Proceeding of the IEEE
International Conference on Robotics and Biomimetics
Dali, China, December 2019

A Two-level Adaptive Target Recognition and Tracking Method Based on Vision for Multi-robot System

Liang Ren[1,2], Zhiqiang Cao[1], Min Tan[1], Peng Zhao[2], Xuechao Chen[3]
1 State Key Laboratory of Management and Control for Complex Systems, CASIA;
University of Chinese Academy of Sciences, Beijing, China
2 The 21st Research Institute of China Electronics Technology Group Corporation, Shanghai, China
3 Beijing Advanced Innovation Center for Intelligent Robots and Systems, BIT, Beijing, China

Abstract – **The vision-based target recognition and tracking have received much attention in the field of robotics. Existing methods mainly focus on the vision perception of individual robot with a single view, however, the performance is susceptible to illumination and occlusion. Multi-robot collaborative perception provides a potential solution to deal with the limitation of single-view observation, however, the challenging of environmental adaptability for multi-robot collaborative decision still remains unsolved. To solve this problem, this paper proposes a two-level adaptive target recognition and tracking method based on vision for multi-robot system. The problem of multi-robot target recognition and tracking is solved under a two-level framework, which contains the features fusion level of individual robot and the cooperation level of multi-robot system. In the first level, the features measuring results that influence the visual perception of individual robot are fused, while the second level combines the voting of each robot together to determine the target for multi-robot system. Both the features measuring weights and robots voting weights are adaptively updated according to their evaluation, which lead to a beneficial result where the features and robots with higher accuracy play major roles in the first and second levels, respectively. Therefore, a good adaptability to the environments can be guaranteed. The experimental results show that the proposed approach can realize the coordination of multi-robot system in target recognition and tracking with an effective performance.**

Index Terms – **multi-robot system, collaborative visual perception, two-level framework, features measuring weights, robots voting weights.**

I. INTRODUCTION

Vision is a main means for robot to perceive the outside world, which provides the mobile robot with intuitive and rich environmental information. At present, target recognition and tracking based on vision have been widely used in robot navigation, video surveillance, and industrial testing, etc.

In the early visual perception field, template matching provides a solution for target recognition [1]-[4], which needs to specify the target template in advance, and then extracts features and calculates the correlation between template and target. Although template matching-based method is simple, it is restricted by threshold selection. It is difficult to meet the accuracy requirement in complex scenes. Nowadays, the object recognition based on machine learning is popular with a high accuracy. This method can deal with complex backgrounds or non-rigid objects recognition. It obtains classifier by training samples, and then classify test samples by the trained classifier, where the feature extraction and classifier learning are important.

The common features include color [5][6], texture [7]-[10], and shape [11]-[13]. To avoid the disturbance of similar objects, weighted summation of measuring results is used to obtain a more discriminative result. In the fusion process of different feature measuring results, the proportion setting of each feature is important. The feature measuring weight is mostly preset according to the prior knowledge. For the fixed feature measuring weights, it is difficult or impossible to find a set of specific weight values to fit all environments.

Classifier is a model that uses the limited training samples set to estimate the posterior probability of object belonging to a certain type as accurately as possible. There are mainly two kinds of models for predicting object class by estimating the posterior probability: generative model and discriminant model. The generative model first models the joint probability distribution, and then obtains the posterior probability. Bayesian classifier [14] is a typical one. Directly estimating joint probability based on limited training samples shall result in the combination explosion in the calculation. Therefore the modeling of the joint probability distribution by generative model depends on some assumptions, but these assumptions do not always hold in practice. The discriminant model can predict category by directly modeling the posterior probability. A representative classifier is the support vector machines (SVM) [15]. SVM is a binary classification model, by finding a classification hyperplane in the feature space, the data sample points of different categories are separated, and the interval between different types of points is maximized.

For single target tracking algorithms, there are mainly two types: generation-based tracking methods and discrimination-based ones. The former is regarded as a search problem and the most similar target is found in the adjacent areas. The latter treats the tracking process as a classification problem and classifies the target from the backgrounds.

The generation-based tracking is to model the target area in the current frame, and look for the most similar area as the prediction position in the next frame. Classical methods include kalman filtering [16]-[18], particle filtering [19][20], mean-shift method [21]. A problem of this method is the interference of background noise. Influenced by the noise, the offset of target tracking happens. The discrimination-based tracking generally takes the target area of the current frame as a positive sample, and take the background region as a negative sample. A classifier is trained to distinguish the target and background in

978-1-7281-6321-5/19//$31.00 ©2019 IEEE

the next frame. Representative methods include TLD [22][23], StruckSVM [24], etc. With the high computational efficiency of correlation filtering, the tracking based on correlation filtering can achieve a fast processing. Henriques *et al*. brought the nuclear space into correlation filtering and proposed KCF [25] tracking algorithm, which constructs the sample set into a cyclic matrix, and obtains the analytic solution of ridge regression classifier in the frequency domain by fast Fourier transform.

At present, there have been abundant theoretical achievements in target recognition and tracking for a single robot. However, few research focuses on collaborative perception of multi-robot system. Every robot in the system observes the target from different views, and an effective collaboration of multiple robots can undoubtedly improve the whole performance. Target recognition and tracking of multi-robot system refers to joint analysis of the visual perception data of each robot to obtain the interested target and its motion process. Compared with the case of a single robot system, the solution of multi-robot system can effectively handle the problems such as limited field of view of a single robot. In this paper, a two-level adaptive target recognition and tracking method based on vision for multi-robot system is proposed. The proposed method weights the results of robot perception and fuses the results of feature measurement, where the voting weight of each robot and the measuring weight of each feature are adaptively updated. This solution can lead to a fact that the robot and feature with high accuracy dominate the visual perception process. The proposed method realizes the collaboration of multi-robot system in target recognition and tracking, which can enhance the adaptability to the environments.

The rest of this paper is organized as follows. Section II gives the problem statement. The proposed approach is described in Section III in detail. The experiment results are given in Section IV and Section V concludes the paper.

II. Problem Statement

This paper is motivated by the problem of target recognition and tracking suitable for multi-robot systems. To improve the multi-robot collaborative visual perception performance in unknown environments, the problem is divided into two levels: the features fusion level of single robot and the cooperation level among the robots.

We label $d_{i,j}(z)$ as the features fusion result of the jth robot at the ith moment where j=1,2,…, J represents the index of the robot. The measuring weight of each feature is given the same value at the initial moment, and the measurement result of each feature can be acquired by SVM [15].

$$d_{i,j}(z) = \sum_{k=1}^{K} \lambda_{i,j,k}(z) * d_{i,j,k}(z) \quad (1)$$

where $d_{i,j,k}(z)$ and $\lambda_{i,j,k}(z)$ are the measurement result and measuring weight of each feature, respectively, and k=1,...K represents the index of the feature.

The target recognition result for the area z at the ith moment is expressed by V_i. The value of V_i is decided by multi-robot system voting, which is represented by (2). Note that the voting weight of each robot is given the same value at the initial moment.

$$V_i = \sum_{j=1}^{J} w_{i,j}(z) * d_{i,j}(z) \quad (2)$$

where $w_{i,j}(z)$ is the voting weight of the robot j.

As mentioned above, a two-level framework for target recognition and tracking is considered. In the first level, the visual perception result of each robot is obtained by combining the measurement results of multiple features. In the second level, according to the weighted voting of robots, the visual perception result of multi-robot system is obtained. Especially, both the measuring weight of each feature and the voting weight of each robot are adaptively updated. The two-level framework for target recognition and tracking of multi-robot system is illustrated in Fig. 1. The visual perception result V_i was obtained by (1) and (2). When the value of V_i is no less than a given non-negative threshold σ, the system determines the target and tracks it. In the process of tracking, $\lambda_{i,j,k}(z)$ and $w_{i,j}(z)$ are updated to recalculate V_i to ensure the quality of visual perception.

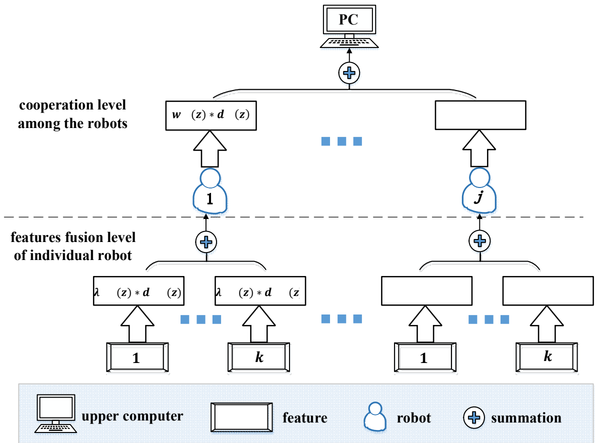


Fig. 1. The two-level framework for target recognition and tracking of multi-robot system.

III. The Proposed Approach

A. *The adaptive updating of voting weight of each robot*

The robot should give respective voting weight according to its current observation to the target. Considering that the accuracy of each robot is not always the same for the changing environment, the voting weight of each robot will be updated adaptively according to the perception accuracy of each robot. A guidance principle is that the robots with higher accuracy dominate the process of visual perception of the robotic system.

The adaptive updating of voting weight can be deduced by minimizing the exponential loss function. In this paper, the exponential loss function instead of 0/1 loss function is selected as objective function, as shown in (3).

The real class of the identified region z is denoted by $y \in \{-1,1\}$, the partial derivative of L_i with respect to V_i is performed, and we have (4).

$$L_i = E_z(\mathrm{e}^{-y*V_i}) = \mathrm{e}^{-V_i}\mathrm{P}(y=1|z) + \mathrm{e}^{V_i}\mathrm{P}(y=-1|z) \quad (3)$$

$$\frac{\partial L_i}{\partial V_i} = -\mathrm{e}^{-V_i}\mathrm{P}(y=1|z) + \mathrm{e}^{V_i}\mathrm{P}(y=-1|z) \quad (4)$$

With $\frac{\partial L_i}{\partial V_i} = 0$, one can obtain (5).

$$V_i = 0.5\log\frac{\mathrm{P}(y=1|z)}{\mathrm{P}(y=-1|z)} \tag{5}$$

Therefore,

$$sign(V_i) = sign\left[0.5\log\frac{\mathrm{P}(y=1|z)}{\mathrm{P}(y=-1|z)}\right]$$
$$= \begin{cases} 1, \mathrm{P}(y=1|z) > \mathrm{P}(y=-1|z) \\ -1, \mathrm{P}(y=1|z) < \mathrm{P}(y=-1|z) \end{cases} \tag{6}$$

$$sign(V_i) = \underset{y\in\{-1,1\}}{argmax}\, \mathrm{P}(y|z) \tag{7}$$

As can be seen from (7), the bayesian optimum is realized with the minimum classification error rate. It means that minimizing the exponential loss function can minimize the error rate of dichotomies. Thus, the exponential loss function L_i is used to replace the 0/1 loss function as the optimization objective of the dichotomy problem.

After confirming the rationality of the exponential loss function, the updating of voting weight for each robot is derived as follows. To minimize the classification error rate of multi-robot system, one only needs to minimize the classification error rate of each robot. Without loss of generality, take the *j*th robot as an example and the target function is shown in (8).

$$L_{i,j} = E_z\left(\mathrm{e}^{-y*w_{i,j}*d_{i,j}}\right)$$
$$= \mathrm{e}^{-w_{i,j}}\mathrm{P}\left(y = d_{i,j}|z\right) + \mathrm{e}^{w_{i,j}}\mathrm{P}\left(y \neq d_{i,j}|z\right) \tag{8}$$

The partial derivative of the target function $L_{i,j}$ with respect to $w_{i,j}$ is given by

$$\frac{\partial L_{i,j}}{\partial w_{i,j}} = -\mathrm{e}^{-w_{i,j}}\mathrm{P}\left(y = d_{i,j}|z\right) + \mathrm{e}^{w_{i,j}}\left(1 - \mathrm{P}\left(y = d_{i,j}|z\right)\right) \tag{9}$$

With $\frac{\partial L_{i,j}}{\partial w_{i,j}} = 0$, and the voting weight of robot can be obtained as follows.

$$w_{i,j} = 0.5\log\frac{\mathrm{P}\left(y = d_{i,j}|z\right)}{1-\mathrm{P}\left(y = d_{i,j}|z\right)} \tag{10}$$

The tracking confidence $T_{i,j}$ acquired by KCF [25] reflects the degree of discrimination in the target space, which can be used to measure $\mathrm{P}\left(y = d_{i,j}|z\right)$. At the same time, it is necessary to ensure that (10) is meaningful and monotonous. Constructing the exponential function $\mathrm{P}\left(y = d_{i,j}|z\right) = 1 - 0.5 * \mathrm{e}^{-T_{i,j}}$ whose value is between 0.5 and 1 with a monotonous increasing. Thus, the updating of voting weight for each robot can be obtained by (11).

$$w_{i,j} = 0.5\log(2 * \mathrm{e}^{T_{i,j}} - 1) \tag{11}$$

It is noted that the voting weight of each robot indicates the importance of the corresponding robot in the decision-making of the multi-robot system, and the sum of the voting weights of all robots may not be required to be 1.

B. The adaptive updating of feature measuring weight

The perception accuracy of individual robot is also an important factor to affect the cooperative perception of multi-robot system. Hence, the weighted sum of features that affect the visual perception of individual robot is carried out, meanwhile, the feature measuring weight will be updated adaptively. As a result, features with higher accuracy will dominate the process of visual perception, which can improve the adaptability to environment for the individual robot.

Similar to the updating solution of voting weight of each robot, the exponential loss function is still used as the optimization goal to derive the measuring weight for each feature. In order to minimize the classification error rate of multiple features for a single robot, one only needs to minimize the classification error rate of each feature. Without loss of generality, take the *k*th feature as an example and we have the following target function:

$$L_{i,j,k} = E_z\left(\mathrm{e}^{-y*\lambda_{i,j,k}*d_{i,j,k}}\right)$$
$$= \mathrm{e}^{-\lambda_{i,j,k}}\mathrm{P}\left(y = d_{i,j,k}|z\right) + \mathrm{e}^{\lambda_{i,j,k}}\mathrm{P}\left(y \neq d_{i,j,k}|z\right) \tag{12}$$

The partial derivative of the target function $L_{i,j,k}$ with respect to $\lambda_{i,j,k}$ is shown in (13).

$$\frac{\partial L_{i,j,k}}{\partial \lambda_{i,j,k}} = -\mathrm{e}^{-\lambda_{i,j,k}}\mathrm{P}\left(y = d_{i,j,k}|z\right) + \mathrm{e}^{\lambda_{i,j,k}}\left(1 - \mathrm{P}\left(y = d_{i,j,k}|z\right)\right) \tag{13}$$

With $\frac{\partial L_{i,j,k}}{\partial \lambda_{i,j,k}} = 0$, the measure weight of *k*th feature can be obtained as follows.

$$\lambda_{i,j,k} = 0.5\log\frac{\mathrm{P}\left(y = d_{i,j,k}|z\right)}{1-\mathrm{P}\left(y = d_{i,j,k}|z\right)} \tag{14}$$

$\frac{d_{i,j,k}}{d_{i,j}}$ expresses the proportion of this feature in the fusion result of multiple features, and it can be used to measure $\mathrm{P}\left(y = d_{i,j,k}|z\right)$. In this paper, $\mathrm{P}\left(y = d_{i,j,k}|z\right) = 1 - 0.5 * \mathrm{e}^{-\frac{d_{i,j,k}}{d_{i,j}}}$ is constructed, whose value is between 0.5 and 1, and it is monotonically increasing. Then, the updating of measuring weight for *k*th feature is expressed as follows.

$$\lambda_{i,j,k} = 0.5\log(2 * \mathrm{e}^{\frac{d_{i,j,k}}{d_{i,j}}} - 1) \tag{15}$$

It is noted that the measuring weight of each feature indicates the importance of the corresponding feature in the decision-making of robot, and the sum of the measuring weights of features is not required to be 1.

C. The collaborative visual perception of multi-robot system

Algorithm 1 illustrates the process of the collaborative visual perception of multi-robot system with adaptive updating of robot voting weight and feature measuring weight.

At the initial moment, all the robots detecting the interested area constitute the set Ω. Then the system determines the target and sends the information to every robot. The robots whose judgments are not consistent with the system will exit from the set Ω, while the others track the target and continue to participate in the voting decision. These robots in Ω update

$w_{i,j}(z)$ and $\lambda_{i,j,k}(z)$ to re-identify the tracking result in the process of tracking. The process of the visual perception for multi-robot system is dominated by the robots and features with high weights. It is worth noting that V_i is only decided by the robots in Ω, and the robots without in the Ω will exercise according to the target information sent by the system.

Algorithm 1. The collaborative visual perception of multi-robot system

Input: The visual information captured by each robot.
Output: the robot voting weight $w_{i,j}(z)$, the feature measuring weight $\lambda_{i,j,k}(z)$, and the visual perception result V_i

```
Initialize Ω, w_{i,j}(z), and λ_{i,j,k}(z);
flag = 0;
for time i=1,…,I;
  if flag = 0
    for each robot in Ω
      compute d_{i,j}(z) according to (1);
    end for
    compute V_i according to (2);
    if V_i ≥ σ then flag = 1;
      for each robot in Ω
        if d_{i,j}(z) < 0 then exit Ω;
        else if d_{i,j}(z) ≥ 0 then track the target;
        end if
      end for
    end if
  else if flag = 1
    for each robot in Ω
      update w_{i,j}(z) , λ_{i,j,k}(z) according to (11), (15);
    end for
    for each robot in Ω
      compute d_{i,j}(z) according to (1);
    end for
    compute V_i according to (2);
    if V_i ≥ σ then flag = 1;
      for each robot in Ω
        if d_{i,j}(z) < 0 then exit Ω; initialize w_{i,j}(z), λ_{i,j,k}(z);
      end for
      for each robot not in Ω
        compute d_{i,j}(z) according to (1);
        if d_{i,j}(z) ≥ 0 then join Ω, track the target;
      end for
    else if V_i < σ then flag = 0;initialize Ω, w_{i,j}(z), λ_{i,j,k}(z);
    end if
  end if
end for
```

IV. EXPERIMENT

A multi-robot system with three robots is used to verify the effectiveness of the proposed method. Each robot uses HOG [13] and H-S to represent the shape and color information of the target, respectively. The HOG feature is obtained by calculating the gradient direction and intensity distribution of the local area of the image. Even when the specific position of the edge is unknown, the distribution of gradient direction can well represent the contour of the target. On the other hand, the hue and saturation in HSV color space are not affected by the brightness component, and they are robust to the illumination changing. The statistical results of the hue and saturation distribution in an image is located in a series of predefined bin to get H-S histogram, which is used for classification recognition.

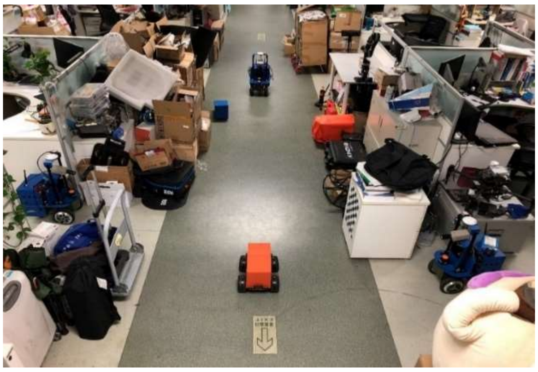


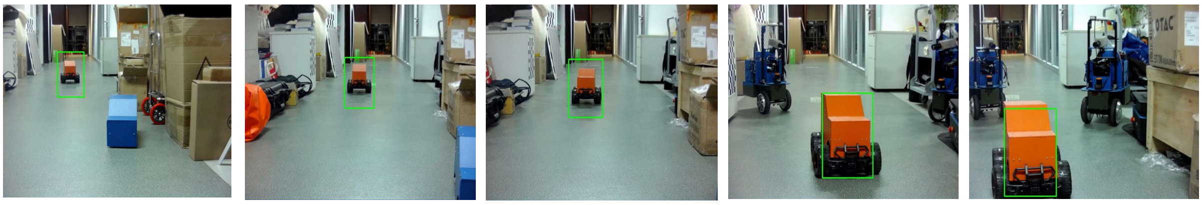
Fig. 2. Three robots recognize and track the target car in the complex scene.

The three robots recognize and track the target car in the complex scene where there are many similar objects and shielding, as shown in Fig. 2. The visual perception process of the entire system is shown in Fig. 3. Fig. 4 gives the voting weights of the three robots to reflect the cooperation of the robots in the process of visual perception. The red, blue and green curves correspond to the voting weights of Nos. 1-3 robots, respectively.

As can be seen from Figs. 3 and 4, initially, only No. 1 robot and No. 2 robot recognize the target, and the target is tracked through the voting decision of these two robots. The No. 2 robot is near to the target, which leads to a higher voting weight than that of No. 1 robot. This situation lasts until the No. 3 robot observes the target at frame 21, as shown in frame 21 of Fig. 3(c). Since the recognition information is consistent with the system, No. 3 robot begins to participate in the voting decision. With the movement of the target, for No. 2 robot, the target is partial blocked, which brings a severe influence of the tracking confidence, and the corresponding voting weight is dramatically reduced. When the target disappears in the field of view of No. 2 robot, its voting weight becomes 0. From frame 40 to frame 60, it can be seen that only No. 1 robot and No. 3 robot are involved in voting. Starting from frame 61, No. 2 robot recognizes the target again, and joins the system decision. Since then, these three robots can identify and track the target until the end of the task.

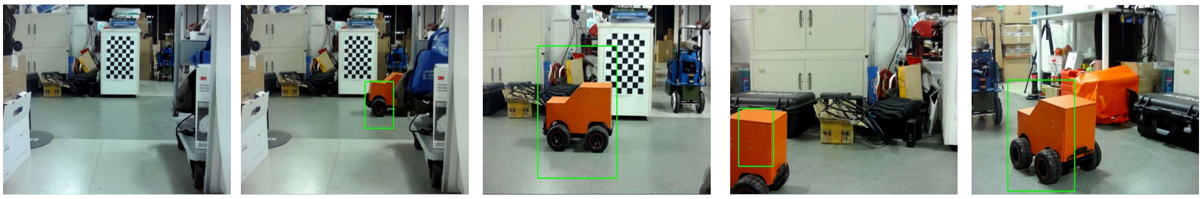
In the whole process of visual perception, objects of similar color or shape sometimes appear in the visual field of three robots. However, by voting decisions and the measurement fusion of features, the interference from similar objects is avoided. The curves of feature measuring weights of the three robots are shown in Fig. 5, where the red and blue curves corresponds to the measuring weight of HOG feature and H-S feature, respectively.

It is seen that the shape information of the target dominates the visual perception of No. 1 robot at the beginning, whereas the color feature began to play a major role after the distance to the target is small. For No. 2 robot, the measuring weights of the two features is close to each other at the initial stage, and the shape feature becomes the dominant one after the target is re-recognized. For No. 3 robot, when the target is found, the target



(a)

(b)



(c)

Fig. 3. The visual perception of the robotic system with 5 selected frames (frame 1, frame 21, frame 43, frame 64, frame 87). (a)-(c) correspond to the Nos. 1-3 robots, respectively.

is partially shaded, and the shape feature provides the main contribution to visual perception. After the target is completely detected, color feature have more weight than that of shape feature.

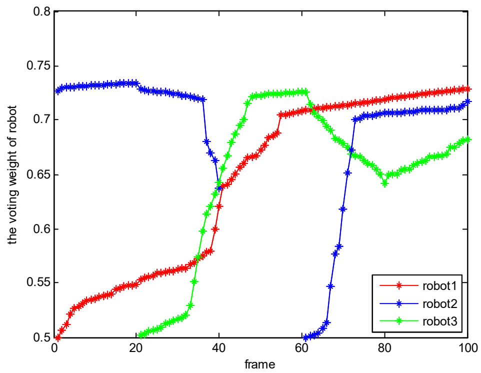
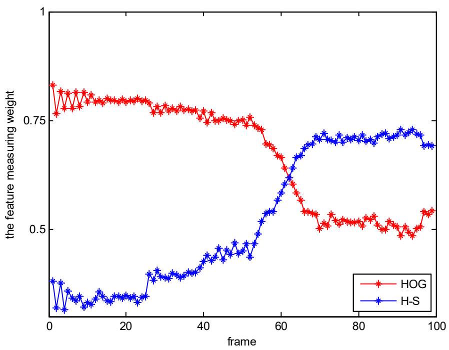
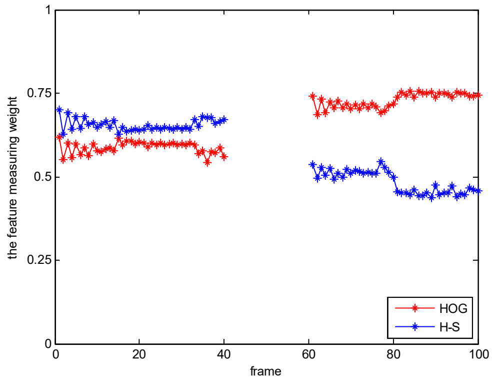


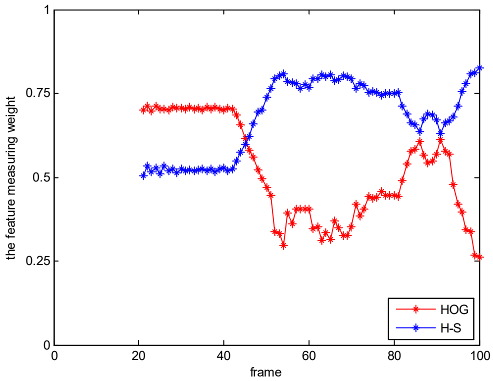
Fig. 4. The voting weights of the three robots.



(a)



(b)



(c)

Fig. 5. The feature measuring weights of the three robots. (a)-(c) are corresponding to Nos.1-3 robots, respectively.

From the above experimental analysis, the voting weights of robots and the measuring weights of features can adaptively update according to the actual environmental information, making the robots and features with better accuracy to play a leading role in visual perception. This solution can effectively improve the quality of recognition and tracking, which is beneficial to the whole robotic system.

V. Conclusion

In this paper, a two-level adaptive target recognition and tracking method based on vision for multi-robot system is proposed. The perception results of multi-robot system are obtained by weighted voting of robots, where the visual perception of each robot is implemented by combining the measuring results of different features. Specifically, the voting weight of each robot and the measuring weight of each feature can be updated adaptively. The proposed method realizes the collaboration of multi-robot system in target recognition and tracking, and the features and robots with higher accuracy play leading roles in visual perception, which enhances the environmental adaptability of the whole system. The validity of the proposed method is verified by experiment.

Acknowledgment

This work was supported in part by the Open Foundation of the State Key Laboratory of Management and Control for Complex Systems, CASIA under Grant 20190106, in part by the National Natural Science Foundation of China under Grants 61633017, 61633020, in part by the 21st Research Institute of China Electronics Technology Group Corporation under Grant ZX-JH-19-0241, and in part by the Beijing Advanced Innovation Center for Intelligent Robots and Systems under Grant 2018IRS21.

References

[1] Hanebeck U D. Template matching using fast normalized cross correlation. *Proceeding of Spie on Optical Pattern Recognition XII*, 2001, 4387:95-102.

[2] Choi M S, Kim W Y. A novel two stage template matching method for rotation and illumination invariance. *Pattern Recognition*, 2002, 35(1):119-129.

[3] Lin Z, Davis L S, Doermann D S, et al. Hierarchical Part-Template Matching for Human Detection and Segmentation. *IEEE 11th International Conference on Computer Vision*, 2007.

[4] Lin Z, Davis L S. Shape-based human detection and segmentation via hierarchical part-template matching. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2010, 32(4):604-18.

[5] Ren H, Zhong Q. A new image segmentation method based on HSI color space for biped soccer robot. *IEEE International Symposium on IT in Medicine and Education*, 2008, 1058-1061

[6] Schmits T, Visser A. An omnidirectional camera simulation for the USARSim world. *Lecture Notes in Computer Science*, 2009, 296-307

[7] Sonka M, Hlavac V, Boyle R. Image Processing Analysis and Machine Vision. Thomson Learning, 2008.

[8] Gotlieb C, Kreyszig H. Texture descriptors based on co-occurrence matrices. *Computer Vision, Graphics, and Image Processing*, 1990, 51(1): 70-86.

[9] Calonder M, Lepetit V, et al. BRIEF: Computing a Local Binary Descriptor Very Fast. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(7): 1281-1298.

[10] Ma W, Manjunath B. A comparision of wavelet transform features for texture image annotation. *International Conference on Image Processing*, 1995, 256-259.

[11] Grimson W, Huttenlocher D. On the Sensitivity of the Hough Transform for Object Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, 12(3): 255-274.

[12] Zhang G, Cheng B, Feng R, et al. Real-time driver eye detection method using Support Vector Machine with Hu invariant moments. *IEEE International Conference on Machine Learning and Cybernetics*, 2008, 2999-3004.

[13] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005: 886-893.

[14] Grossman D, Domingos P. Learning Bayesian network classifiers by maximizing conditional likelihood. *International Conference on Machine Learning*, *ACM*, 2004.

[15] Cortes C, Vapnik V. Support-vector networks. *Machine Learning*, 1995, 20(3):273-297.

[16] Sinopoli B, Schenato L, Franceschetti M, et al. Franceschetti, et al. Kalman filtering with intermittent observations. *IEEE Transactions on Automatic Control*, 2004, 49(9): 1453-1464.

[17] Simon D, Chia T. Kalman filtering with state equality constraints. *IEEE Transactions on Aerospace and Electronic Systems*, 2002, 38(1): 128-136.

[18] Haykin S, Arasaratnam I. Cubature kalman filters. *IEEE Transactions on Automatic Control*, 2009, 54(6): 1254-1269.

[19] Gustafsson F, Gunnarsson F, Bergman N, et al. Particle filters for positioning, navigation, and tracking. *IEEE Transactions on Signal Processing*, 2002, 50(2): 425-437.

[20] Hue C, Le Cadre J, Perez P. Tracking multiple objects with particle filtering. *IEEE Transactions on Aerospace and Electronic Systems*, 2002, 38(3): 791-812.

[21] Zhou H, Yuan Y, Shi C. Object tracking using SIFT features and mean shift. *Computer Vision and Image Understanding*, 2009, 113(3):345-352.

[22] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(7): 1409-1422.

[23] Kalal Z, Mikolajczyk K, Matas J. Face-TLD: Tracking-learning-detection applied to faces. *IEEE International Conference on Image Processing*, 2010, 3789-3792.

[24] Hare S, Golodetz S, Saffari A, et al. Struck: Structured Output Tracking with Kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(10): 2096-2109.

[25] Henriques J, Caseiro R, Martins P, et al. High-speed Tracking with Kernelized Correlation Filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 583-596.