizontal image of the environment and the image will supplement the barrier information in front of UGV. UAV can provide a global image and position information for the avoidance of UGV. Inside the heterogeneous collaborative system, UAV and UGV need collaborative positioning. Based on the collaborative awareness of the UAV/UGV, the heterogeneous collaborative system can complete several complex tasks such as cluster formation, collaborative avoidance and information fusion (Fig. 1).

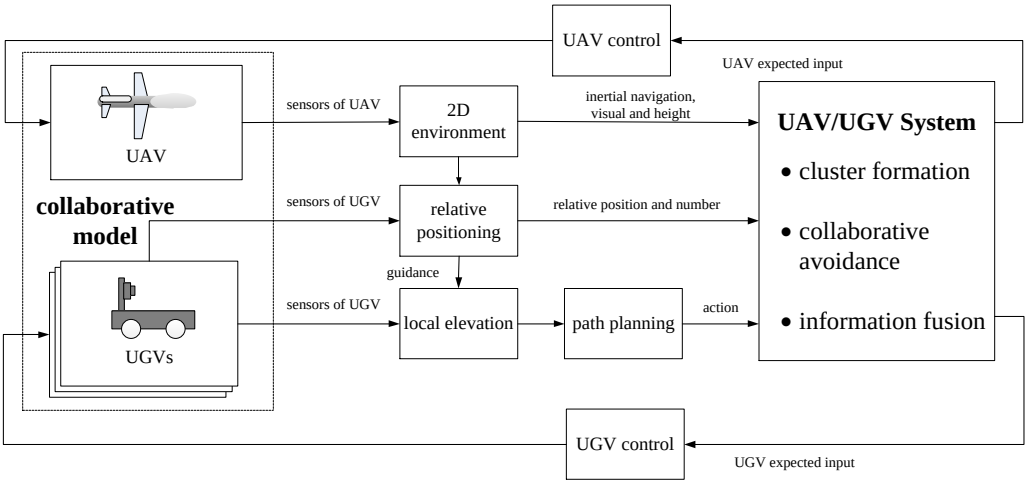


Fig. 1. UAV/UGV heterogeneous collaborative system

3. Correlation Filter Tracking Algorithm

The general idea is that basing on ridge regression, a cyclic matrix is calculated to train the classifier. By the classifier, the position of the target in new frame can be got. In order to accelerate the calculation, the cyclic matrix is transformed into the frequency domain.

3.1 Training Classifier

A ridge regression objective function from Reference [2] is

$$\min_{w} \sum_{i} (f(x_i) - y_i) + \lambda \|w\|^2, \tag{1}$$

$$f(x_i) = w^T x_i, \tag{2}$$

In Equations (1) and (2), w is filter template, y is ideal regression label (it is in Gaussian type) and λ is a regularization factor. When λ tends to zero, the model obtained will be complex and the model will be simple when λ tends to 1. x is the training sample and $f(x_i)$ is the actual regression value.

In Equation (1), calculate the derivative about w and let it equal to zero. So the extreme value is:

$$w = (X^T X + \lambda I)^{-1} X^T y, \tag{3}$$

Here, $X = (x_1, x_2 \cdots x_n)$ is a cyclic matrix and $X = Fdiag(\hat{x})F^H$ where $\hat{x} = F(x)$. F is a Fourier transform matrix and I is a unit matrix. The superscript H denotes conjugate transpose. Combine cyclic matrix and Equation (3) and

make Fourier transforms on both sides of the new equation, so there is Equation (4):

$$\hat{w} = \frac{\hat{x}^{*} \Theta \hat{y}}{\hat{x}^{*} \Theta \hat{x} + \lambda}, \tag{4}$$

In Equation (4), Θ represents the hadamard product.

3.2 Nonlinear Classifier

The correlation filter uses linear regression and the linear input can be mapped to the non-linear space of high dimension by kernel function method.

Filter template w is expressed as a linear combination of samples:

$$w = \sum_{i} \alpha_i \psi(x_i), \tag{5}$$

In Equation (5), α is the variable which needs to be optimized for dual space, and w is the variable for original space optimization, and $\psi(\ )$ is the mapping function for mapping x to the high-dimensional space.

Combine Equation (1), Equation (2) and Equation (5), then calculate the derivative about α. Let the result equals to zero, so:

$$\alpha = \frac{y}{\psi^{H}(X)\psi(X) + \lambda I}, \tag{6}$$

By the kernel function $K_{xx'} = k(x, x') = \psi^{H}(x)\psi(x')$, Equation (6) can be changed as:

$$\alpha = \frac{y}{K + \lambda I}, \tag{7}$$

When X is a cyclic matrix, the Gaussian kernel function is also a cyclic matrix and it has the following properties:

$$K = F diag(\hat{k}) F^{H}, \tag{8}$$

Combine Equations (7) and (8), then make Fourier transform on both sides of the new equation. So Equation (9) is as follow.

$$\hat{\alpha} = \frac{\hat{y}}{\hat{k}^{xx} + \lambda}, \tag{9}$$

3.3 Target Detection

When using linear filter template w, the detection function is:

$$f(z) = w^{T} z, \tag{10}$$

When using nonlinear filter template α, the detection function is:

$$f(z) = w^T z = \sum_{i=1}^{n} \alpha_i k(z, x_i), \tag{11}$$

In the Equations (10) and (11), z is the target area where needs to be detected. In order to accelerate the calculation, both of Equations (10) and (11) can be transformed into frequency domain.

4. Improved Correlation Filter Tracking

In the case of occlusion and deformation, the filter template will update error. In context-aware correlation filter, the robustness of filter template is enhanced by adding more information around target (such as upper, lower, left and right background information). In this paper, the target is divided into two components which are equivalent to double the positive and negative samples. Both of components with different features can be trained in a filter template. Then, the two results are compared with that in context-aware correlation filter. The confidence map with maximum confidence value is used to locate the position of target.

In this section, two components are selected. It is because the related filtering algorithm performs windowing on the target frame. The more the components are, the smaller the size of the components. When the windowing process is performed, the components of the true target are further reduced, which is not conducive to tracking the target.

Suppose the two components of target are a_1 and a_2, and their training samples are A_1 and A_2, respectively. y is the Gaussian regression label and the target equation is:

$$f_p(w, B) = \left\| Bw - \bar{y} \right\|_2^2 + \lambda_1 \left\| w \right\|_2^2, \tag{12}$$

Here $B = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$, $\bar{y} = \begin{bmatrix} y \\ y \end{bmatrix}$ and $\lambda_1 \in [0,1]$.

In Equation (12), calculate the derivative about w and let it equal to zero. So there is Equation (13):

$$w = \frac{A_1 y + A_2 y}{A_1{}^H A_1 + A_2{}^H A_2 + \lambda_1 I}, \tag{13}$$

Equation (13) can be converted to the Fourier domain:

$$\hat{w} = \frac{\hat{a}_1 \Theta \hat{y} + \hat{a}_2 \Theta \hat{y}}{\hat{a}_1{}^* \Theta \hat{a}_1 + \hat{a}_2{}^* \Theta \hat{a}_2 + \lambda_1}, \tag{14}$$

After mapping to non-linear high-dimensional space, Equation (14) can be rewritten as:

$$\alpha = \frac{K_{a_1a_1} y + K_{a_2a_2} y}{K^H_{a_1a_1} K_{a_1a_1} + K^H_{a_2a_2} K_{a_2a_2} + \lambda_1 I}, \tag{15}$$

In Equation (15), K is the inner product of a_1 or a_2 in high-dimensional space. Equation (15) is transformed to the Fourier domain as:

$$\hat{\alpha} = \frac{\hat{k}_{a_1a_1} \Theta\hat{y} + \hat{k}_{a_2a_2} \Theta\hat{y}}{\hat{k}_{a_1a_1}{}^{*} \Theta\hat{k}_{a_1a_1} + \hat{k}_{a_2a_2}{}^{*} \Theta\hat{k}_{a_2a_2} + \lambda_1}, \tag{16}$$

By Equations (10) and (13), the confidence map $map1$ can be got when detecting the image in a new frame. Similarly, by Equation (11) and (15), there is the confidence map $map2$. The confidence map $map3$ is obtained by context-aware correlation filter.

The maximum confidence value m in the three-confidence map is:

$$m = \max\{\max(map1), \max(map2), \max(map3)\}, \tag{17}$$

In Equation (17), the position of maximum confidence value is the target position.

5. Simulation and Experiment

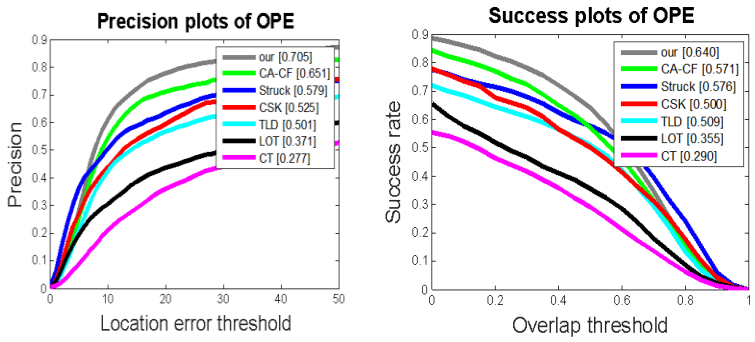
5.1 Simulation and Analysis

In simulation, several popular algorithms are selected to compare with the proposed method: Context-Aware Correlation Filter (CA-CF) [21], Structured Output Tracking with Kernels (struck) [22], Circulant Structure Kernel (CSK) [23], Tracking-Learning-Detecting (TLD) [24], Locally Orderless Tracking (LOT) [25] and Real-time Compressive Tracking (CT) [26].

Gradient histogram feature is taken as image feature. The test data are 50 groups of serial pictures from the common samples OTB-50 and OTB-100 and the samples include light changes, occlusion and deformation. The main hardware of simulation computer includes Intel Core i5-3230M CPU (2.60GHz) and 4GB running memory. The software platform is MATLAB2017a in 64bit Windows10 operating system.

Fig. 2 shows the precision plot and success plot of overall results. By the precision plot of each method, OPE (one-pass evaluation) robustness assessment is used to evaluate the performance. OPE measures the European distance error between the estimated and real center of target. In precision plots, the abscissa shows different thresholds of distance (pixel unit) while the ordinate is the percentage of qualified frames in total frames. Here, qualified frame means the center error of estimated target is less than the threshold. The percentage of threshold in 20 pixels is used as the evaluation score (CLE). The success plot

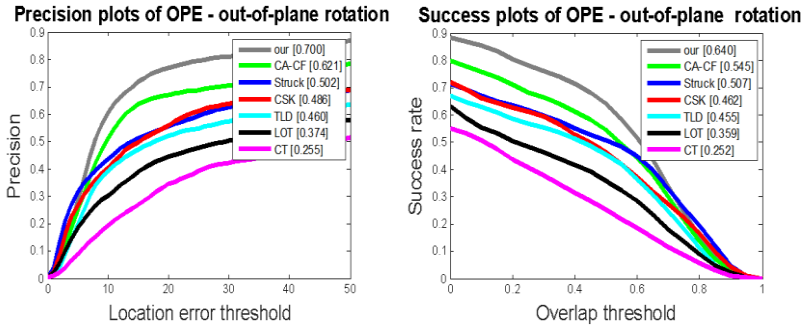
shows the rate of overlap between the estimated rectangle and the real rectangle of target. The calculation of success plot uses the area of the intersecting two rectangles divided by that of the union. In success plots, the abscissa is a continuous threshold from 0 to 1 while the ordinate is the percentage of qualified frames in total frames. Here, qualified frame means the overlap rate is greater than the threshold. The area below the curve represents the evaluation score.



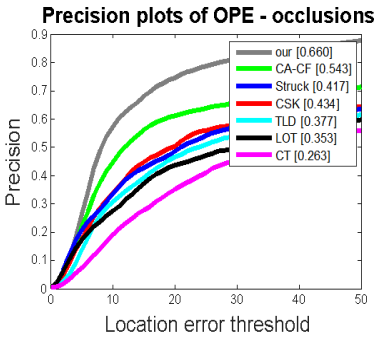
(a) (b)

Fig. 2. Precision plot and success plot of overall results

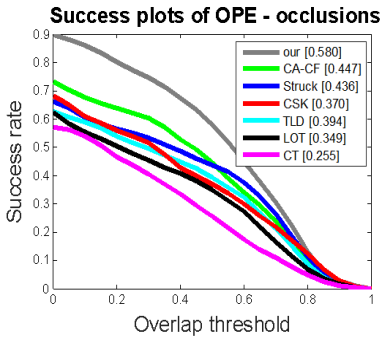
Here, gray line represents the proposed method, green line represents CA-CF, dark blue line represents struck, red line represents CSK, light blue line represents TLD, black line represents LOT and purple line represents CT, respectively. In Fig. 2(a), precision plots of OPE show that the algorithm has the highest accuracy rate than others. Compared with CA-CF, struck, CSK, TLD, LOT and CT, the accuracy rate increases by 5.4%, 12.6%, 18.1%, 20.4%, 33.4% and 42.8%, respectively. In Fig. 2(b), success plots of OPE show that the algorithm ranks first in success rate of all methods. Compared with CA-CF, struck, CSK, TLD, LOT and CT, the success rate increases by 6.9%, 6.4%, 14.1%, 13.1%, 28.5% and 35%, respectively. Overall, the proposed method has a better performance on accuracy and success rate.



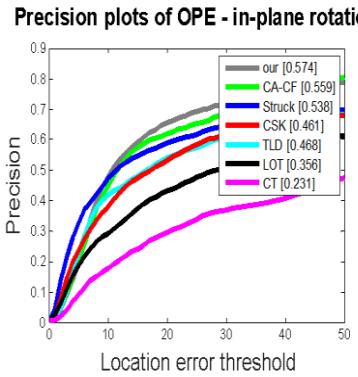
(a) (b)


Precision plots of OPE - occlusions
our [0.660]
CA-CF [0.543]
Struck [0.417]
CSK [0.434]
TLD [0.377]
LOT [0.353]
CT [0.263]
Precision
Location error threshold

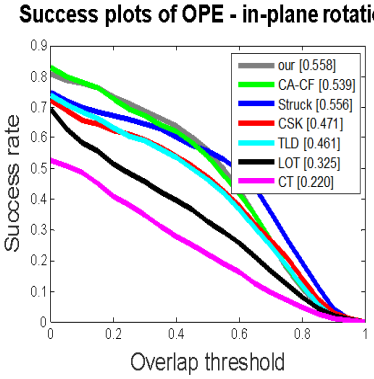
(c)


Success plots of OPE - occlusions
our [0.580]
CA-CF [0.447]
Struck [0.436]
CSK [0.370]
TLD [0.394]
LOT [0.349]
CT [0.255]
Success rate
Overlap threshold

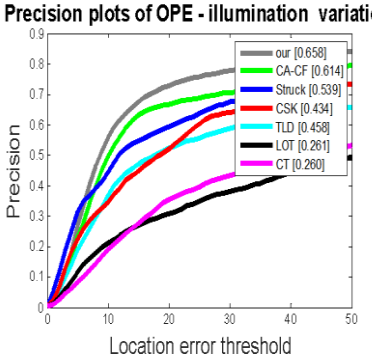
(d)


Precision plots of OPE - in-plane rotation
our [0.574]
CA-CF [0.559]
Struck [0.538]
CSK [0.461]
TLD [0.468]
LOT [0.356]
CT [0.231]
Precision
Location error threshold

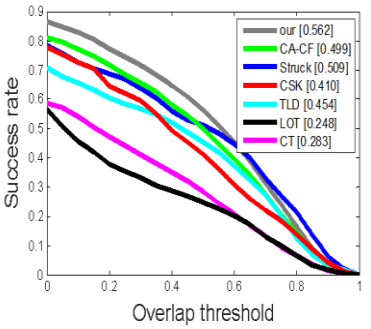
(e)


Success plots of OPE - in-plane rotation
our [0.558]
CA-CF [0.539]
Struck [0.556]
CSK [0.471]
TLD [0.461]
LOT [0.325]
CT [0.220]
Success rate
Overlap threshold

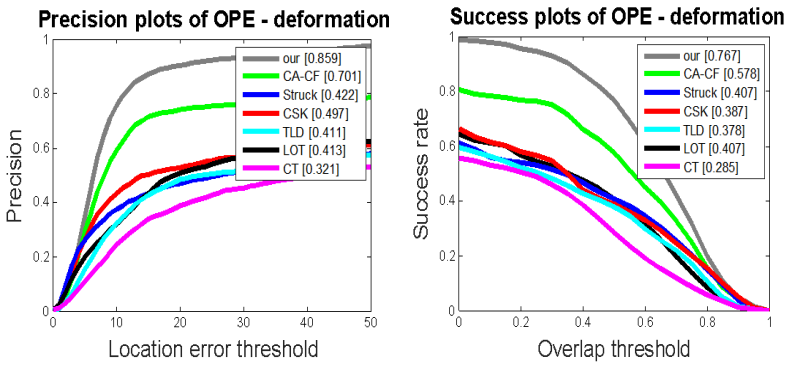
(f)


Precision plots of OPE - illumination variation
our [0.658]
CA-CF [0.614]
Struck [0.539]
CSK [0.434]
TLD [0.458]
LOT [0.261]
CT [0.260]
Precision
Location error threshold

(g)


Success plots of OPE - illumination variations
our [0.562]
CA-CF [0.499]
Struck [0.509]
CSK [0.410]
TLD [0.454]
LOT [0.248]
CT [0.283]
Success rate
Overlap threshold

(h)

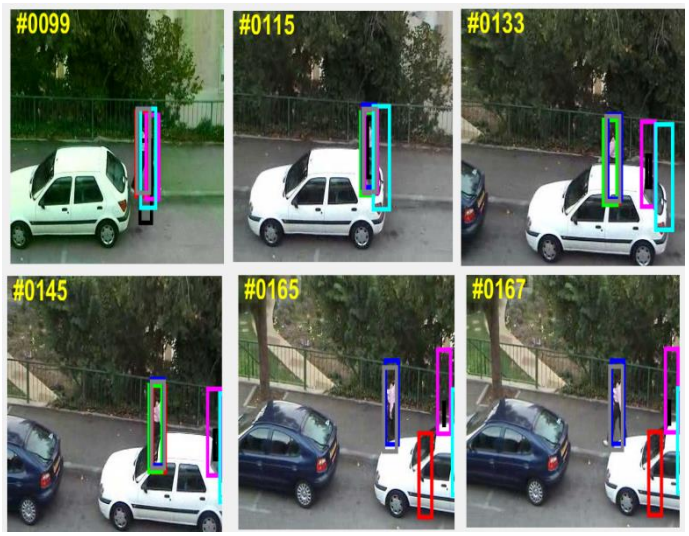


(i) (j)

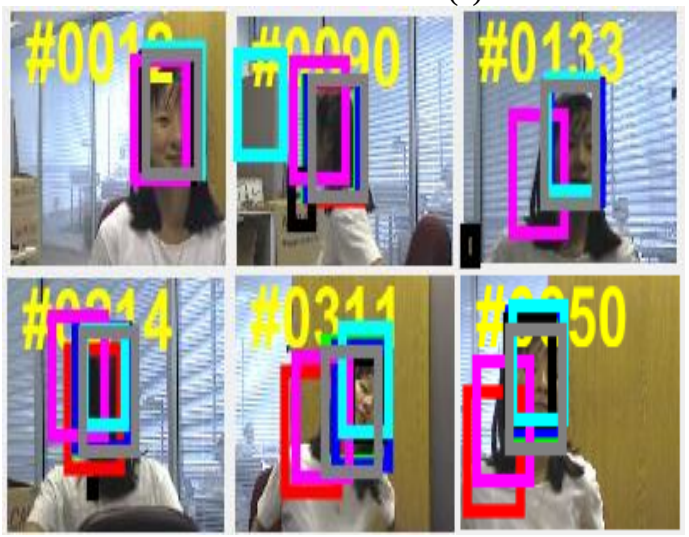
Fig. 3. Precision plot and success plot in different situations

The accuracy rate and success rate of out-plane rotation, occlusion, in-plane rotation, illumination variation and deformation are shown by Fig. 3(a) and (b), Fig. 3(c) and (d), Fig. 3(e) and (f), Fig. 3(g) and (h) and Fig. 3(i) and (j), respectively. From Fig. 3(a)(c)(e)(g)(i) and Fig. 3(b)(d)(f)(h)(j), the proposed method ranks first both on accuracy rate and success rate. In case of out-plane rotation, occlusion, in-plane rotation, illumination variation and deformation, the accuracy rate of the proposed method increases by 7.9%, 11.7%, 1.5%, 4.4% and 15.8% than CA_CF, respectively. In addition, the success rate of the proposed method increases by 9.5%, 13.3%, 1.9%, 6.3% and 18.9% than original context sensing correlation filter, respectively. Moreover, compared with the best performance among struck, CSK, TLD, LOT and CT, the accuracy rate increases by 19.8%, 3.6%, 22.6%, 11.9% and 36.2%, respectively. The accuracy rate increases by 13.3%, 0.2%, 14.4%, 5.3% and 36.1%, respectively.

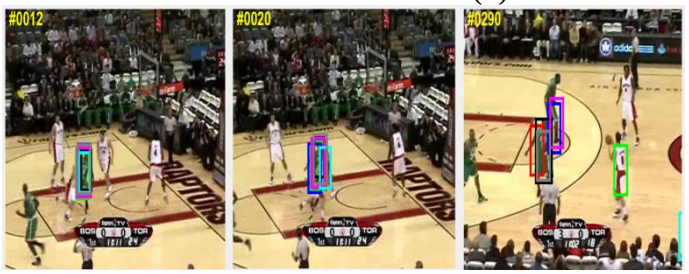
Furthermore, four sets of data including occlusion, deformation, rotation and light changes are used to test CA-CF, struck, CSK, TLD, LOT, CT and the proposed method. The four sets are named by Woman, Girl, Basketball and David3, respectively. Similar to Fig. 2 and Fig. 3, gray frame represents the proposed method, green frame represents CA-CF, dark blue frame represents struck, red frame represents CSK, light blue frame represents TLD, black frame represents LOT and purple frame represents CT, respectively.

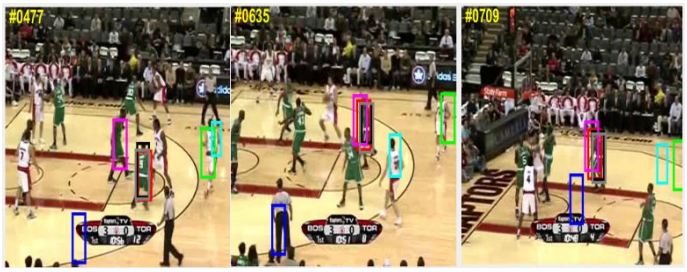


(a) Woman

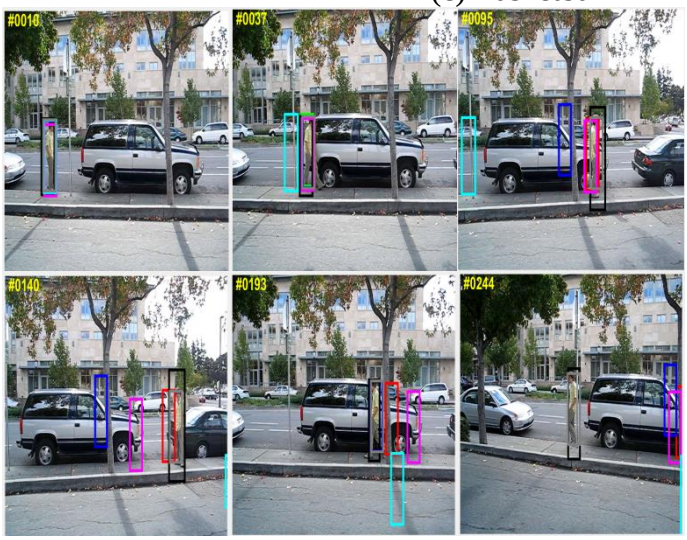


(b) Girl





(c) Basketball



(d) David3

Fig. 4. Results of moving target tracking in video

In woman dataset of Fig. 4(a), all algorithms can track a woman in initial stage. Once the woman is blocked by a car, TLD and CT will lose target. When occlusion lasts several frames, only struck and the proposed method can still track the woman. In girl dataset of Fig. 4(b), CT and CSK will fail to track in the case of rotation and deformation. In basketball dataset of Fig. 4(c), there is not only occlusion and deformation but also similar moving target. However, the proposed method still keeps tracking successfully. In David3 dataset of Fig. 4(d), occlusion, deformation and light changes exist in the same video. After twice of occlusion, it can be seen that struck and TLD lose the target. Then the target returns back, so it goes through twice of occlusion again. Finally, only the proposed method and LOT track the target successfully. From Fig. 4(a) to 4(d), the proposed method shows a better tracking performance in complex background including occlusion, deformation and light changes.

5.2 Indoor Experiment

The UAV in collaborative system is DJI Matrice 100 with N1 flight control module (Fig. 5). Matrice 100 is an open source flight platform developed by DJI. The platform fully supports SDK and external devices. By flexible API interface, custom commands can achieve autonomous flight.

N1 flight control module is a stand-alone component which also has onboard SDK. Its core parts include flight computer, Inertial Measurement Unit (IMU), gyroscope, accelerometer, barometer, GPS and so on.

Fig. 5. UAV in Collaborative System

The UGV in collaborative system is an omni-directional mobile robot which can reach any position by its own motion mechanism. Omni-directional mobile robots usually use Omni Wheel or Mecanum Wheel. The UGV in Fig. 6 is equipped with Mecanum Wheels.

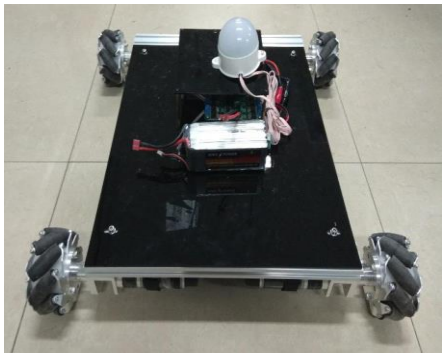


Fig. 6. UGV in Collaborative System

Basing on the kinematic analysis, the mathematical model of UGV is established. Then a control law is designed and downloaded to a control chip. To simplify the mathematical model, there are some assumptions:

1. With enough friction, UGV will not skid.

2. Four wheels are mounted on each corner of UGV respectively and parallel to each other.

3. The coordinates of UGV coincide with geographic coordinate.

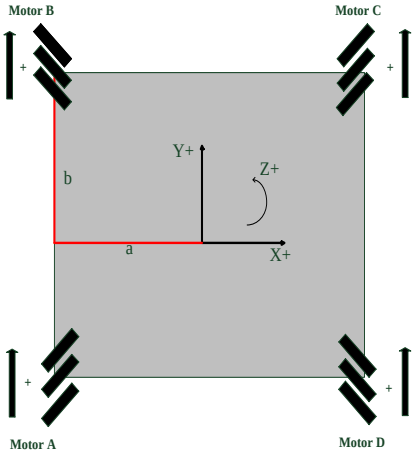


Fig. 7. Motion of UGV

According to Fig. 7, the motion of UGV can be linearly decomposed into three components. V_A, V_B, V_C and V_D represents the velocity of four wheels A, B, C and D, respectively. V_x is the velocity of UGV along X axis, V_y is the velocity of UGV along the Y axis and ω is the angular velocity around Z axis.

When UGV goes along X axis, there is

$$V_A = -V_x, V_B = +V_x, V_C = -V_x, V_D = +V_x, \tag{18}$$

When UGV goes along Y axis, there is

$$V_A = +V_y, V_B = +V_y, V_C = +V_y, V_D = +V_y, \tag{19}$$

When UGV rotates around X axis, there is

$$V_A = -\omega(a+b), V_B = -\omega(a+b), V_C = +\omega(a+b), V_D = +\omega(a+b), \tag{20}$$

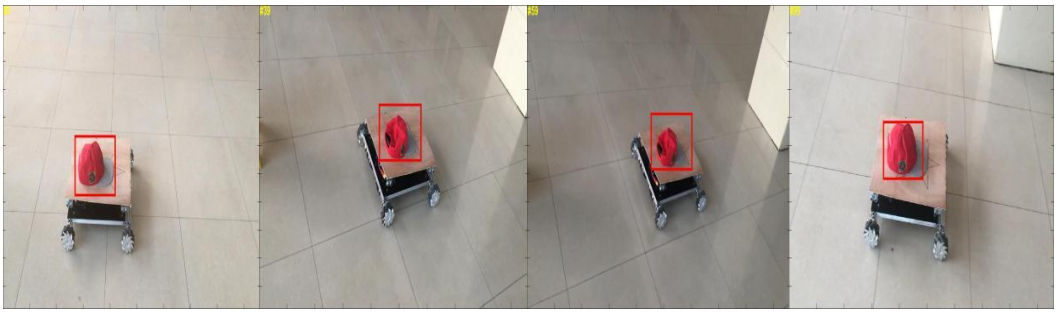
Basing on Equation (18), (19) and (20), the velocity of four wheels can be calculated according to the status of UGV.

$$\begin{cases} V_A = -V_x + V_y - \omega(a+b) \\ V_B = +V_x + V_y - \omega(a+b) \\ V_C = -V_x + V_y + \omega(a+b) \\ V_D = +V_x + V_y + \omega(a+b) \end{cases}, \tag{21}$$

After the design of hardware and control law, a collaborative guidance experiment is used to verify the proposed method as shown in Fig. 8.

(a) light changes

(b) occlusion



(c) deformation

Fig. 8. Collaborative guidance experiment

Fig. 8(a) shows that the proposed method can track the target when light changes. In Fig. 8(b), the proposed method will not lose the target when occlusion occurs. In Fig. 8(c), the target deforms first and then returns to initial state. However, the proposed method still shows a better performance. The precision plots of Fig. 8 correspond to Fig. 9(a), (b) and (c). When light changes in Fig. 8(a)

or occlusion occurs in Fig. 8(b), the accuracy rate of the proposed method is nearly 100% if the allowable error threshold is greater than 15 pixels.

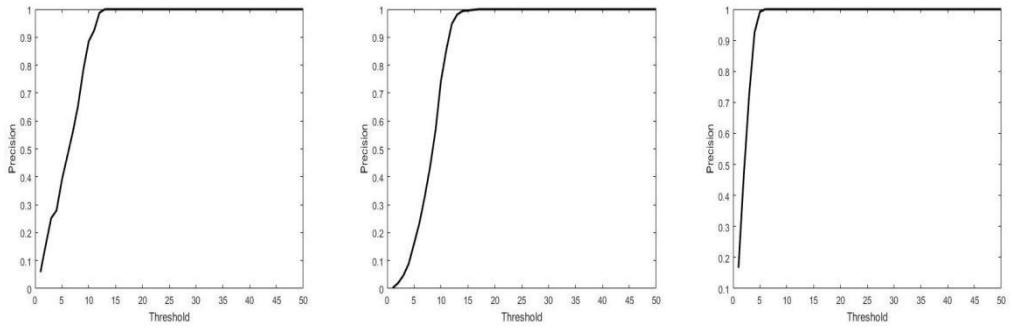


Fig. 9. Precision plots of collaborative guidance experiment

When deformation occurs in Fig. 8(c), the accuracy rate of the proposed method is nearly 100% if the allowable error threshold is greater than 5 pixels. In practical application, the allowable error threshold of 5 pixels or 15 pixels almost has no influence on tracking. The collaborative guidance experiment shows that the proposed method fulfills the requirement of tracking a moving target (about 50fps).

6. Conclusions

This paper first introduces a UAV/UGV heterogeneous collaborative system and then studies the method of tracking target for collaborative positioning which is the key part of the system. Since complex environment such as occlusion, deformation and light changes exists, the proposed method improves correlation filter by dividing the target into two components. Then the components are trained into a filter template so as to improve its robustness. In simulation, compared with several popular algorithms, the proposed method shows a better performance on both of accuracy and success rate. In indoor experiment with occlusion, deformation and light changes, the proposed method which is downloaded to a control chip can still track a moving target. The simulation and experiment show that the proposed method is effective and feasible for the UAV/UGV heterogeneous collaborative system. The real-time and accuracy of the tracking algorithm is crucial to the heterogeneous collaborative system, and the accuracy has a strong relationship with the feature extraction. Therefore, further work is to extract better features to improve the accuracy.

Acknowledgment

This work is supported by National Natural Science Foundation of China under Grant 61503255, Aeronautical Science Foundation of China under Grant 2016ZC54011 and Natural Science Foundation of Liaoning Province under Grant 2015020063. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

The authors declare that there is no conflict of interest regarding the publication of this article.

R E F E R E N C E S

[1]. *Z. Chen, Z. Hong and D. Tao*, "An experimental survey on correlation filter-based tracking", in Computer Science, **vol. 53**, no. 6025, 2015, pp. 68-83.

[2]. *J. F. Henriques, R. Caseiro, P. Martins and J. Batista*, "High speed tracking with kernelized correlation filters", in IEEE Transactions on Pattern Analysis and Machine Intelligence, **vol. 37**, no. 3, 2015, pp. 583-596.

[3]. *J. Jimenez, A. Martin, V. Uc and A. Espinosa*, "Mexican sign language alphanumerical gestures recognition using 3D Haar-like features", in IEEE Latin America Transactions, **vol. 15**, no. 10, 2017, pp. 2000-2005.

[4]. *C. Singh, E. Walia and K. P. Kaur*, "Enhancing color image retrieval performance with feature fusion and non-linear support vector machine classifier", in Optik-International Journal for Light and Electron Optics, **vol. 158**, 2018, pp. 127-141.

[5]. *O. Z. Kraus, J. L. Ba and B. J. Frey*, "Classifying and segmenting microscopy images with deep multiple instance learning", in Bioinformatics, **vol. 32**, no. 12, 2016, pp. i52-i59.

[6]. *T. Li, B. Ni, X. Wu, Q. Gao, Q. Li and D. Sun*, "On random hyper-class random forest for visual classification", in Neurocomputing, **vol. 172**, no. C, 2016, pp. 281-289.

[7]. *J. Rajeshwari and K. Karibasappa*, "Adaboost modular tensor locality preservative projection: face detection in video using Adaboost modular-based tensor locality preservative projections", in let Computer Vision, **vol. 10**, no. 7, 2017, pp. 670-678.

[8]. *C. Ma, X. Yang, C. Zhang and M. Yang*, "Long-term correlation tracking", Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, Massachusetts, USA, 2015, pp. 5388-5396.

[9]. *R. Yang and Z. Wei*, "Extended kernelised correlation filter tracking", in Electronics Letters, **vol. 52**, no. 10, 2016, pp. 823-825.

[10]. *Y. Li and J. Zhu*, "A scale adaptive kernel correlation filter tracker with feature integration", Proceedings of 2014 IEEE International European Conference on Computer Vision, Zurich, 2014, pp. 6-12.

[11]. *J. Kwon, K, Kim and K. Cho*, "Multi-target tracking by enhancing the kernelised correlation filter-based tracker", in Electronics Letters, **vol. 53**, no. 20, 2017, pp. 1358-1360.

[12]. *M. Danelljan, G. Hager, F. S. Khan and M. Felsberg*, "Learning spatially regularized correlation filters for visual tracking", Proceedings of 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015, pp. 4310-4318.

[13]. *S. A. Wibowo, H. Lee, E. K. Kim and S. Kim*, "Collaborative learning based on convolutional features and correlation filter for visual tracking", in International Journal of Control Automation and Systems, **vol. 16**, no. 1, 2018, pp. 335-349.

[14]. *J. F. Henriques, R. Caseiro, P. Martins, and J. Batista*, "High-Speed Tracking with Kernelized Correlation Filters", in IEEE Transactions on Pattern Analysis & Machine intelligence, **vol. 1**, no. 3, 2017, pp. 125-141.

[15]. *Y. Chen, X. Yang, B. Zhong, S. Pan, D. Chen and H. Zhang*, "CNNTracker: Online discriminative object tracking via deep convolutional neural network", in Applied Soft Computing, **vol. 38**, 2016, pp. 1088-1098.

[16]. *S. Bell and K. Bala*, "Learning visual similarity for product design with convolutional neural networks", in ACM Transactions on Graphics, **vol. 34**, no. 4, 2015, pp. 1-10.

[17]. *R. Tapu, B. Mocanu and T. Zaharia*, "DEEP-SEE: Joint object detection, tracking and recognition with application to visually impaired navigational assistance", in Sensors, **vol. 17**, no. 11, 2017, pp. 2473-2497.

[18]. *A. Elhayek, E. D. Aguiar, A. Jain, J. Thompson, L. Pishchulin, M. Andriluka, C. Bregler, B. Schiele and C. Theobalt*, "MARCOnI-ConvNet-Based MARker-Less motion capture in outdoor and indoor scenes", in IEEE Transactions on Pattern Analysis and Machine Intelligence, **vol. 39**, no. 3, 2017, pp. 501-514.

[19]. *J. Chung and K. Sohn*, "Image-based learning to measure traffic density using a deep convolutional neural network", in IEEE Transactions on Intelligent Transportation Systems, **vol. 19**, no. 5, 2018, pp. 1670-1675.

[20]. *I. B. Akkaya and U. Halici*, "Mouse face tracking using convolutional neural networks", in let Computer Vision, **vol. 12**, no. 2, 2018, pp. 153-161.

[21]. *M. Mueller, N. Smith and B. Ghanem*, "Context-aware correlation filter tracking, In IEEE Conference on Computer Vision and Pattern Recognition", Honolulu, Hawaii, USA, 2017, pp. 1387-1395.

[22]. *S. Hare, S. Golodetz, A. Saffari, V. Vineet, M. Cheng, S. L. Hicks and P. H. S. Torr*, "Struck: structured output tracking with kernels", in IEEE Transactions on Pattern Analysis and Machine Intelligence, **vol. 38**, no. 10, 2016, pp. 2096-2109.

[23]. *G. Kim, S. Jeong and S. Lee*, "Improved circulant structure with kernel tracker using correlation kernel based on wavelet feature", in TechArt Journal of Arts and Imaging Science, **vol. 3**, no. 2, 2016, pp. 43-46.

[24]. *Z. Kalal and K. Mikolajczyk and J. Matas*, "Tracking-learning-detection", in IEEE Transactions on Pattern Analysis and Machine Intelligence, **vol. 34**, no. 7, 2012, pp.1409-1422.

[25]. *S. Avidan, D. Levi, A. Barhillel and S. Oron*, "Locally orderless tracking", in International Journal of Computer Vision, **vol. 111**, no. 2, 2012, pp. 213-228.

[26]. *S. Pan, L. Shi and S. Guo*, "A Kinect-Based Real-Time Compressive Tracking Prototype System for Amphibious Spherical Robots", in Sensors, **vol. 15**, no. 4, 2015, pp. 8232-8252.