Research on Sports Video Content Analysis Method Based on Clustering Extraction Algorithm

Yu Wang, Zhu Mei, Tianbo Yu*

Shenyang Jianzhu University, Shenyang, China

*Correspondence Author

Abstract. The algorithm first uses the time domain information of the sports video image to segment the foreground image containing multiple sports targets through sample variance for background modeling. Then the spatial connectivity rate of pixels is defined, and the initial clusters are adaptively split and merged. The self-organizing iterative clustering algorithm can complete the segmentation of multiple moving targets without setting the number of clustering divisions in advance. Experimental results prove that the algorithm has a good segmentation effect on multiple moving targets, and the segmentation results are consistent with the judgment of human vision. The use of spatial connectivity information makes the algorithm iteratively converge fast and has good real-time performance.

Keywords: Clustering extraction algorithm; Sports video; Fuzzy clustering algorithm; Membership degree.

1. Introduction

The rapid development of multimedia technology and network information has produced a large amount of video data, which makes it very difficult to obtain interesting information from the vast array of video information. In the process of solving this problem, people found that key frames can support the rapid speed of the entire video content. Query, retrieval and browsing. The application of key frames greatly reduces the amount of video data, and at the same time provides an organizational framework for processing video content. Current key frame extraction technologies are almost all based on shots, which can be roughly divided into 5 categories: (1) The method based on the lens boundary, this kind of method is simple but does not well reflect the complexity of the lens content, and the selected key frames are often not stable enough, and the representativeness is not very strong; (2) the method based on the visual content, It can select the corresponding number of key frames according to the degree of change of the lens content, but the selected key frames may not have representative meaning, and when the lens is moving, it is easy to select too many key frames; (3) Based on the method of motion analysis, This method requires a large amount of calculation when analyzing motion; (4) The method based on lens motion, camera motion is useful for professional photographers, but not meaningful for general users; (5) Method based on clustering [1] .Clustering is an unsupervised classification technology, which has been widely used in various fields such as pattern recognition, speech analysis, and information retrieval. This paper proposes a new clustering-possibility C-pattern (PCP) based on Class key frame extraction algorithm. Simultaneously apply it to sports video analysis.

2. Sports video compression method based on clustering algorithm

Video compression is to divide the video entities in the video database into several video clusters with a certain meaning, so that the entities in each video cluster have the greatest similarity, and the entity difference between clusters is the largest [2]. The best effect of video compression is to maximize the cohesion between entities in the video cluster, and minimize the cohesion between different clusters. Therefore, if the cohesive force between entities in a video cluster is stronger, the tendency of "clumping" is more obvious, and the existence of a video cluster is more reasonable. Different clusters are distinguished because the mutual cohesion of the entities between the clusters

is too small. Video outliers can be regarded as a special video cluster, so the cohesion of the outliers should also be small. Based on the above assumptions, the specific definition of the evaluation index of video compression effectiveness is given below:

For video cluster C, the formation process can be described as starting from any video entity, and continuously "adsorbing" the entity that has the greatest cohesion with the current entity, and finally all entities in the cluster form a tree structure connected by cohesion, called Condensed tree, denoted as AT(C). The condensed tree and the minimum spanning tree have the same geometric structure. The cohesion between all entities in the condensed tree is the strongest, which can better reflect the tightness of the video cluster entities, and then give the definition of cohesion within the cluster:

For a video cluster C containing n entities, the generated aggregation tree is AT(C), and all the cohesion between entities in the aggregation tree constitutes the cohesion set AF(C), the average value of the cohesion is recorded as μAF, and the variance is recorded as σAF. Furthermore, the intra-cluster cohesion FI(C) can be expressed as:

$$F_I(C)=\begin{cases}\dfrac{\sum F_{agg}(p_i,p_j)}{m}, F_{agg}(p_i,p_j)\le \mu_{AF}-\sigma_{AF}\\ \mu_{AF}, \forall F_{agg}(p_i,p_j)>\mu_{AF}-\sigma_{AF}\end{cases} \tag{1}$$

In the formula: m represents all cohesive forces less than μAF-σAF. It can be seen that the definition of cohesion within a cluster takes into account the uniformity of entity distribution and avoids the interference of individual extreme outliers. For any video cluster (or video outlier) C, its inter-cluster cohesion is defined as the average value of the cohesion of entities in other clusters to entities in C, denoted as FE(C), expressed as:

$$F_E(C)=\frac{\sum F_C(p_i,p_j)}{w} \tag{2}$$

pi∈C,pj∉Cand pj∈ND (pi). Where: w represents the number of entities in other video clusters (including video outliers) that have a cohesive effect with entities in video cluster C. For a given video database SDB, let's suppose that all its entities are divided into M video clusters, denoted as C1,C2,…,CM; at the same time, N video outliers are obtained, denoted as O1,O2,…,ON. Therefore, the video compression effectiveness metric can be defined as the ratio of the minimum value of cohesion in all clusters to the average value of cohesion between clusters, denoted as FTSCV, expressed as:

$$FTSCV=\frac{\min\limits_{i=1,\ldots,M}F_I(C_i)}{\dfrac{\sum_{i=1}^{M}F_E(C_i)+\sum_{i=1}^{N}F_E(O_i)}{M+N}}=\frac{(M+N)\min\limits_{i=1,\ldots,M}F_I(C_i)}{\sum_{i=1}^{M}F_E(C_i)+\sum_{i=1}^{N}F_E(O_i)} \tag{3}$$

It can be seen from formula (3) that for a video compression result, if the cohesion of entities in the video cluster is stronger and the cohesive effect of entities between clusters is weaker, the larger the FTSCV value, the better the clustering effect [3]. At the same time, it can be found that the influence of video outliers is taken into account in the evaluation process of this article, and when calculating the cohesion between clusters, the video clusters are regarded as a whole, which fully reflects the cohesive relationship between different clusters and avoids The traditional evaluation index of separation between clusters (such as shortest distance, centroid distance) can not be applied to the shortcomings of arbitrary shape video clusters.

3. Kalman filter sports video tracking processing

Kalman filter is to estimate the current value of the signal based on the previous estimate of the system and the latest observation data. Therefore, in addition to the state equation, it is also necessary to establish an observation equation [4]. Assuming that there is a linear relationship between the state

variables of the system and the observed data, use y_k to represent the observed signal vector, then the relationship between the state variable x_k and the observed signal vector y_k is:

$$y_k = C_k x_k + \upsilon_k \tag{4}$$

In the formula, C_k is the observation matrix, which is a matrix of $m \times n$, and m and n are the dimensions of y_k and x_k respectively. The first term $C_k x_k$ on the right side of the equation represents the useful signal s_k, and the second term υ_k is noise, which is a random vector representing the observation error. In general, it can be assumed that υ_k is normal white noise, and the mean value is zero.

We have the multi-dimensional signal model of Kalman filter as shown in Figure 1. If the dashed box in the figure is represented by the transfer function $A(z)$, the one-dimensional signal model of Kalman filter as shown below is obtained.

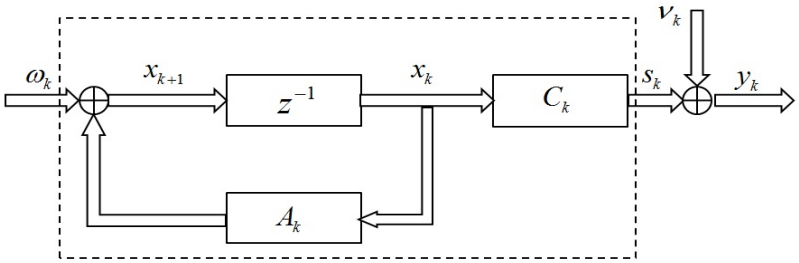


Figure 1. Kalman filtering signal model

The essence of Kalman filter is to solve the estimated value $\hat{x}_k$ of x_k under the minimum mean square error criterion [5]. Its characteristic is that x_k can be calculated by recursive method. Assume that the state equation and observation equation of a linear dynamic system are:

$$x_k = A_k x_{k-1} + \omega_{k-1} \tag{5}$$

$$y_k = C_k x_k + \upsilon_k \tag{6}$$

In the formula, x_k is a multi-dimensional state vector, y_k is a multi-dimensional observation vector, A_k is a state transition matrix, C_k is an observation matrix, ω_k and υ_k are process noise and observation noise, respectively.

Assumptions: (1) ω_k and υ_k are normal white noise with zero mean, and they are not correlated with each other. (2) The initial state vector x_0 is a random vector, and its statistical characteristics are given and independent of ω_k, υ_k A_k and C_k in formula (5) and formula (6) are known, and y_k is the observed data, which is also known. Therefore, solving the state equation and the observation equation is the process of solving the estimated value $\hat{x}_k$ of x_k by knowing y_k and x_{k-1}. The specific solution process is as follows:

(1) Assuming ignoring ω_k and υ_k, then x_k and y_k obtained according to formulas (5) and (6) are represented by $\hat{x}_k^{'}$ and $\hat{y}_k^{'}$ respectively, then we get

$$\hat{x}_k^{'} = A_k \hat{x}_{k-1} \tag{7}$$

$$\hat{y}_k^{'} = C_k \hat{x}_k^{'} = C_k A_k \hat{x}_{k-1} \tag{8}$$

Where $\hat{x}_{k-1}$ is the estimated value of x_{k-1},

(2) Compare the $\hat{y}_k^{'}$ calculated by the formula (8) with the actual observation value y_k, and record the difference between the two as $\tilde{y}_k = y_k - \hat{y}_k^{'}$. $\tilde{y}_k$ contains the information of the noise ω_k

and υ_k and the information of the current observation value y_k. If $\tilde{y}_k$, which contains noise and observation information, is multiplied by a filter gain matrix K_k to correct $\hat{x}_k^{'}$, the error caused by ignoring noise in step (1) will be reduced, and a better estimate will be obtained

$$\hat{x}_k = A_k\hat{x}_{k-1} + K_k(y_k - \hat{y}_k^{'}) = A_k\hat{x}_{k-1} + K_k(y_k - C_kA_K\hat{x}_{k-1}) \tag{9}$$

According to the above formula, if we can obtain K_k under the condition of the minimum mean square error between the estimated value $\hat{x}_k$ and the true value x_k, and then substitute K_k into formula (9), the obtained $\hat{x}_k$ is the value of x_k under the minimum mean square error criterion Linear optimal estimation.

(3) Solve K_k under the condition of minimum mean square error. The mean square error is a square error matrix, denoted by P_k as:

$$P_k = E\left[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T\right] = E\left[\tilde{x}_k\tilde{x}_k^T\right] \tag{10}$$

Assuming that ω_k and υ_k are normal white noise with zero mean, and they are not correlated with each other, we get

$$\begin{aligned} &E\left[\omega_k\right] = 0;\ E[\omega_k\omega_j^T] = Q_k\delta_{kj} \\ &E\left[\upsilon_k\right] = 0;\ E[\upsilon_k\upsilon_j^T] = R_k\delta_{kj} \\ &E\left[\omega_k\upsilon_j^T\right] = 0;\ k, j = 0,1,2,... \end{aligned} \tag{11}$$

Among them, $\delta_{kj} = \begin{cases} 0 & ,i \neq k \\ 1 & ,i = k \end{cases}$, $Q_k = \mathrm{var}[\omega_k]$, $R_k = \mathrm{var}[\upsilon_k]$. Solving the minimum mean square error matrix P_k, one can derive a series of Kalman one-step recursive formulas: state one-step prediction:

$$\hat{x}_k^{'} = A_k\hat{x}_{k-1} \tag{12}$$

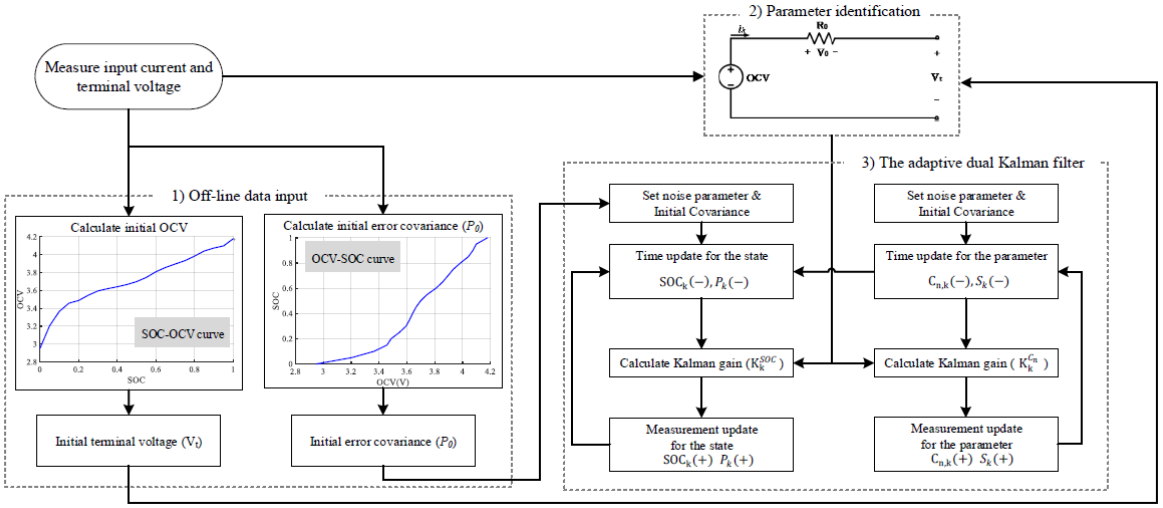


Figure 2. Two calculation loops of the Kalman filter

One-step prediction mean square error matrix $P_k^{'} = A_kP_{k-1}A_k^T + Q_{k-1}$, filter gain matrix $K_k = P_k^{'}C_k^T(C_kP_k^{'}C_k^T + R_k)^{-1}$, filter mean square error matrix $P_k = (I - K_kC_k)P_k^{'}$, and state filter $\hat{x}_k = A_k\hat{x}_{k-1} + K_k(y_k - C_kA_k\hat{x}_{k-1})$. If the system characteristics $E[x_0]$ and $\mathrm{var}[x_0]$ of the initial state x_0 are known, let $\hat{x}_0 = E[x_0]$, $P_0 = E[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T] = \mathrm{var}[x_0]$, and then according to the above series of formulas, the state estimation value $\hat{x}_k$ at all times can be obtained. Figure 2 shows the filter calculation loop and gain calculation loop of the Kalman filter (picture quoted from Integrated

Approach Based on Dual Extended Kalman Filter and Multivariate Autoregressive Model for Predicting Battery Capacity Using Health Indicator and SOC/SOH).

4. Experimental analysis

The experiment extracts two continuously changing moving images from the final video of the 2016 National College Basketball League to implement segmentation, and performs morphological filtering and frame data correction on each region. Generally speaking, the smaller the SA, the higher the spatial accuracy of the segmentation method [6]. The SA calculation results of different segmentation methods are shown in Table 1. The fuzzy clustering algorithm used in this paper has the smallest SA value among the four segmentation methods, which can effectively fill the gaps in the sports video space, prevent vacancies in the image segmentation area, and can be accurate Image with more complicated factors such as segmentation environment and color degree, with high spatial accuracy. Then use the FTSCV index and Dunn index proposed in this paper to evaluate the clustering results. The Dunn index uses the diameter of the cluster to indicate the tightness of the video clusters, and uses the distance between the clusters to indicate the degree of separation between the video clusters. For a clustering result with k video clusters, the Dunn index is specifically expressed as follows:

$$D_k = \min_{i=1,\ldots,k}\{\min_{j=i+1,\ldots,k}[\frac{d(C_i,C_j)}{\max_{m=1,\ldots,k}(diam(C_m))}]\},$$
$$d(C_i,C_j) = \min_{x\in C_i, y\in C_j}\{d(x,y)\}, \quad (13)$$
$$diam(C_m) = \max_{x,y\in C_m}\{d(x,y)\}$$

In the formula, d(Ci,Cj) is the shortest distance between entities in clusters; diam(Cm) is the diameter of clusters, that is, the maximum distance between entities in clusters [7]. The larger the Dunn index value, the better the clustering result.

Table 1. Effectiveness evaluation results of DBSCAN and K-means clustering

Database	DBSCAN			K-means
	Parameter	*FTSCV*	Dunn	*FTSCV*
Database 1	Eps=28, MinPts=5	1.35	0.1	0.09
	Eps=29, MinPts=5	1.66	0.35	0.14
	Eps=35-109, MinPts=5	37.03	0.35	37.03
	Eps=110-119, MinPts=5	2.32	0.26	0.07
Database 2	Eps=20, MinPts=6	2.21	0.05	0.1
	Eps=25, MinPts=6	15.73	0.05	0.18
	Eps=28-64, MinPts=6	28.68	0.12	0.16
	Eps=65, MinPts=6	2.53	0.13	0.11
Database 3	Eps=30, MinPts=5	2.28	0.11	0.03
	Eps=35, MinPts=5	9.63	0.11	0.07
	Eps=45, MinPts=5	15.21	0.69	0.01
	Eps=190, MinPts=5	2.69	0.39	1.32

In order to analyze the role of the pixel connectivity rate in the spatial clustering algorithm proposed in this paper, the changes in the pixel connectivity rate in the algorithm are counted and compared. With the increase of the number of iterations, the change trend of the average and intermediate value of the pixel points in each cluster is to gradually adjust the pixels in the initial cluster so that the pixels in each cluster are connected to the cluster center the rate reaches the iterative convergence condition [8]. A total of 3 iterations of clustering segmentation were performed on the multi-moving target in the foreground image 2. After each iteration, the change trend of the average and the median value of the pixel connectivity rate in each cluster is shown in Figure 3 (picture quoted from Super voxel Segmentation with Voxel-Related Gaussian Mixture Model). It can be seen from

Figure 3 that compared with the second iteration, the average and the median value of the connectivity rate of the pixels in different clusters have changed a lot in the first iteration, and after the second and third iterations The connectivity rate of the pixels in the obtained clusters has a small change and has stabilized. It can be seen that the effect of cluster splitting and merging in the spatial clustering algorithm is obvious, and as the number of iterations increases, the cluster splitting and merging tend to be stable.

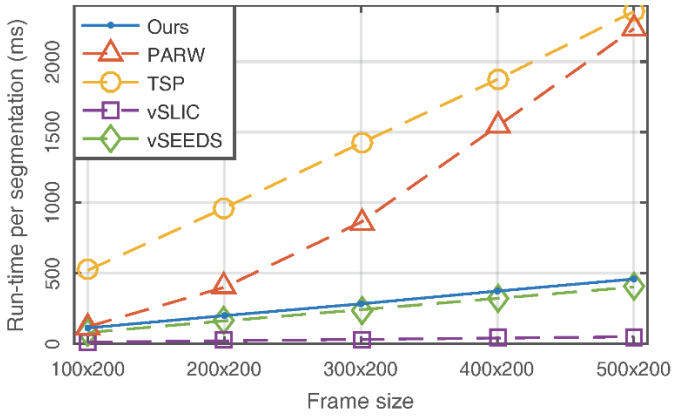


Figure 3. Iterative convergence graph of sports video image clustering segmentation

5. Conclusion

This paper presents an algorithm for extracting sports foreground region information of sports frequency sequences based on frame difference intersection clustering. In the simulation experiment, relatively satisfactory results have been achieved. In terms of real-time performance, the algorithm does not involve complicated mathematical operations. Compared with the existing methods, the algorithm has certain advantages in operation speed and can meet the real-time requirements. Experimental results prove that the algorithm is a better algorithm for extracting prospect athletes from sports video sequences.

References

[1] Zhang Yifan, Li Jiachen, Kuang Yuanyou, et al. Research on video key frame extraction technology based on K-means clustering. Computer and Information Technology, vol. 29, pp. 4-14, January2021.

[2] Huang Xuan. Chinese video text recognition based on clustering and multiple neural networks. vol. 5, pp. 7-11, January 2021.

[3] Ding Ruixin, Chen Yuanrui. Analysis of Key Frame Extraction Algorithms in Content Video Retrieval. Information and Computer, vol. 32, pp. 25-27, April 2020.

[4] Zhou Ju, Luo Bing. Video abstract extraction algorithm based on multi-feature layering. Journal of Wuyi University: Natural Science Edition, vol. 11, pp. 39-45, February 2020.

[5] Li Daxiang, Zhu Zhiyu, Liu Ying. Surveillance video summary algorithm based on motion trajectory clustering. Computer Engineering and Design, vol. 40, pp. 55-60, June 2019.

[6] Zhang Dong. Point cloud data classification and extraction based on FCM clustering algorithm. Engineering Technology Research, vol. 4, pp. 2-9, Fifteen 2019.

[7] Zhang Lu, Lu Jinlai. Video abstract based on hierarchical clustering and TextRank. Computer Engineering and Design, vol. 40, pp. 55-59, July 2019.

[8] Chen Zhiwei, Liu Yuhan. Photo classification technology based on density clustering algorithm. Heilongjiang Science and Technology Information, vol. 11, pp. 75-76, March 2019.