

Hierarchical Association Model Based Target Tracking

Wei Sun, Songhao Zhu, and Chengjian Sun

*School of Automatic, Nanjing University of Post and Telecommunications,
Nanjing, 210046, China*

zhush@njupt.edu.cn

Abstract

This paper proposes a hierarchical association multi-target tracking trajectory generation method. The AdBoost method combined with online discriminant analysis apparent model is first utilized to achieve initial tracking trajectories; then, the Hungarian method is utilized to optimize fragmented and discontinuous trajectories to achieve accurate trajectories fragments; finally, the intelligent extrapolation based on the energy minimization utilized to achieve the final smoother and continuous trajectories. Experimental results on PETS 2009/2010 benchmark demonstrate the effectiveness and efficiency of the proposed scheme.

Keywords: Hierarchical Association Model, Online Apparent Model, Target Tracking.

Introduction

Multi-target tracking has emerged as an active research topic in the past two decades due to its widespread applications in many areas^[1-3]. However, robust tracking of multiple targets remains a huge challenge, especially in the case of mutual occlusion. To solve the issue of false tracking caused by inter-target mutual occlusion, Andriyenko and Schindler propose a method based on the continuous minimum energy^[4], where the tracking informations of an objects are utilized to construct an energy function and the optimal tracking trajectory is obtained by solving the energy function based on the association strategy.

Inspired by the idea in literature [4], this paper proposes a tracking method based on hierarchical correlation combined with apparent model. Specifically, tracking fragments are first obtained according to the information of color, texture and gradient; then, the double threshold method is utilized to obtain discontinuous but correlation tracking fragments; finally, tracking fragments are further correlated with respect to extracted positive and negative training samples to obtain smooth and continuous tracking trajectories. Different from the tracking trajectories achieved based on the continuous minimum energy method, the proposed method takes into account not only the tracking trajectory information of the past and current frames, but also the tracking trajectory information of future frames.

Online Apparent Model Based Initial Correlation

The aim of the online discrimination apparent model (OLDAMs) [5] is to implement the similarity comparison between objects using the AdaBoost algorithm, and to achieve the initial correlation of tracking fragments. The online apparent discrimination model mainly consists of the following four steps: (1) Online sample collection; (2) Apparent feature extraction; (3) Similarity measurement, and (4) Algorithm learning. Among the above four steps, training samples are collected within a time sliding window to adapt to the changes of appearance and background; and apparent feature, as the combination of similarity measure among image descriptors, is utilized to distinguish different targets well. The obtained tracking fragment set is denoted as $T^L = \{T_1^L, T_2^L, \dots, T_i^L, \dots\}$.

Hungarian Algorithm Based Second Correlation

Second Correlation Process

The tracking target set is denoted as $\mathcal{H} = \{r_1, r_2, \dots, r_j, \dots\}$, where $r_j = (x_j, y_j, s_j, t_j, a_j)$, (x_j, y_j) is the location information of target, s_j is the size of the j^{th} tracking target, t_j is the frame number where the j^{th} tracking target appears, and a_j is the color histogram of the j^{th} tracking target. The tracking trajectory set is denoted as $T = \{T_1, T_2, \dots, T_i, \dots\}$, where $T_i = \{r_{ij} | \forall j, t_{ij} < t_{ij+1}\}$ is the i^{th} tracking trajectory.

The second correlation process of tracking trajectory is an iterative process. That is, the correlation set of tracking fragment obtained using the OLDAMs in 2.1 is denoted as $C^L = \{C_1^L, C_2^L, \dots, C_k^L, \dots\}$, where $C_k^L = \{T_{k0}^L, T_{k1}^L, \dots, T_{kk}^L, l_k\}$ is the number of tracking fragment for the k^{th} tracking target. After the second correlation, the tracking fragment set is denoted as $T^M = \{T_1^M, T_2^M, \dots, T_N^M, \dots\}$.

Under the assumption each tracking trajectory obtained using the OLDAMs is independent of each other, the trajectory optimization problem can be considered as a maximum a posteriori problem:

$$L^* = \arg \max_L P(L | T^L) = \arg \max_L P(T^L | L) P(L) = \arg \max_L \prod_{T_i^L} P(T_i^L | L) \prod_{L_k \in L} P(L_k) \quad (1)$$

If the false alarm rate of multi-target tracking follows the Bernoulli distribution, the likelihood probability of tracking trajectory in equation (1) is formulated as:

$$P(T_i^L | L) = \begin{cases} P_+(T_i^L) = \Phi^{|T_i^L|}, & \text{if } \exists L_k \in L, T_i^L \in L_k \\ P_-(T_i^L) = (1 - \Phi)^{|T_i^L|}, & \text{if } \forall L_k \in L, T_i^L \notin L_k \end{cases} \quad (2)$$

where Φ represents the detection accuracy, $|T_k^L|$ denotes the number of tracking objects in the k^{th} tracking trajectory T_k^L , $P_+(T_k^L)$ and $P_-(T_k^L)$ denote the likelihood probability of tracking objects and false targets in T_k^L respectively.

$P(L_k)$, the posterior probability of tracking trajectory in Equation (1) can be

written as the forms of Markov chain:

$$P(L_k) = P_{init}(T_{i_0}^L) P_{link}(T_{i_1}^L | T_{i_0}^L) \cdots P_{link}(T_{i_{k-1}}^L | T_{i_k}^L) P_{term}(T_{i_k}^L) \quad (3)$$

where $P_{init}(T_{i_0}^L)$, $P_{link}(T_{i_{k-1}}^L | T_{i_k}^L)$ and $P_{term}(T_{i_k}^L)$ denote the initial probability, connection probability and termination probability respectively.

Assuming that there is only one tracking trajectory for a tracking target, and can be formulated as:

$$\forall \xi \in \{L, M, H\}, \forall T_i, T_j \in T^\xi, T_i \cap T_j = \emptyset \quad (4)$$

The meaning of Equation (4) is that there is no overlap between any two tracking trajectories, namely there is no mutual occlusion between any two tracking trajectories. Therefore, the correlation of the i^{th} tracking trajectory is unique, and Equation (3) can be further written as:

$$L^* = \arg \max_L \prod_{\forall L_k \in L, T_i^L \notin L_k} P_-(T_i^L) \prod_{L_k \in L} \left(P_{init}(T_{i_0}^L) P_+(T_{i_0}^L) P_{link}(T_{i_1}^L | T_{i_0}^L) \cdots P_{link}(T_{i_{k-1}}^L | T_{i_k}^L) P_+(T_{i_k}^L) P_{term}(T_{i_k}^L) \right) \quad (5)$$

Compared with other methods, the proposed maximum a posteriori probability equation as shown in Equation (5) has the following advantages: (1) Removing false tracking fragments, and eliminating false target tracking effectively; (2) Dealing with the inter-change between initial correlation and termination correlation of tracking trajectory well by using the likelihood probability.

Solving Of Maximum A Posteriori

Suppose there are n tracking trajectories, then the maximum a posterior probability as shown in Equation (5) can be converted to the solving of the problem of standard bipartite graph matching, and the transition matrix is formulated as:

$$C = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} & C_{1(n+1)} & -\infty & \cdots & -\infty \\ C_{21} & C_{22} & \cdots & C_{2n} & -\infty & C_{2(n+2)} & \cdots & -\infty \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} & -\infty & -\infty & \cdots & C_{n(2n)} \\ \hline C_{(n+1)1} & -\infty & \cdots & -\infty & 0 & 0 & \cdots & 0 \\ -\infty & C_{(n+2)2} & \cdots & -\infty & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ -\infty & -\infty & \cdots & C_{(2n)n} & 0 & 0 & \cdots & 0 \end{bmatrix}_{2n \times 2n} \quad (6)$$

It should be noted that $P_+(T_i^L)$ is divided into two components, and then these two components are integrated into two adjacent transfer matrices respectively.

As mentioned before, one of the aim of constructing Equation (5), a maximum a posteriori probability equation, is to reduce the false alarm rate, which is also

reflected in the constitution of diagonal elements of the transfer matrix as shown in Equation (6): (1) Each diagonal element is set to be a logarithmic likelihood function to deal with the false alarm of each tracking fragment, since each false alarm tracking fragment is not associated with other true tracking fragments track, either not considered as an initial trajectory or a termination trajectory; (2) Each diagonal element represents the "self-correlation" of the corresponding tracking fragment.

Solving Of Transition Matrix

The connection probability in Equation (3) is determined by the information of apparent contour, motion state, and time interval of tracking targets:

$$P_{link}(T_j | T_i) = A_a(T_j | T_i) A_m(T_j | T_i) A_t(T_j | T_i) \quad (7)$$

Computing the apparent contour information of tracking targets. To achieve the optimal tracking results, the Kalman filtering method is first adopted to extract the position and size information of tracking targets to estimate the movement speed; then, the random sampling consensus algorithm is utilized to obtain the integrated color histogram a_i^* ; finally, the Gaussian distribution function is adopted to obtain the information of apparent contour:

$$A_a(T_j | T_i) = G(\text{corr}(a_i^*, a_j^*); 0, \sigma_c) \quad (8)$$

where $\text{corr}(\cdot)$ represents the correlation coefficient of the color histogram a_i^* and a_j^* , and σ_c represents the corresponding variance.

Computing the time interval information of tracking targets. The maximum time interval between two correlated tracking trajectories is measured with respect to the rate of missed detection during the time interval between these two correlated tracking trajectories:

$$A_t(T_j | T_i) = \begin{cases} Z_\theta \alpha^{\Delta t - 1 - \omega}, & \text{if } \Delta t \in [1, \theta] \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where α is the missing rate, θ is the threshold of time interval, and Z_θ is the normalization factor. Within the time interval $[1, \theta]$, ω is the number of frames where one tracking target is occluded by other tracking targets, $\Delta t - 1 - \omega$ is the number of frames which are missed.

Minimum Energy Based High Lever Correlation

The tracking trajectories obtained based on the initial and second correlation are short and rough, therefore minimum energy based high correlation is here utilized to achieve longer and smoother tracking trajectories

The energy function is here defined as follows:

$$E(X) = \alpha E_{app}(X) + \beta E_{vis}(X) + \gamma E_{dyn}(X) + \kappa E_{exc} + \lambda E_{mat}(X) + \mu E_{cor}(X) \quad (10)$$

where $\alpha, \beta, \kappa, \gamma, \lambda, \mu$ denote the weight of different model respectively. E_{app}

denotes the energy of appearance model, and E_{det} denotes the energy of visibility model and helps to eliminate the correlation between tracking targets. E_{exc} , E_{dyn} and E_{per} denote the energy of exclusion model, dynamic model, trajectory persistence model respectively, which help to optimize the tracking results. E_{cor} denotes the energy of trajectory correction model, which helps to prevent the over-fitting in the process of iterating to a certain extent.

The optimal tracking trajectory X^* is the optimal solution of the following equation:

$$X^* = \arg \min_{X \in R} E(X) \quad (11)$$

where $E(X^*)$ is consecutive minimum energy in the search space R . The value of $E(X^*)$ depends on the length of video sequences and the number of tracking target, and generally between 10^3 and 10^4 .

Experiments

The performance comparison result of different target tracking methods is shown in table 1, where (1) Recall (Re): The number of targets correctly detected divided by the number of ground truth targets; (2) Precision (Pre): The number of targets correctly detected divided by the number of targets detected using one method; (3) False Alarms Per Frame (FAF): The number of targets detected falsely in each frame; (4) Mostly Tracked Trajectories (MT): The number of trajectories which are successfully tracked for more than 80%; (5) Fragments (Frag): The number of interruptions for the same trajectory during the tracking process; (6) Id Switch (IDS): The number of trajectories whose labels change with other..

Table 1: The performance comparison result of three different target tracking methods.

Databas e	Re	Pre	FAF	MT	Frag	IDs
EM ^[4]	74.1	87.4	0.62	84.2	75	56
AM ^[6]	80.4	86.1	0.99	76.1	37	31
Ours	93.5	90.8	0.55	94.7	35	25

It can be seen from table 1 that the proposed method achieves the best tracking performance in comparison with the other two methods.

Conclusions

To deal with the issue of multi-target tracking, this paper proposes a hierarchical correlation method combine with the apparent model. Given the detection results of tracking targets, the AdBoost algorithm combined with online discriminant apparent model is first utilized to achieve initial tracking trajectories; then, the

Hungarian algorithm is here utilized to optimize fragmented and discontinuous tracking trajectories to achieve stable and accurate trajectories fragments; finally, the intelligent extrapolation based on energy minimization is here utilized to achieve the final smoother and longer tracking trajectories. Experimental results on PETS 2009/2010 benchmark demonstrate the effectiveness and efficiency of the proposed scheme.

Acknowledgments

This work is supported by Postdoctoral Foundation of China under No. 2014M550297, Postdoctoral Foundation of Jiangsu Province under No. 1302087B, Education Reform Research and Practice Program of Jiangsu Province under No. JGZZ13_041, and Graduate Bilingual Teaching-Learning Pilot Program of Pattern Recognition and Intelligent Systems of Jiangsu Province.

References

- D. Mitzel, E.r Horbert, A. Ess, and B. Leibe. Multi-person Tracking with Sparse Detection and Continuous Segmentation. *European Conference on Computer Vision*, 2010: 397-410.
- C. Kuo and R. Nevatia. How does person identity recognition help multi-person tracking? *IEEE Conference on Computer Vision and Pattern Recognition*, 2011: 1217-1224.
- B. Yang and R. Nevatia. Online learned discriminative part-based appearance models for multi-human tracking. *European Conference on Computer Vision*, 2012: 484-498.
- A. Andriyenko and K. Schindler. Multi-target tracking by continuous energy minimization. *IEEE Conference on Computer Vision and Pattern Recognition*, 2011: 1265-1272.
- J. Yeh and T. Hsu. Online Selection of Tracking Features Using AdaBoost. *IEEE Transactions on Circuits and Systems for Video Technology*, 2009, 19(3): 442-446.
- C. Kuo, C. Huang, and R.Nevatia. Multi-target tracking by on-line learned discriminative appearance models. *IEEE Conference on Computer Vision and Pattern Recognition*, 2010: 685-692.