Multi-Object Tracking Algorithm Based on Spatial Constraints

Yuanhang Cheng
College of Information Engineering
Shenyang University
Shenyang 110044, China
cyh518@163.com

Jing Wang
College of Information Engineering
Shenyang University
Shenyang 110044, China
sdwj91@163.com

***Abstract*—For multi-object tracking in complex environments, this paper presents an improved tracking algorithm based on spatial constraints. With basic framework of Dalal-Triggs detector (which uses HOG features to describe the image blocks and an SVM to predict the existence of objects), we use a graph structure model to constrain the spatial relationship between the multiple objects that are being tracked. Experiments show that spatial constraints among the objects make greatly improved performance of the tracker in multiple objects tracking. In the video of camera significant movement, fast-moving objects, objects change appearance and occlusion, the tracker performs well.**

Keywords—Histogram of oriented gradients;multi-object tracking;spatial constraints;online learning

I. INTRODUCTION

Object tracking is the most popular research direction computer vision, and also has a wide range of applications in both military and civilian fields. Because of the complexity of actual application scenario, the objects tracking will be affected by posture diversity, illumination changes, occlusion, motion blur and such factors. So design a tracking system with high-precision and strong robustness is a very challenging job.

Currently, the framework and the thought of objects tracking have two main types: one is Generative Method, the other is Discriminative Method [1]. Generative Methods model with some features of objects and estimate the joint probability distribution. The distribution of the data can be reflected from the perspective of statistical and the similarity of similar data itself can be represented. Finding the region of the closest model in the range of the objects may be presented to achieve tracking. Its essence is to view the problem of tracking as a problem of searching and matching. Discriminative Methods, which basic idea of tracking algorithm is that it will transform object tracking problem into a classification problem (mostly binary classification problem) to deal with. Any part of foreground (i.e., the objects) is a category, background (all content of non-objects location) is another, which used to training the classifier and looking for the right features and appropriate classification method to distinguish foreground from background region. Discriminative capability between foreground and background owned by the algorithm thanks to the introduction that learning of the background area. For tracking moving objects in complex environments, the inhibitory effect of the drift is better. These two kinds of objects tracking methods can be represented by the basic framework that as shown in Fig.1.

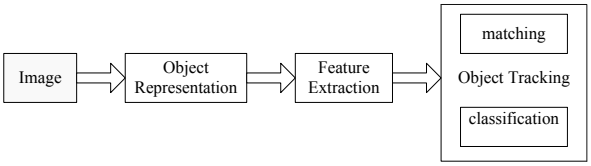

Fig. 1. Basic framework of object tracking

II. FEATURE EXTRACTION

Feature extraction is an important part of the objects tracking, which affects the performance of the entire system directly. In this paper, HOG features as the object features that have been successfully applied in the field of objects detection in recent years. HOG (Histogram of Oriented Gradient) feature is proposed by Navneet Dalal and Bill Triggs [2], which used as the feature set of pedestrian detection to perform pedestrians detecting identification. HOG is widely used in objects detection due to the nature of HOG features with geometric constant distance and illumination change direct adaption. In gesture recognition, face recognition, and pedestrian detection, vehicle detection and other fields, excellent performance are showed by use of HOG features.

Herein, steps of HOG feature calculation are:

1) The input color image is converted into grayscale image.

2) Gamma correction method is adopted to accomplish the standardization (normalized) of color space of the input image. The purpose is to adjust the contrast of the image, reduce the impact caused by change of image local shadows and illumination, and at the same time suppress noise interference.

3) First-order differential is used to calculate the gradient. Mainly in order to capture the contour information, at the same time, further weaken the interference of light.

4) The gradient shadowed into gradient direction of cells. The purpose is to provide a code for local image region.

5) All cells are normalized in the block. Normalization can be further compressed on the illumination, shadow and edge. Usually, each cell shared by a number of different blocks, but the normalization is based on the different blocks, so the calculation results is also different. Therefore, features of a cell will be appeared several times in the final vector with different results. After block descriptor normalized is called HOG descriptor [3].

6) Collected HOG features in all blocks of detection space. This step is to collect the HOG features of all overlapping blocks in the detection window, and combine them into the final feature vectors for classification.

In this paper, parameters are as follows: no gamma-corrected RGB color space; the gradient operator is [-1, 0, 1] and no smoothing; the gradient direction discretized (voted) into 9 bins between 0-180°; block size is 16×16, cell size of 8×8; L2-norm block normalization; block move step is 8 pixels.

III. GRAPH STRUCTURE MODEL

Image has the strong structure, namely the strong correlation between pixels, and these correlation contain important information about the image structure. But the structural information can't be described by images of gray, color, texture and so on. However, the graph as a tool for describing the data can keep the relationship between structures and regions, and therefore it is a very important and effective representation for information of the structure.

We define a graph $G=(V,E)$ for all objects $i \in V$, and want to track with the edge $(i,j) \in E$ between objects. The edges in the graph model can be viewed as springs that on behalf of spatial constraints between the tracked objects. Graph structure models represent spatial relations between each object and the springs can be stretched or compressed when the object being considered is deformed. This article uses the minimum spanning tree modeling structure constraints [4]. It is a search for all possible collection of tree models with fully connected, make to minimize $\sum_{(i,j)\in E} \left\| \mathrm{X}_i - \mathrm{X}_j \right\|^2$, where X_i and X_j is the locations of objects i and j in the first frame. In a connected undirected weighted graph $G=(V,E)$, the weight $w(i,j)$ of each edge (i,j) is given. A spanning tree is found from the graph G, which has the minimum weights, and is called a minimum spanning tree (MST) [5]. Where, the weight of an edge is the Euclidean distance between two nodes connected by the edge. The minimum spanning tree of a weighted graph is the spanning tree with minimum weights in the graph. We use the online method learning spring parameters during tracking.

IV. ONLINE LEARNING

Due to the dramatic changes in the environment surrounding the objects, test samples that little similar to train samples can't be classified effectively by offline learning classifiers and then lead to performance degradation and drift, so this paper adopts the SVM online learning to design the classifier.

First, the positive and negative samples are sampled according to the objects positions in the first frame, feature vectors are extracted from two types of samples for SVM classifier training, the support vectors in the feature vector samples are found by SVM learning from samples, so the optimal separating hyperplane is established. Then the detection result of moving objects in current video are viewed as the test sample, and their feature vector are extracted and input SVM, finally the output of the classifier is the tracking result of the current frame. We use the configuration of tracked objects in the previous frame as a positive sample to update our model, the appearance models of all objects and the structural constraints between them are trained in an online structured SVM framework. Our classifier uses this method in online learning and training to complete the objects tracking.

We use the passive-aggressive algorithm to complete the parameters update. Passive-Aggressive (PA) training algorithm [6] is similar to general online training algorithm. However, the difference is aggressive parameters added when amending the weight. Don't need to correct weight when the prediction is correct, this algorithm in the passive state; however, when the prediction is error, this algorithm in the active state. Therefore, it is called passive-aggressive algorithm. For the samples of linear separable and linear inseparable, on the basis of the original objective function $\begin{cases} w_{t+1} = \arg\min\limits_{w\in R^n} \frac{1}{2}\|w-w_t\| \\ s.t.\ \ l\left(w;\left(x_{t,}y_t\right)\right)=0 \end{cases}$, two times amendments for the objective function and add slack variables ξ ,get PA- Ⅰ : $\begin{cases} w_{t+1} = \arg\min\limits_{w\in R^n} \frac{1}{2}\|w-w_t\| + C\xi \\ s.t.\ \ l\left(w;\left(x_{t,}y_t\right)\right)\le \xi, \xi \ge 0 \end{cases}$, add slack variables ξ^2 , get PA- Ⅱ : $\begin{cases} w_{t+1} = \arg\min\limits_{w\in R^n} \frac{1}{2}\|w-w_t\| + C\xi^2 \\ s.t.\ \ l\left(w;\left(x_{t,}y_t\right)\right)\le \xi \end{cases}$ .Three different aggressive parameters, which $\tau_t = \frac{l_t}{\|x_t\|^2}$, $\tau_t = \min\left\{C, \frac{l_t}{\|x_t\|^2}\right\}$, $\tau_t = \frac{l_t}{\|x_t\|^2 + \frac{1}{2C}}$ are obtained when weight correction according to the three objective functions[7].

V. EXPERIMENTAL RESULTS AND ANALYSIS

We used a video segment with multiple objects to evaluate the performance of our tracker, and compared with traditional HOG + SVM tracker without structural constraints between objects. We based on two indicators to show performance of the tracker: average location error and correct detection rate.

Average location error (ALE) refers to the average distance of the center of the identified bounding box to the center of the ground truth bounding box, the smaller the better .Correct detection rate (CDR) means the percentage of frames for which the overlap between the identified bounding box and the ground truth bounding box is at least 50 percent, the larger the better.

We showed the classification results based on confidence map, and online updated the positive and negative samples and objects edges on the basis of confidence level. Objects tracking problem can be regarded as an "objects / background" binary classification problem. Confidence map in the form of gradation expressed confidence level for each pixel to objects in the candidate region, that is the judgment result of certain feature or group of features on the "objects / background" classification problem [8]. The process of generating confidence map is to determine the location information of the objects and background in the next frame, the confidence map is a binary image, i.e., one part is the objects, the other part is the background.

In flower video, several similar flowers are moving in direction and changing in appearance because of the influence of wind, sometimes partial occlusion each other occurs. In this paper, the tracker structural constraints can be used to distinguish flowers with similar appearances. Tracking results are shown in Fig.2. The first figure is ground-truth annotations of the objects in first frame of the video. Right side of each figure shows the confidence map of the current frame. A perfect tracking can be seen during the entire length of the video (2,249 frames). The comparison of the proposed new method and the traditional algorithm in tracking performance is shown in Table.1.

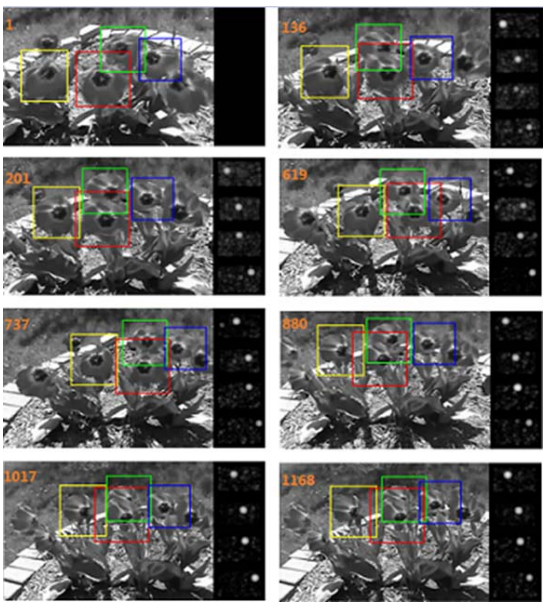

Fig. 2. Tracking results

TABLE I. A COMPARISON OF THE PROPOSED METHOD AND THE TRADITIONAL ALGORITHM

Traditional HOG + SVM algorithm	ALE	50.6
	CDR	0.38
Proposed new method	ALE	9.5
	CDR	0.99

In hunting video, it is very challenging for tracking robustness, because the appearance and relative location of the cheetah and the gazelle change significantly over time. The structural constraints can be used to prevent loss of tracking. Tracking results are shown in Fig.3. A perfect tracking can be seen during the entire length of the video (1,805 frames). The comparison of the proposed new method and the traditional algorithm in tracking performance is shown in Table.2.

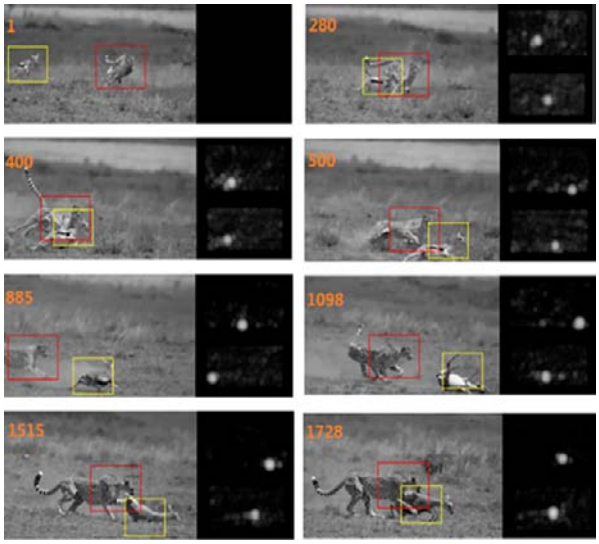

Fig. 3. Tracking results

TABLE II. A COMPARISON OF THE PROPOSED METHOD AND THE TRADITIONAL ALGORITHM

Traditional HOG + SVM algorithm	ALE	171.7
	CDR	0.07
Proposed new method	ALE	19.4
	CDR	0.87

VI. CONCLUSIONS

Based on HOG + SVM tracking framework, this paper put forward an improved multiple objects tracking algorithm based on space constraints. Firstly, HOG features are extracted as the objects features, then a minimum spanning tree model was established between the objects for mutual structural constraints, finally SVM classifier online update for training and classification, thus the objects are tracked effectively. The

experimental results show that our tracker performance is superior to the traditional tracker.

REFERENCES

[1] Wenxin Mao, "Research on feature extraction and object tracking algorithm based on SVM", D. Chong Qing University ,2014.

[2] Navneet Dalal, Bill Triggs, "Histograms of oriented gradients for human detection," CA: CVPR 2005, pp. 886-893,2005.

[3] Rui Su, "The research of human detection based on histogram of orientation gradient,"D. South China University of Technology ,2010.

[4] Lu Zhang, Laurens van der Maaten, "Preserving structure in model-free tracking,"J.IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.36, pp.756-767 , April 2014.

[5] Wenxian Bao, "Algorithms for image matching and its application based on structural feature,"D. An Hui University ,2010.

[6] Crammer K, Dekel O, Singer Y, "Online passive aggressive algorithms,"J. Machine Learning Research,vol. 7, pp.551-585,2006.

[7] Jianwei Lin, Fanglin Shen, Xionglin Luo, "Research on perceptron learning algorithm,"J. Computer Engineering, vol.36, pp.190-192,2010.

[8] Peng Xiao, Miyi Duan, Qi Zhao, Kedai Zhang, "Visual object tracking based on confidence map adaptive fusion,"J. Radio Engineering, vol,43, pp.20-23,2013.