Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2022.030102
ISSN 2790-0983 Vol. 3, Issue 1: 10-17

Comparison between Generative Model and Discriminant Model Based on Support Vector Nachine Algorithm

Shuang Guo*

Hebei Chemical & Pharmaceutical College, Shijiazhuang, China

**corresponding author*

Keywords: Support Vector Machine, Machine Learning, Head Count, Target Tracking

Abstract: Video target tracking is an important research direction in the field of computer vision, and plays an important role in artificial intelligence and big data applications. This paper mainly studies the comparison between generative model and discriminant model based on support vector machine (SVM) algorithm. This paper constructs MSA model and Structured SVM algorithm model according to the generation model and discriminant model classical target tracking algorithm respectively. This paper mainly uses OTB and TC-128 data sets as test data sets to conduct performance comparison experiments on target tracking algorithms of generative model and discriminant model. It can be seen from the experimental results that the Structured SVM model is superior to the MSA model in all aspects.

1. Introduction

The essence of target tracking is to process the continuous video frame images. By tracking and positioning the targets in each frame image, the change information of the targets in the continuous video frames can be obtained [1-2]. Feature extraction is a mathematical description process of the object of interest in the video image from the image to the number, and then build the object description model according to the obtained object features. Target feature extraction and observation model establishment are two important steps of target tracking technology. According to the different methods used between them, tracking methods are generally divided into two categories: generation method and discrimination method [3-4]. The description of the target appearance and the establishment of the target model are important steps in the generation method, which will affect the efficiency and accuracy of the tracker. The model description methods are different according to the difficulty of the target [5]. Discriminant method tries to distinguish the target and background in the image, and directly studies the prediction model without considering the sample generation model. In computer vision, the main idea of these methods is to use appropriate feature extraction methods to represent the objects in video images, and then select appropriate machine learning methods to classify the extracted features and select the final target location according to the classification results. Because the learning process is not only about the object appearance, but also introduces the background information of the video, the algorithm performance is more robust than the generative method [6]. Its advantage is that their speed, accuracy and robustness are completely superior to the generation method.

The target tracking algorithm based on the generation model method focuses on how the data is generated. First, a target generation model that can represent the target is learned from the target samples given in the initial frame; Then, the target is searched globally in each subsequent video image frame, and the searched signal is classified; Finally, the classification signal and the target generation model are matched for similarity, and the signal with the highest matching degree is used as the target position in the current frame. For example, in the current frame, it is learned that the appearance of the target is 90% red and 10% blue. In the subsequent frames, when the tracking algorithm searches the whole image, it will use the image block area that best matches this proportion as the predicted target position. Common tracking algorithms include normalized cross-correlation matching target tracking algorithm, KLT optical flow method, Kalman filter, mean shift method, etc. [7-8]. The discriminant tracking box considers tracking as a frame by frame detection problem. Select the target frame manually from the first frame, and then model the target. During tracking, train a classifier to distinguish the target in the foreground from the background information in the video sequence. If the classifier considers the foreground target information of the sample, it will continue the following tracking steps, that is, give the target location and update the model. If the classifier considers the samples as background information, it will resample and update the samples and adjust the weights. With the development of correlation technology, the discrimination algorithm has gradually formed two branches. One is the regression discrimination method under the framework of correlation filtering. After feature extraction, the method regresses the target to the Gaussian distribution form of the target model, and determines the target location according to the response map of the two-dimensional Gaussian distribution. One is the classification discrimination method under the depth learning method, which sends the extracted image features to the trained classifier and tracks the target according to the confidence level of the classifier output [9-10].

Although the video target tracking technology based on discriminant model has become the mainstream algorithm at present, the tracking algorithms under different frameworks have the challenges of tracking accuracy under different attributes, and tracking speed under actual scene applications. Therefore, the comparative study of video target tracking technology based on generative and discriminant models in this paper has important theoretical significance and practical application value.

2. Generative and Discriminant Target Tracking Algorithms

2.1. Generative Target Tracking Algorithm

Mean Shift algorithm (MSA) is a nonparametric kernel density estimation method. It does not need to rely on any prior information, but only uses the sampling points in the feature space to calculate the density function of the target area. Its essence is an iterative process. It can find the local extreme value in the density distribution of a group of data, and then iterate to the densest data. The final effect of the iteration process of MSA is that starting from the starting point, the local extreme value obtained through each iteration reaches the point with the densest feature points step by step, which is related to the set threshold [11-12]. When the MSA is applied to target tracking, the kernel function and weight coefficient need to be referenced to improve the accuracy and tracking ability of the algorithm. The specific tracking process mainly includes two steps: first, calculate the eigenvalue probabilities of all pixels in the target image block and candidate target image block to obtain the descriptions of the two models. Then, the similarity measurement method is used to calculate the similarity of the two models. According to the calculation results, the candidate region with the highest similarity to the target model is selected and used as the Mean Shift vector of the current frame target region. This vector points from the initial position of the target to the position of the current frame target, so the algorithm will eventually point to the actual position of the current frame target, thus realizing target tracking [13].

The specific tracking steps of MSA are as follows: manually select the target to be tracked, and record the coordinates, length and width of the rectangular box; Calculate the probability density of the target model, the estimated position of the target and the size of the kernel window; Use the target information obtained from the previous frame to estimate the position of the target in the current frame, and calculate the candidate target model; Calculate the position of the target according to the kernel function and weight coefficient and output [14].

The advantage of MSA is that it is simple to calculate and can achieve more accurate tracking when the target has been selected; In addition, because MSA uses color histogram model, it has good tracking effect even if the edge of the target is occluded to a certain extent. However, MSA also has some defects, such as: it cannot update the template in real time during the tracking process, it cannot solve the problem of tracking failure caused by the change of target scale, and the tracking accuracy will decrease when the target moves rapidly [15].

2.2. Discriminant Target Tracking Algorithm

The SVM method uses the structured SVM (SVM) to directly predict the best candidate as the target, and does not classify the target and background. This method simplifies the process that standard SVM needs to label samples to train classifiers, and shows good performance in tracking [16-17]. Based on the analysis of its tracking mechanism, y is defined as a tracking box in this paper, so the target tracking problem can be expressed as predicting an optimal tracking box y by using SVM F (x, y, w) from a given input image x ∈ X ˆ ∈ Y, the formula is expressed as:

$$\hat{y} = \arg\max_{y \in Y} F(x, y, w) \tag{1}$$

Since the offset b in the standard SVM can be constant in practice, the SVM model can represent:

$$F(x, y, w) = w^T \Phi(x, y) \tag{2}$$

Where w is the model parameter of SVM, Φ (x, y) is the characteristic mapping function.

Unlike standard SVM, which assigns sample xi a tag value of 1 or - 1 (where i=1,2,..., n), structured SVM directly takes the box yi corresponding to sample xi as a tag, that is, a structured tag. Therefore, for the training sample set {(x1, y1,..., (xn, yn))}, the model parameter w in equation (2) can be obtained by solving the following quadratic programming problem.

$$\begin{aligned} &\min_{y \in Y} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} \zeta_i \\ &s.t.\ \forall i, \forall y \neq y_i : w^T \Psi_i(y) \geq \Delta(y_i, y) - \zeta_i \\ &\forall i : \zeta_i \geq 0 \end{aligned} \tag{3}$$

Among them, Ψ i(y)= Φ (xi,yi)- Φ (xi, y) represents the difference between the eigenvectors of samples (xu, yi) and (xi, y). Δ (yi, y) is the loss function, defined as:

$$\Delta(y_i, y) = 1 - O_{pi}(y_i, y) \tag{4}$$

Wherein, Opi (yi, y) represents the overlap rate between frames yi and y, and pi represents the center position of the current frame target.

The key of the target tracking algorithm based on SVM is to optimize the solution model expressed in equation (3). According to the analysis of the kernel method, when the nonlinear kernel is used, the support vector opportunity will map the original sample data to a higher dimensional space, which will inevitably bring a huge amount of computation. Although the kernel technique can be used for explicit calculation, there are still problems such as complex optimization, which makes SVM not suitable for large-scale data training. To solve this problem, scholars use linear kernel instead of nonlinear kernel, and use a binary representation method to expand the dimensions of the original features to obtain equivalent nonlinear classification effect. However, since the original features cannot fully represent the information of the target image, the feature points that can be used to distinguish the target from the background are not obvious, resulting in limited improvement in tracking performance [18].

In addition, the method based on SVM adopts the method of directly updating the SVM model after the result prediction is completed, which makes the model vulnerable to occlusion and fails to track. The main reason is that when occlusion occurs, the continuous updating method will incorrectly update the occluded object as the target. Short term occlusion has little impact on the algorithm, but long term accumulation will lead to model drift, which will lead to tracking failure. Therefore, if we can quickly detect whether the target is occluded before updating, and stop updating the model when occlusion is detected, we can effectively avoid tracking failure.

3. Comparison Experiment of Generating and Discriminant Target Tracking Algorithms

3.1. Experimental Data Set

This paper mainly selects OTB and TC-128 datasets as test datasets, and compares and analyzes different algorithms on this basis.

OTB can be divided into three versions: OTB-2013, OTB-2015 and OTB-50. OTB-50 and OTB-2013 both contain tracking sequences of 50 specific objects. OTB-2015 is a combination of OTB-2013 and OTB-50, including 100 specific target tracking sequences. OTB divides target tracking into 11 types of visual challenge attributes, and labels the challenge attributes for each video sequence. Each video sequence has more than one attribute tag corresponding to it, which can analyze the tracking ability of the algorithm in different challenge attributes.

The TC-128 includes 128 video sequences that are manually labeled without repetition. This dataset mainly analyzes and compares the impact of color information on target tracking, so all video sequences are color images. Like OTB, each video sequence contains multiple visual challenge attributes. Among them, 50 video sequences overlap with the OTB. In addition, 78 manually labeled video sequences are added, which increases the diversity and complexity of the target tracking algorithm evaluation dataset on the basis of the OTB dataset.

3.2. Evaluating Indicator

This paper mainly conducts experiments based on OTB and TC-128 datasets, and mainly compares and analyzes the algorithms with distance accuracy and overlapping success rate evaluation indicators.

Distance accuracy. First, calculate the Euclidean distance between the predicted position and the real position of all video frames. The ratio of the number of video frames that meet the threshold condition to the total number of frames is the distance accuracy. The calculation proportion varies with the threshold value.

Overlapping success rate refers to the proportion of video frames whose overlap rate is greater than a given threshold in the total number of frames. The calculated proportion varies with the threshold value. Generally, the threshold value is set as 0.5.

4. Analysis of Experimental Results

4.1. Overlapping Success Rate And Distance Accuracy Rate

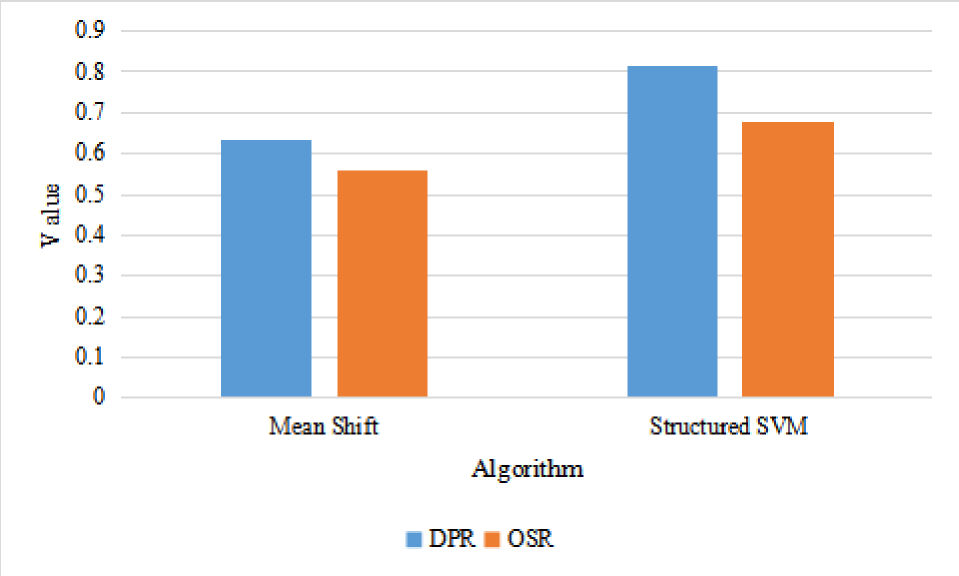

Figure 1. Comparison of distance accuracy and overlapping success rate of algorithms on OTB2015 dataset

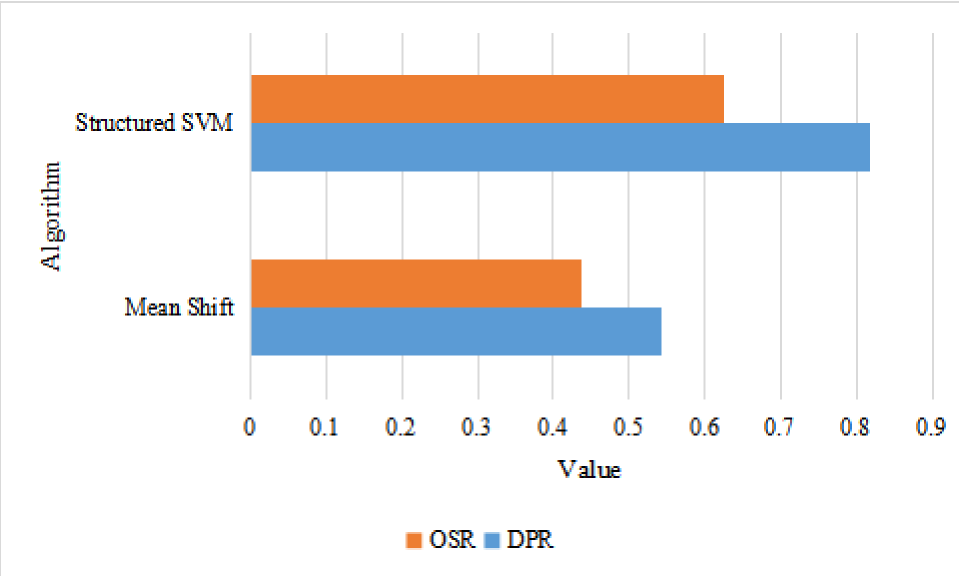

Figure 2. Algorithm comparison on TC-128 dataset

As shown in Figure 1 and Figure 2, the overlapping success rate of Structured SVM is significantly higher than that of Mean Shift. The reason is that the algorithm improves the success rate of the algorithm in terms of filter model optimization from the use of generative model to the use of discriminant model; In the aspect of object appearance representation, from the use of traditional manual features to the use of hierarchical depth features, the discrimination ability of the algorithm in different scenes is improved. The distance accuracy obtained by the algorithm on different data sets shows a trend of increasing chapter by chapter. The TC-128 dataset is more challenging than some video sequences in the OTB2015 dataset, and the tracking effect of the algorithm in the TC-128 dataset is slightly worse. Compared with Mean Shift, Structured SVM has been greatly improved, highlighting the advantages of discriminant correlation filter.

Table 1. Distance accuracy and overlap success rate of the algorithm

	OTB2015		TC-128	
Algorithm	DPR	OSR	DPR	OSR
Mean Shift	0.631	0.557	0.543	0.438
Structured SVM	0.814	0.675	0.817	0.624

As shown in Table 1, under the test data set, the distance accuracy and overlapping success rate of algorithm 1 are the lowest, but in practical applications, when the number of samples is large, it can converge to the target more quickly, and it can solve the target tracking problem in scenarios with hidden variables.

4.2. Tracking Speed

Table 2. Tracking performance of the algorithm

	Success rate	Tracking speed
Mean Shift	0.517	32
Structured SVM	0.689	35

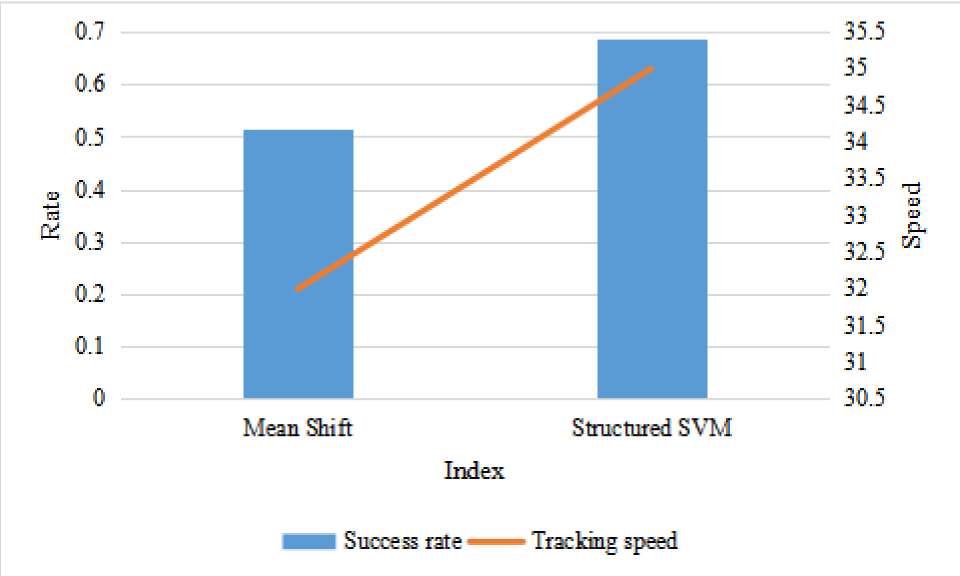

Figure 3. Comparison of algorithm tracking speed and success rate

As shown in Table 2 and Figure 3, Structured SVM is faster than Mean Shift because it uses the principle of correlation filter for tracking, generates a large number of training samples using the cyclic matrix, and improves the processing speed by using the fast Fourier transform in the frequency domain. Structured SVM has a higher success rate than Mean Shift.

5. Conclusion

Video target tracking is one of the important research topics in the field of computer vision and pattern recognition. In practical application scenes, targets are often interfered by such factors as illumination changes, rapid movement, occlusion, low resolution, rotation, deformation, low illumination and motion blur. It is difficult to design tracking algorithms that can maintain high robustness and certain real-time performance in various complex scenes. Based on the generative model and discriminant model, this paper discusses and designs the target tracking algorithm from the mathematical modeling method of the model, modeling the appearance of the target using manual features and depth features, relocating and re tracking strategy, and updating the target model, aiming to enable the algorithm to achieve the balance between accuracy, robustness and real-time in complex scenes.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Kamkar S, Ghezloo F, Moghaddam H A, et al. Multiple-target tracking in human and machine vision. PLoS computational biology, 2020, 16(4): e1007698. https://doi.org/10.1371/journal.pcbi.1007698

[2] Dendorfer P, Osep A, Milan A, et al. Motchallenge: A benchmark for single-camera multiple target tracking. International Journal of Computer Vision, 2021, 129(4): 845-881. https://doi.org/10.1007/s11263-020-01393-0

[3] Mahmoudi N, Ahadi S M, Rahmati M. Multi-target tracking using CNN-based features: CNNMTT. Multimedia Tools and Applications, 2019, 78(6): 7077-7096. https://doi.org/10.1007/s11042-018-6467-6

[4] Penin B, Giordano P R, Chaumette F. Vision-based reactive planning for aggressive target tracking while avoiding collisions and occlusions. IEEE Robotics and Automation Letters, 2018, 3(4): 3725-3732. https://doi.org/10.1109/LRA.2018.2856526

[5] Tian Y, Dehghan A, Shah M. On detection, data association and segmentation for multi-target tracking. IEEE transactions on pattern analysis and machine intelligence, 2018, 41(9): 2146-2160. https://doi.org/10.1109/TPAMI.2018.2849374

[6] Briñón-Arranz L, Seuret A, Pascoal A. Circular formation control for cooperative target tracking with limited information. Journal of the Franklin Institute, 2019, 356(4): 1771-1788. https://doi.org/10.1016/j.jfranklin.2018.12.011

[7] Aryankia K, Selmic R R. Neuro-adaptive formation control and target tracking for nonlinear multi-agent systems with time-delay. IEEE Control Systems Letters, 2020, 5(3): 791-796. https://doi.org/10.1109/LCSYS.2020.3006187

[8] Daniyan A, Lambotharan S, Deligiannis A, et al. Bayesian multiple extended target tracking using labeled random finite sets and splines. IEEE Transactions on Signal Processing, 2018, 66(22): 6076-6091. https://doi.org/10.1109/TSP.2018.2873537

[9] Upadhyay J, Rawat A, Deb D. Multiple Drone Navigation and Formation Using Selective Target Tracking-Based Computer Vision. Electronics, 2021, 10(17): 2125. https://doi.org/10.3390/electronics10172125

[10] Tesfaye Y T, Zemene E, Prati A, et al. Multi-target tracking in multiple non-overlapping cameras using fast-constrained dominant sets. International Journal of Computer Vision, 2019, 127(9): 1303-1320. https://doi.org/10.1007/s11263-019-01180-6

[11] Chandrakala S, Jayalakshmi S L. Generative model driven representation learning in a hybrid framework for environmental audio scene and sound event recognition. IEEE Transactions on Multimedia, 2019, 22(1): 3-14. https://doi.org/10.1109/TMM.2019.2925956

[12] Yashchenko A V, Potapov A S, Rodionov S A, et al. Application of generative deep learning models for approximation of image distribution density. Journal of Optical Technology, 2019, 86(12): 769-773. https://doi.org/10.1364/JOT.86.000769

[13] Najar F, Bourouis S, Bouguila N, et al. A new hybrid discriminative/generative model using the full-covariance multivariate generalized Gaussian mixture models. Soft Computing, 2020, 24(14): 10611-10628. https://doi.org/10.1007/s00500-019-04567-2

[14] Ding Z, Shao M, Fu Y. Generative zero-shot learning via low-rank embedded semantic dictionary. IEEE transactions on pattern analysis and machine intelligence, 2018, 41(12): 2861-2874. https://doi.org/10.1109/TPAMI.2018.2867870

[15] Ko T, Kim H. Fault classification in high-dimensional complex processes using semi-supervised deep convolutional generative models. IEEE Transactions on Industrial Informatics, 2019, 16(4): 2868-2877. https://doi.org/10.1109/TII.2019.2941486

[16] Buscombe D, Grams P E. Probabilistic substrate classification with multispectral acoustic backscatter: A comparison of discriminative and generative models. Geosciences, 2018, 8(11): 395. https://doi.org/10.3390/geosciences8110395

[17] Kim J, Oh S, Kwon O W, et al. Multi-turn chatbot based on query-context attentions and dual wasserstein generative adversarial networks. Applied Sciences, 2019, 9(18): 3908. https://doi.org/10.3390/app9183908

[18] Lopez R, Gayoso A, Yosef N. Enhancing scientific discoveries in molecular biology with deep generative models. Molecular Systems Biology, 2020, 16(9): e9198. https://doi.org/10.15252/msb.20199198