Video synopsis algorithm based on two-stage target tubes grouping

Yuting Wang[1,*], *Zhixiang* Zhu[1,2], *Chenwu* Wang[2], and *Pei* Wang[2]

[1]School of Computing, Xi'an University of Posts & Telecommunications, 710061 Xi'an, Shaanxi Province, China
[2]School of Modern Postal, Xi'an University of Posts & Telecommunications, 710061 Xi'an, Shaanxi Province, China

Abstract. Video synopsis generates a concentrated video that can be browsed quickly. With the increase of condensation ratio, more pseudo collisions between target tubes will occur. To solve this problem, this paper proposed a video synopsis algorithm based on two-stage target tubes grouping. In the first stage, using the hypergraph to analyze the collision relationship between target tubes, and the target tubes are grouped according to the hyper-edges. In the second stage, a clustering algorithm based on equal distance nearest neighbor sampling is proposed to group the target tubes. Then, selecting target tubes according to the selection principle of quantity priority between groups and length priority within groups (QPB-LPG) . Finally, these target tubes are rearranged to generate concentrated videos with smaller pseudo collisions. The experimental results show that this algorithm can significantly reduce the pseudo collision between target tubes without reducing the frame condensation ratio and frame compact rate compared with existing video concentration algorithms, and the feasibility of the method is fully verified.

1 Introduction

As an important form of data representation, video has been more used in criminal investigation, security management and other fields. People urgently need a way to efficiently obtain the main content of video and occupy less memory. Video synopsis technology is one of the effective means to solve this problem.

The collision of target tubes in video synopsis is one of the important problems in current research. Literature [1-2] reduces the collisions between targets by moving targets or changing the size and speed of targets. However, some target motion information will be lost while changing target attributes. Literature [3-5] use graph to analyze the collision relationship between two target tubes, and then uses graph coloring method to rearrange the target tubes. But, these methods can only analyze the collision relationship between two targets. To solve this problem, this paper attempts to analyze and deal with the collisions between target tubes by using the hypergraph which can analyze the collision relationship between multiple targets. Some researchers [6-8] used hypergraph to solve the problem of

* Corresponding author: 1361764030@qq.com

video summarization from multiple perspectives. This paper attempts to use hypergraph to associate the colliding target tubes and group them, so as to avoid the colliding targets in the same frame in the process of rearrangement.

In order to optimize concentrated videos, this paper proposed a video synopsis algorithm based on two-stage target tubes grouping. On the basis of hypergraph grouping, the target tubes in the known grouping are grouped by clustering. Then selecting and rearranging them, and calculating the energy loss, avoid different targets with collision at the same time.

The main contributions of this paper are as follows:

(1) Using hypergraph to solve the problem of collisions between target tubes;

(2) Proposing a two-stage target tube grouping algorithm;

(3) Proposing a target tube concentration strategy of selecting first and then rearrangement.

2 Two-stage grouping and rearrangement of target tubes

In order to reduce collision, a video synopsis algorithm based on two-stage target tube grouping is proposed in this paper. The basic algorithm is shown in Figure 1.

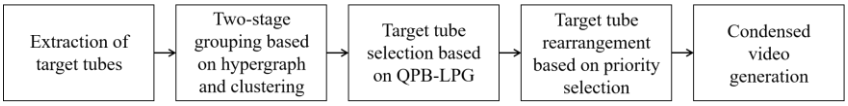

Fig. 1. Basic framework of algorithm.

2.1 Two-stage target tube grouping algorithm

2.1.1 Stage 1: grouping based on the hypergraph

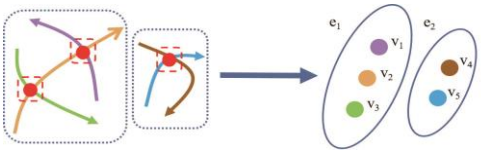

Fig. 2. Correspondence of hypergraphs.

$$B_{ij}=\begin{bmatrix} t_{11} & t_{12} & \cdots & \cdots & t_{1n} \\ t_{21} & \ddots & \ddots & \ddots & \cdots \\ \cdots & \ddots & \ddots & \ddots & \cdots \\ \cdots & \ddots & \ddots & \ddots & \cdots \\ t_{m1} & \cdots & \cdots & \cdots & t_{mn} \end{bmatrix} \quad (1)$$

Hypergraph is a generalization of graph. A hyperedge can connect multiple vertices. Therefore, this paper applied hypergraph to the video synopsis to solve the collision problem between multi-target tubes. Hypergraph is composed of vertex set $V=\{v_1,v_2,\cdots,v_m\}$ and hyper-edge set $E=\{e_1,e_2,\cdots,e_n\}$. As shown in Figure 2, the vertex set V is used to represent the target tube set, and the hyper-edge set E is used to represent the collision relationship between multi-target tubes, then target tubes with collision relationship are connected with a hyper-edge. And the collision frame length is used to

describe the collision of target tubes. Using a matrix B_{ij} represent the collision degree between two target tubes in the group, as shown in formula (1):
Where the value t_{ij} represents the frame length of the collision between the target tube vertex i and the target tube vertex j. We set a collision threshold t_0. When t_{ij} is greater than t_0, it is recorded as a serious collision, otherwise, it is recorded as a slight collision. The vertices determined as serious collision are connected by a hyper-edge. The target tubes are grouped according to the collision relationship to obtain group set $G=\{G_1,G_2,\ldots\ldots,G_M\}$. There was no collision relationship between the target tubes from different groups.

2.1.2 Stage 2: cluster grouping of target tubes based on equidistant nearest neighbor sampling

Because the direction and position of the targets in the groups are often not unified. During rearrangement, the moving targets in opposite directions may appear alternately, making the video length longer. Therefore, the grouping of the second stage is added. The group $G=\{G_1,G_2,\ldots\ldots,G_M\}$ obtained in the first stage is grouped by DBSCAN clustering to obtain group $C=\{C_{M1},C_{M2},\ldots\ldots,C_{MN}\}$, similar targets are grouped into one class according to the distance measurement and direction measurement between the target tubes, and the target tubes are grouped according to the class. As shown in Figure 3, if no second stage grouping, the target tubes in different directions appear alternately during rearrangement, like tube 1 and tube 2, tube 3 and tube 4, resulting in a lot of redundancy.

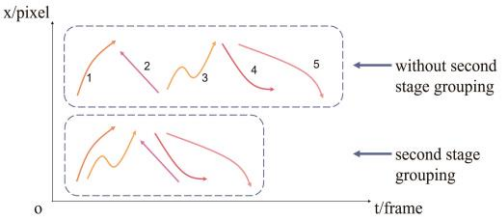

Fig. 3. Second stage grouping.

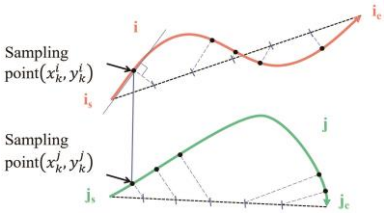

Fig. 4. Equidistant nearest neighbor sampling method.

In reference [9], the distance measurement between two targets is calculated by using the starting point and the ending point. Considering the number of points selected by the above method is too small, an equal distance nearest neighbor sampling method is proposed. The specific method is to take the bisection points on the connecting line between the starting point and the ending point, and find the points closest to the bisection points on the target tubes, and take these points as the sampling points. As shown in Figure 4, i and j are two independent target tubes, i_s and i_e are the starting point and end point of the target

tube i, j_s and j_e are the starting point and end point of the target tube j, and k represents the k-th sampling point from the starting point, $k \in N$. As shown in formula (2):

$$d(i,j) = a * \left(\sum_{k=1}^{n} \sqrt{\left(x_k^i + x_k^j\right)^2 + \left(y_k^i + y_k^j\right)^2} \right) + (1-a)\,\mathrm{rad}(\theta) \tag{2}$$

Where $d(i,j)$ is the similarity of the two target tubes, a is the weight coefficient, $d_\alpha(i,j)$ and $d_\beta(i,j)$ are the distance measurement and direction measurement of the two target tubes, θ is the included angle of the connecting line between the starting point and the ending point of the two target tubes, and $rad()$ represents the radian system of θ.

2.2 Selection of target tubes

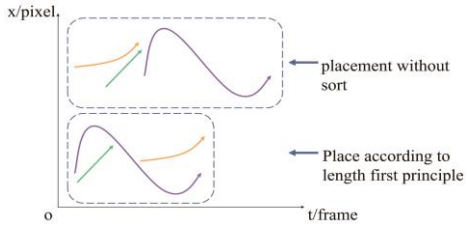

Fig. 5. Length priority within groups.

The target tubes extracted from the video often have different shapes, lengths and number of classes. During rearrangement, the video length becomes longer due to the lack of compactness of the targets. Therefore, a QPB-LPG selection principle is designed in this paper. For the group sets with more than 2 groups, give priority to group $C_1 \in C$ with the largest number of target tubes in G_1. Then, for the selected group C_1 and group $G_2 \in G$ with only one group, give priority to the longest target tube in each group in these groups. Considering that such selection may lead to poor robustness, roulette selection strategy is added to the selection [10].

For the length in the group, as shown in Figure 5, according to the principle of length first, the shorter tube placed later can be inserted into the front video, and the video length is significantly shorter.

2.3 Rearrangement of target tubes

Considering that there will be some losses in the process of grouping, selection and rearrangement. Therefore, using the energy function to reflect the loss in these process. The energy function $E(F)$ is calculated as equation (3):

$$E(F) = E_c(F) + E_t(F) + E_p(F) \tag{3}$$

Where, F represents the set of starting frames of all target tubes, $E_c(F)$ is the collision loss item, $E_t(F)$ is the long loss item, and $E_s(F)$ is the sorting priority loss item, where:

(1)Collision loss item: this loss item is defined as the number of target active pixels that are not occluded in the original videos but are occluded in the concentrated video. As shown in equation (4):

$$E_c(F)=\omega_c \cdot \sum_{i,j}\sum_{k=1}^{n} Mask(T_i(k)) \cap Mask(T_j(k)) \tag{4}$$

ω_c is the normalized parameter of the collision loss item, and $Mask(T_i(K))$ is the pixel after segmentation of the instance of the k-th frame of the i-th target tube.

(2)Duration loss item: this loss item is defined as the maximum end time of all target tubes during rearrangement, that is, the length of the last concentrated video. As shown in equation (5):

$$E_t(F)=\omega_t \cdot \max\left(\left\{F_{T_i}^{end}\right\}\right) \tag{1}$$

ω_t is the normalized parameter of the long loss item.

(3)Sorting priority loss item: this loss item is defined as the reverse order of the time order and priority order of the target tube during rearrangement. The larger the value, the greater the degree of destruction of the current priority order, as shown in equation (6):

$$E_p(F)=\omega_p \cdot \sum_{i,j} \delta\left(\left(P_{T_i}-P_{T_j}\right)\left(F_{T_i}^{start}-F_{T_j}^{start}\right)<0\right)\left|P_{T_i}-P_{T_j}\right| \tag{2}$$

ω_p is the normalization parameter of sorting priority loss item, which is 1 when $\delta(cond)$ is true and 0 when $\delta(cond)$ is false. P_{T_i} is the priority sequence number of the i-th target tube.

3 Experimental design and result analysis

3.1 Experimental data sets

The data set scenes are shown in Figure 6. Figure 6 (a) shows the data set Library which has a total of 15200 frames. Figure 6 (b) shows the data set Corridor which has a total of 15884 frames. Figure 6 (c) shows the data set Road1 which has a total of 4650 frames. Figure 6 (d) shows the data set Road2 which has a total of 4500 frames. Figure 6 (E) shows the data set Shopping mall which has a total of 9275 frames.

Fig. 6. Dataset scenario diagram

3.2 Comparison of target tube grouping algorithms

Our grouping algorithm is compared with the DBSCAN clustering algorithm based on Hausdorff distance [11] (DBSCAN clustering algorithm based on Hausdorff distance, HD-DCA) and DBSCAN clustering algorithm based on symmetric distance function [9] (DBSCAN clustering algorithm based on symmetric distance function, SDF-DCA) in five different scenes.

In this section, the effect of grouping algorithm is measured by adjusting Rand index [12] (ARI) and normalized mutual information [13] (NMI).

(1)Adjusted rand index (ARI): reflect the overlapping degree of the two division methods. If the value is close to 1, the better the clustering effect.

(2)Normalized mutual information (NMI): an information theory used to measure the degree of mutual prediction between clustering results and predefined clusters based on internal information, and to measure the similarity of two clustering results.

Table 1. Comparison of grouping algorithms.

		Library	Corridor	Road1	Road2	Shopping mall
HD-DCA	ARI	0.469	0.239	0.342	0.219	0.257
	NMI	0.815	0.651	0.460	0.519	0.285
SDF-DCA	ARI	0.693	**0.538**	0.238	0.928	0.422
	NMI	0.910	**0.790**	0.468	0.879	**0.721**
Our	ARI	**0.997**	0.339	**0.735**	**0.971**	**0.510**
	NMI	**0.961**	0.730	**0.903**	**0.948**	0.573

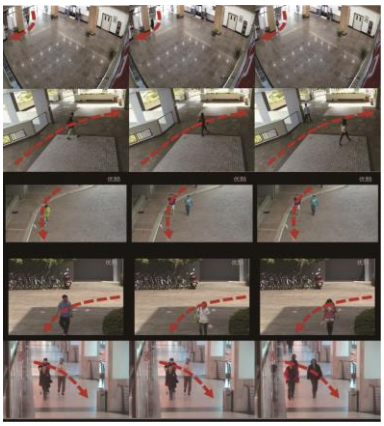

Fig. 7. Display of target trajectory in the same direction of original videos

Fig. 8. Grouping effect of each algorithm

Figure 7 shows different algorithms in different scenarios from top to bottom. Figure 8 shows the effects of HD-DCA, SDF-DCA and the algorithm grouping in this paper from left to right. The red line indicates that the targets shown in Figure 7 are correctly divided into same groups, the green line indicates the target tracks in other directions divided

together wrongly, and the orange line indicates that groups lacks targets in the same directions. The experimental data are shown in Table 1.

According to the Figure 8 and Table 1, the experimental results of the two-stage target tube grouping algorithm are closer to the expected grouping effect, and there are many deficiencies in the other two algorithms.

3.3 Comparison of video synopsis algorithms

Our video synopsis algorithm, the video synopsis algorithm combining object speed and size change [14] (OSSC-VC) Video synopsis method considering trajectory geographic direction[15] (TGD-VC) is compared in five different scenes.

(1) Frame compression ratio [4] (FR): the ratio of the number of frames in the video summary and the original video, $FR = T_S / T_I$, where T_S is the number of frames in the video summary and T_I is the number of frames in the input original video.

(2) Frame compact rate [4] (CR): it is used to judge whether the target tube rearrangement in the summary video is compact. The calculation formula is as follows (9):

$$CR = \frac{1}{w \cdot h \cdot T_S} \sum_{t=1}^{T_s} \sum_{x=1}^{w} \sum_{y=1}^{h} \left\{ 1 \middle| p(x, y, t) \in \text{foreground} V_S \right\} \tag{3}$$

(3) Overlap ratio [4] (OR): used to indicate the collision degree of the target tube in the summary video. The calculation formula is as follows (10):

$$OR = \frac{1}{w \cdot h \cdot T_S} \sum_{t=1}^{T_s} \sum_{x=1}^{w} \sum_{y=1}^{h} \left\{ 1 \middle| p(x, y, t) \in \text{foreground} V_S \right\} \tag{10}$$

Table 2. Comparison of video synopsis algorithms.

		Library	Corridor	Road1	Road2	Shopping mall
OSSC-VC	FR	0.066	0.126	0.215	0.388	**0.323**
	CR	0.023	0.022	0.062	0.137	0.116
	OR	0.002	0.006	0.012	0.005	0.021
TGD-VC	FR	0.035	**0.054**	0.377	0.198	0.220
	CR	0.042	**0.063**	0.030	0.116	**0.235**
	OR	0.003	0.007	0.095	0.014	0.065
Our	FR	**0.031**	0.088	**0.103**	**0.156**	0.307
	CR	**0.050**	0.042	**0.107**	**0.162**	0.194
	OR	**0.001**	**0.001**	**0.003**	**0.002**	**0.015**

It can be seen from Table 2 that in most scenarios, the video concentration algorithm based on two-stage target tube grouping proposed in this paper has better experimental results.

4 Conclusion

Our algorithm divided tubes into several groups by analyzing the collision and position relationship between targets, and the two-stage grouping method in this paper effectively group target tubes. Later, the QPB-LPG selection principle is adopted to determine the position of the target tube with larger space occupation first, so as to facilitate the subsequent insertion of the target tube with smaller space occupation. Then, rearranged target tubes according to the energy loss. Finally, experiments show that the two-stage

grouping video synopsis method proposed in this paper groups effectively reduces the collision rate of concentrated video, improves the condensation ratio of concentrated video, and achieves good concentration effect. However, this method also has some shortcomings. The next step will focus on how to improve the execution efficiency of the algorithm.

References

1. Y. W. Nie et al. Collision-Free Video Synopsis Incorporating Object Speed and Size Changes[J]. IEEE Transactions on Image Processing, 2019: 1-1.
2. Y. W. Nie, C. X. Xiao et al. Compact video synopsis via global spatiotemporal optimization[J]. IEEE transactions on visualization and computer graphics, 2013.
3. Y. He, Z. G. Qu, C. X. Gao, N. Sang. Fast Online Video Synopsis Based on Potential Collision Graph[J]. IEEE Signal Processing Letters, 2016: 1-1.
4. Y. He, C. X. Gao, N. Sang, Z. G. Qu, J. Han. Graph coloring based surveillance video synopsis[J]. Neurocomputing, 2016, 225: 64-79.
5. T. Ruan, S. K. Wei, J. Li, Y. Zhao. Rearranging Online Tubes for Streaming Video Synopsis: A Dynamic Graph Coloring Approach[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2019.
6. N. A. Arafat, S. Bressan. Hypergraph Drawing by Force-Directed Placement[C]//International Conference on Database and Expert Systems Applications, 2017.
7. Y. W. Fu et al. Multi-View Video Summarization[J]. IEEE Transactions on Multimedia, 2010, 12: 717-729.
8. Z. Ji, S. F. Fan. Video Summarization with Random Walk on Hypergragh[J]. Journal of Chinese Computer Systems, 2017, 38: 2535-2540.
9. Q. Lu, G. B. Yang, J. T. Tan, Y. Yu, X. H. Yuan. Multi-representation Trajectory Clustering Method in Visualization of Road Traffic Trend[J]. Journal of Computer-Aided Design & Computer Graphics, 2019, 031. 1194-1202.
10. S. K. Feng, L. Zhen, Y. Dong, S. Z. Li. Online Content-aware Video Condensation[C]. IEEE Conference on Computer Vision & Pattern Recognition, 2012: 2082-2087.
11. B. Sai, Z. Q. Cao, Y. J. Tan, X. Lv. Pedestrian data mining with object tracking and trajectory clustering[J]. Systems Engineering - Theory & Practice, 2021, 41: 9.
12. N. X. Vinh, J. Epps, J. C. Bailey. Information theoretic measures for clusterings comparison: is a correction for chance necessary? [J]. Journal of Machine Learning Research, 2010, 11: 2837-2854.
13. T. P. Q. Nguyen, R. J. Kuo. Partition and merge based fuzzy genetic clustering algorithm for categorical data[J]. Applied Soft Computing, 2018, 75.
14. X. L. Li, Z. G. Wang, X. Q. Lu. Surveillance Video Synopsis via Scaling Down Objects[J]. Transactions on Image Processing, 2015.
15. Y. J. Xie, X. J. Liu. Surveillance Video Synopsis Considering Trajectory Geographic Direction[J]. Geomatics and Information Science of Wuhan University, 2017, 042: 70-76.