Multi-object Visual Tracking Algorithm Based on Grey Relational Analysis and Generalized Linear Assignment

Yanzhao Su(✉), Aihua Li, Zhigao Cui, Hao Fang, and Tao Wang

502 Faculty, Xi'an Institute of High Technology, Xi'an, Shan Xi, China
syzlhh@163.com, {aihuali,zhigaocui,haofang}@126.com,
taowang@hotmail.com

Abstract. In the view of multi-object tracking in video sequences affected by the issues of similar objects and occlusion in objects, etc., a hierarchy fusion visual tracking algorithm based on gray relational analysis were proposed in this paper. In the algorithm, object trajectory was associated step by step and the video sequences was processed by adding time windows. First, tracklets were provided by a conservative association of the detections. Then, in every time window, combined with the improved grey degree of incidence and moving information, the similarity of two trajectory was calculated. In the end, the optimal association of the tracklets was achieved according to the generalized linear assignment. By comparison with typical algorithms, experimental results show that the algorithm is applicable to multi object tracking in the scenes without reliable appearance characteristic provided with higher tracking accuracy, and adapt to the effect of object occlusion, similar appearance, camera motion and so on.

Keywords: Grey relational analysis · Generalized linear assignment · Multi-object · Visual tracking

1 Introduction

Multi-object visual tracking[1-2] is important for many computer vision applications including intelligent control, human-computer interaction, virtual reality, etc. Compared with single object tracking, multi-object tracking faced more challenges (e.g. unknown number of objects, appearance changes and mutual occlusion)[3].With the development of object detection technology (such as background modeling, pedestrian detection, etc.), most current approaches to multi-target tracking are based on tracking by detection. A complete motion trajectory is formed by linking the detections in every frame with motion and appearance features of objects. Zhang. et al. [4] formulate multi-object tracking as the minimum cost flow problem in networks. The algorithm could reduce the number of tracklets significantly and maintain the integrity of the object trajectory with the global objective function. To calculate the similarity of object detections, both the appearance and motion characteristics were adopted. The metric methods are relatively simple, such as the Euclidean distance, Bhattacharya distance, etc. As the appearance characteristics of the object in the actual mon-

H. Zha et al. (Eds.): CCCV 2015, Part II, CCIS 547, pp. 448–457, 2015.
DOI: 10.1007/978-3-662-48570-5_43

itoring scene were not stable and discriminative, many scholars use motion features to achieve mul-ti-object tracking. Wen et al. [5] measure the motion feature similarity of objects by adopting the forward and backward prediction information with the assumption of uniform motion. However, the accuracy of tracking results relied on the movement pattern of the object. Dicle et al. [6] construct a Hankel matrix with the current tracklets and represent the dynamics-based similarity the matrix rank. To calculate the rank, an improved Hankel total least squares (IHTLS) algorithm was proposed. The algorithm could reduce the influence of data noise and predict the missing information between two tracklets. However, the real-time ability need to be further improved due to the mass computation of the rank minimization estimation.

In this paper, we proposed a multi-object tracking algorithm based on the combination of grey relational analysis and generalized linear assignment. In our method, just the motion feature is adopted to associate the tracklets, and the motion feature similarity was measured by grey relational analysis without assumption on the motion mode of objects and the scenes. Then, the data association was optimized with the generalized linear assignment. Our approach can track the similar appearance objects with high accuracy and few time consumption.

2 Moving Track Association Based on Generalized Linear Assignment

As the interference of object detection and scene factors, tracklets formed by conservative association are discontinuous in time and space. When one tracklet is associated with another, a variety of factors need to be considered such as similarity, time, etc. Therefore, multi-object tracking is usually formulated as an optimization problem, and the above factors are regarded as constraints or parameters of the optimization process.

Assume that N tracklets exist in the scene during a period of time, let $T = \{T_1, T_2, ...T_N\}$ represent the tracklets set, X_{ij} represent the association relationship between tracklet T_i and T_j. X_{ij} means that the two tracklets belong to the same object, otherwise means the opposite. C_{ij} is defined to represent the degree of similarity between two tracklets. Then, the optimization object function of multi-object tracking can be derived as follows:

$$\begin{aligned} &\arg\max_X C_{ij} X_{ij} \ \ or \ \arg\min_X C_{ij} X_{ij} \\ &st.\ X_{ij} \in (0,1) \end{aligned} \tag{1}$$

Using maximization or minimization relies on the value of . However, the above function couldn't be optimized directly and some constraints should be added in. For example, one tracklet could be only associated with one successor or predecessor. These constraints were adopted in [4], and two virtual nodes were produced to simulate the emergence and disappearance of object trajectory. However, as the randomness of object movement, it is difficult to accurately estimate the probability of objects emergence and disappearance without prior knowledge. Therefore, the constraint condition needs to relax that the object track cannot match the trajectory, and the optimization object function was revised as follows:

$$\begin{aligned} &\arg\max_{X} C_{ij} X_{ij} \\ &st.\ X_{ij} \in \{0,1\}; \\ &\quad \sum_{i=1}^{N} X_{ij} \le 1; \sum_{j=1}^{N} X_{ij} \le 1; \end{aligned} \tag{2}$$

The above formula is a special form of the generalized linear assignment problem. Although such a relaxation does not need to estimate the object emergence and disappearance, it is an "NP-hard" problem completely and can only be solved by an approximate feasible solution. A "soft partition" algorithm called deterministic annealing was proposed in [7]. An optimal estimation solution could obtain by the method.

In fact, the core of tracking by data association is calculating the similarity parameters. Its accuracy would greatly influence the tracking performance. And the constraints could only affect the solving process and the optimal degree of solution. If the similarity parameters are error, it will not be able to get the right association result with the optimization algorithm. Therefore, we propose to use the gray correlation analysis to measure the motion feature similarity of the tracklets.

3 Multi-object Tracking Based on Grey Relational Analysis

3.1 Grey Relational Analysis

As an important part of the grey theory, grey relational analysis is widely used in image engineering, decision analysis, etc. The essence is to find the complicated relationships among various factors of the system through the geometric similarity between data sequence curves [8][9]. The degree of grey incidence is a specific index of the gray correlation analysis. Its value indicates the degree of data correlation, higher for more correlation, whereas lower for few correlation. The original degree of grey incidence proposed by Deng [10], and the absolute degree of incidence proposed by Liu [11] are commonly used in practical problems. Define two behavior sequences: $X_i = (x_i(1), x_i(2), ..., x_i(n))$, $X_j = (x_j(1), x_j(2), ..., x_j(n))$. Let the sequences $X_i^0 = (x_i^0(1), x_i^0(2), ..., x_i^0(n))$, $X_j^0 = (x_j^0(1), x_j^0(2), ..., x_j^0(n))$ represent another two sequences, and the elements of which are generated by subtracting the start point of X_i and X_j, e.g. $x_i^0(k) = x_i(k) - x_i(1)$. Denote the absolute degree of incidence between X_i and X_j as ε_{ij}, which can be calculated as follows:

$$\varepsilon_{ij} = \frac{1 + |s_i| + |s_j|}{1 + |s_i| + |s_j| + |s_i - s_j|} \tag{3}$$

where $s_i - s_j = \int_1^n (X_i^0 - X_j^0)dt$ is the integral of the difference on two sequences, $s_i = \int_1^n (X_i - x_i(1))dt$ and $s_j = \int_1^n (X_j - x_j(1))dt$ are the integral of the difference in the sequences each other. As the absolute degree of incidence is symmetrical and unrelated to the order of X_i and X_j, we use this method to measure the tracklets similarity in this paper.

3.2 Tracklets Similarity Based on Grey Relational Analysis

Reliable tracklets of the objects could be acquired by conservative association method (such as bipartite graph method [12]). And these tracklets need to be associ-ated again for a complete trajectory of the object.

3.2.1 Tracklets Similarity Based on the Absolute Degree of Incidence

Define $T_i = \{T_i^k, k = 1...n\}$ as a tracklet, where $T_i^k = (x_i^k, y_i^k, w_i^k, h_i^k)$ contains the center coordinates, width and height information of object i, n is the length of the tracklet. So the similarity φ_{ij} between T_i and T_j could be calculated as follows:

$$\varphi_{ij} = \frac{\varepsilon_{ij}^x + \varepsilon_{ij}^y}{2} \tag{4}$$

where ε_{ij}^x and ε_{ij}^y represent the absolute degree of incidence in the x , y direction between the center coordinates of the two tracks respectively. And this method is called grey relational analysis (GRA). The specific calculation steps are as follows:

(1) Suppose that the end time t_i^e of T_i is less than the start time t_j^s of T_j, and the track length is n_i and n_j respectively.

(2) Extract center coordinates data of a tracklet with length n_m . For the tracklet T_i, the data is extracted backward, which is started from the time t_i^e . And for the tracklet T_j, the data is extracted forward, which is started from the time t_j^s, The parameter n_m can be calculated as $n_m = \min(\min(n_i, n_j), K)$, where K is a constant, and usually set to 5.

(3) Let X_i , Y_i represent the behavior sequences taken from tracklet T_i, and X_j , Y_j as the behavior sequences taken from tracklet T_j. Then calculate the grey degree of incidence ε_{ij}^x and ε_{ij}^y according to equation (3).

3.2.2 Revised Track Similarity Based on Corrected Degree of Incidence

Using grey relational analysis to measure the similarity mainly rely on the geometrical characteristics of two tracklets only. And there is no need to make any assumption about the movement of objects.

However, the disadvantage of this method is not able to represent a negative correlation [13]. Two tracklets with the same geometrical characteristics, may have the opposite direction and belong to different objects. Furthermore, the contribution of degree of incidence in and direction may be different. In equation (4), it is simply combined with equal weight. Then, an error may yield on the similarity measure. In order to solve the above issue, the moving direction and speed change of objects were taken to revise equation (4), which termed as weighted grey relational analysis(WGRA). The specific process is as follows:

(1) Define $v_x^i, v_y^i, v_x^j, v_x^j$ as the speed in x, y direction of tracklets T_i and T_j which can be derived as follows：

$$\begin{cases} v_x^i = x_i^{n_i} - x_i^{n_i - 1} \\ v_y^i = y_i^{n_i} - y_i^{n_i - 1} \\ v_x^j = x_j^2 - x_j^1 \\ v_y^j = y_j^2 - y_j^1 \end{cases} \tag{5}$$

(2) Define θ as the angle between motion the directions of object and calculate its cosine as below.

$$\cos(\theta) = \frac{v_x^i v_x^j + v_y^i v_y^j}{\sqrt{v_x^i v_x^i + v_y^i v_y^i}\sqrt{v_x^j v_x^j + v_y^j v_y^j}} \tag{6}$$

When the difference in the motion direction of two objects is insignificant, the possibility of them from the same object is larger, and vice versa is two objects. Then, this corrected degree of incidence between two tracks can be calculated as follows:

$$\begin{cases} \overline{\varepsilon_{ij}} = \mathrm{sgn}[\cos(\theta) < \lambda] * \varepsilon_{ij} \\ \mathrm{sgn}[\cos(\theta) < \lambda] = \begin{cases} 1 & ,\cos(\theta) < \lambda \\ -1 & ,else \end{cases} \end{cases} \tag{7}$$

(3) Using the speed difference of the traclets to weight the degree of incidence in x and y direction. Then the revised similarity coefficient $\overline{\varphi_{ij}}$ could be calculated as follows:

$$\cos(\theta) = \frac{v_x^i v_x^j + v_y^i v_y^j}{\sqrt{v_x^i v_x^i + v_y^i v_y^i}\sqrt{v_x^j v_x^j + v_y^j v_y^j}} \tag{8}$$

The corrected grey degree of incidence can represent the positive and negative correlation relationship of two tracklets based on the difference of motion direction . The similarity coefficient weighted with speed can provide better discriminate

Compared to the algorithm based on Hankel matrix, measuring the similarity of tracklets based on grey relational analysis has the advantages of more fast and high accuracy.

3.3 Cascade Optimization

After the similarity of tracklets is required, set $C_{ij} = \varphi_{ij}$ to optimize the GLAP. In order to speed up the calculation process, the multi-object tracking was conducted on the cascade optimization method similar with that in [6]. The whole process is as follows: firstly, divide the video sequence into a series of equally spaced clips and associate the object tracklets in each clip; secondly, according to a certain offset, slide a constant width in time and associate tracklets again; thirdly, in double time window clips, associate the exist tracklets again. This method can tolerate various intervals on the motion of objects effectively, so that the trajectory is more and more complete.

4 Experimental Results

4.1 Evaluation Criteria and Experiment Condition

4.1.1 Evaluation Criteria

Many evaluation criteria could be used to judge the performance of multi-object tracking. The author of [14] puts forward the most-used performance-evaluation index. We used four indexes: Multi-Object Tracking Accuracy (MOTA), False Negative (FN), False Positive (FP) and Miss Match (MM), which could be calculated as follows:

$$\begin{cases} \text{MOTA=1-}\dfrac{\sum_t (fn_t + fp_t + mm_t)}{\sum_t gt_t} \\ FN = \sum_t fn_t \\ FP = \sum_t fp_t \\ MM = \sum_t mm_t \end{cases} \tag{9}$$

Where fn_t, fp_t, mm_t, gt_t represent the number of incorrectly-correlated tarcklets, the remaining number of positive tarcklets, the number of correlation-changed tracklets and the number of referenced tracklets respectively in frame t. And MOTA combined FN, FP and MM, is a relatively comprehensive reflection on the accuracy of multi-object tracking algorithms. Besides, we also use average processing time per frame (TF) to measure the speed of an algorithm.

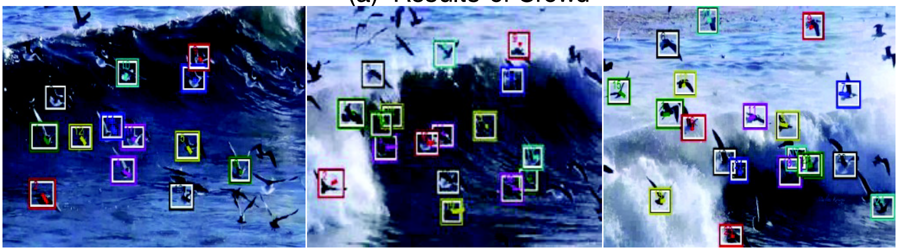
4.1.2 Experiment Condition

Our algorithm was developed under Matlab. The tested dataset and parameters are the same with [6]. The parameter λ was always set between (-1,0), such as -0.9 can get good results. Our dataset SMOT has eight videos and targets have similar features in these videos. All the detection results of moving targets were manually annotated. And we compare the results with [6] by the proposed two algorithm, which denote as GRA and WGRA for convenience.

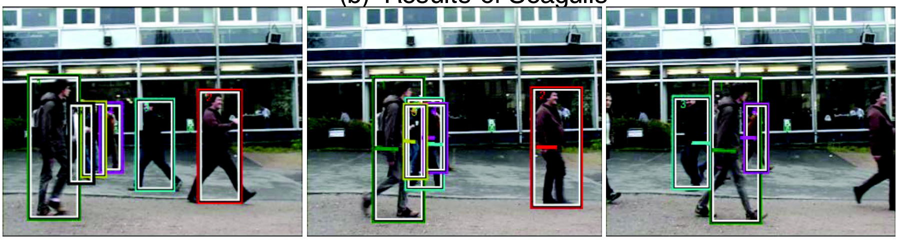
4.2 Experiments Results

The results of our algorithm are shown in Fig. 1. From the results, we can see that utilizing the motion feature, similar appearance objects can be tracked accurately, and the algorithm has the more strong robustness to occlusions.

(a) Results of Crowd



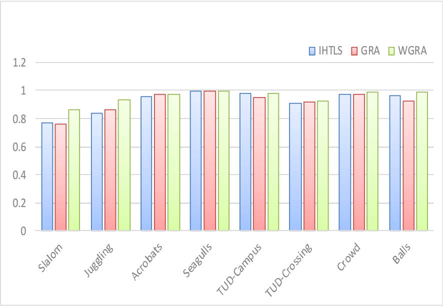
(b) Results of Seagulls



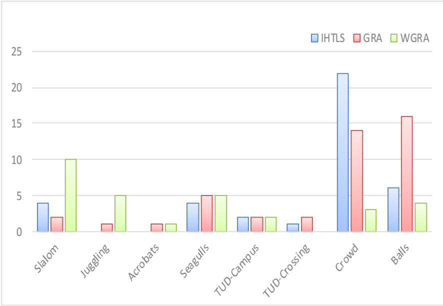
(c) Results of TUD-campus

(d) Results of TUD-crossing

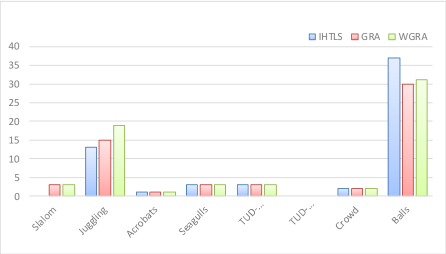
Fig. 1. The tracking results of proposed algorithm



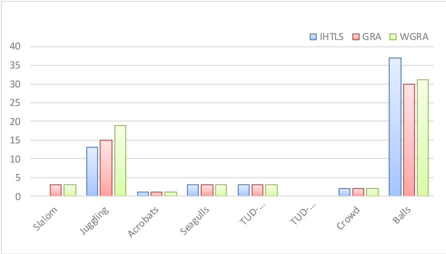
(a) MOTA



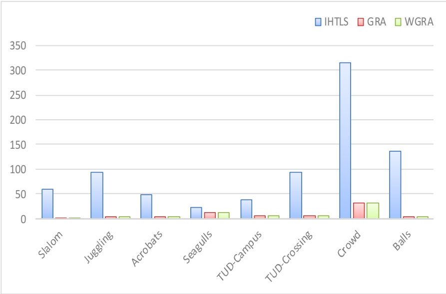
(b) MM



(c) FN



(d) FP



(b) TF

Fig. 2. The metrics comparison of three algorithms

Fig. 2 shows a quantitative comparison of three algorithms. From the four histograms, we can see that WGRA performs best on MOTA, FP and MM while GRA is the optimal algorithm on FN. Furthermore, the average TF of WGRA is about 8 ms, 12 times shorter than that of IHTLS.

We also found that WGRA had excellent performance in distinguishing alternating movement (such as in TUD-crossing) because the corrected degree of incidence with the different direction of movement. However, for a single object performing back and forth movement (such as in Juggling), WGRA would handle it as two separate objects, but GRA could have better performance. So the calculation method of grey correlation influenced the tracking results largely.

5 Conclusion and Future Work

Under the influence of mutual occlusion, similar objects and other factors, the appearance features are not stable and discriminative. In this paper, we use motion features only to track similar appearance objects. The data association problem was resolved by generalized linear assignment optimization. To measure the similarity of tracklets, grey relational analysis is adopting to calculate the similarity on motion feature. Furthermore, we associate the discrete tracklets in a hierarchical processing way. Our method could deal with occlusion and camera motion effectively without appearance features, and is applicable to multi-object tracking in complicated scenes. Comparison experimental results with IHTLS show that our algorithm has obvious advantages. The average tracking accuracy rate reached 95% and the processing time per frame was just 8 ms. Our method can be applied in the offline video analysis. In the future, we will study new calculation method of grey degree of incidence and apply it to online tracking.

References

1. Nawaz, T., Poiesi, F., Cavallaro, A.: Measures of effective video tracking. IEEE Transactions on Image Processing **23**(1), 376–388 (2014)
2. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Computing Surveys (CSUR) **38**(4), 1–46 (2006)
3. Wang, X.: Intelligent multi-camera video surveillance: A review. Pattern Recognition Letters **34**(1), 3–19 (2013)
4. Zhang, L., Li, Y., Nevatia, R.: In: 2008 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, USA, p. 1, June 23–28, 2008
5. Wen, L., Li, W., Yan, J., et al.: Multiple target tracking based on undirected hierarchical relation hypergraph. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, pp. 1282–1289. IEEE Computer Society (2014)
6. Dicle, C., Campsm O.I., Sznaier, M.: The way they move: tracking multiple targets with similar appearance. In: Proceedings of IEEE International Conference on Computer Vision (ICCV), New Jersey, pp. 2304–2311. IEEE (2013)
7. Gold, S., Rangarajan, A.: Softmax to softassign: Neural network algorithms for combinatorial optimization. Journal of Artificial Neural Networks **2**(4), 381–399 (1996)
8. Jiang, B.C., Tasi, S.L., Wang, C.C.: Machine vision-based gray relational theory applied to IC marking inspection. IEEE Transactions on Semiconductor Manufacturing **15**(4), 531–539 (2002)
9. Lin, Y., Pang, J., Li, Y.: A New Grey Relational Fusion Algorithm Based on Approximate Antropy. Journal of Computational Information Systems **9**(20), 8045–8052 (2013)
10. Julong, D.: Introduction to grey system theory. The Journal of Grey System **1**(1), 1–24 (1989)
11. Naiming, X., Sifeng, L.: Research on evaluations of several grey relational models adapt to grey relational axioms. Journal of Systems Engineering and Electronics **20**(2), 304–309 (2009)

12. Cox, L.J., Hingorani, S.L.: An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking. PAMI **18**(2), 138–150 (1996)
13. Ma, Y.: Study on the second grey relational grade and its properties. Kybernetes **39**(8), 1330–1335 (2010)
14. Bernardin, K., Stiefelhagen, R.: Evaluating multiple object tracking performance: the CLEAR MOT metrics. EURASIP Journal on Image and Video Processing **2008** (2008)