Multi dimension query of sports video for cloud service environment

LV Kun, Wang Qingbin

Zhuhai College of Jilin University, Zhuhai Guangdong, china 519000

Email:20416674@qq.com

Keywords: Sports Video, Least Squares Support Vector Machine, Divider Design, Features Extraction, Evidence Theory

Abstract. A kind of sports video classification method (D-S PT-LSSVM) combined with least squares support vector machine (LSSVM) and evidence theory has been proposed for the shortages of low accuracy rate and weak stability etc. of sports video with single features. First, four kinds of features of color, texture, brightness and motion vectors reflecting sports video category have been extracted; second, basic probability assignment has been constructed by taking the LSSVM initial category results of four kinds of single features as independent evidence and finally simulation experiment has been made. Simulation results have shown that the accuracy rate of sports video classification of D-S PT-LSSVM is as high as 97.9%; compared with reference method, the specific sports video classification of D-S PT-LSSVM is with advantages of high accuracy rate and good stability etc.

Introduction

Scholars both at home and abroad have made deep research on the classification of sports video classification and many effective classification methods have been proposed [2]. The classification of sports video mainly includes features extraction, divider design and other key contents; at present, there are sports video extraction methods of mainly static features, movement features and integration of these two etc [3-5]; single static features or movement static features can only describe part and clip information of sports video category and misclassification phenomenon happen easily, therefore, multi-feature has become the main sports video classification method [5]. Compared with single features, multiple features method can describe sports video category information from various perspectives and the accuracy rate of sports video classification has been improved accordingly, however, traditional features combination algorithm merely combines several features together in a simple way but not realize effective integration, which has caused high complexity of feature space and high dimension of divider input but the accuracy, timeliness and stability of sports video classification is still very low. In addition to features extraction, the classification results of sports video is related to divider design; sports video divider constructs with support vector machine (SVM) at the moment. For the huge amount of data, the SVM training speed is slow while least squares support vector machine (LSSVM) is an improved support vector machine, which has overcome the shortages of over-fitting of traditional machine learning algorithm and slow SVM training speed effectively. Therefore, sport video divider has been constructed in this paper with LSSVM.

Sports Video Algorithm of D-S PT-LSSVM

Extraction of sports video characteristics

Extract motion vector field

(1) Set the size of sports video as M×N×T, M×N is resolution, T is the length of video sequence, divide video into K×L blocks and size of each block is h×v, in which h=M/K, C is the block quantity within each block, as shown in picture 1 (K=L=4).

(2) Set up Cartesian coordination system and map motion vector into this coordinate system, as shown in picture 2. MV(i,j) is block with position (i,j), $\theta \in [0,2\pi)$ is the direction of motion vector C. If use Cx to express the weight of No. C motion vector at horizontal direction (x), Cy expresses the

weight of No. C motion vector at vertical direction (y), ρ expresses the motion intensity of block C, and then

$$\begin{cases} \rho = \sqrt{c_x^2 + c_y^2}, \\ \sin(\theta) = c_y / \rho \\ \tan(\theta) = c_y / c_x \end{cases} \tag{1}$$

(3) Arrange coordinate system of continuous video frame based on time order, divide it into Q sogonal fans along x forward direction, quantize p into R intervals and then make histogram for p and θ, it can obtain

$$\begin{cases} Hist_q = \dfrac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{C} q_i^t \quad q \in [1, Q] \\ Hist_r = \dfrac{1}{q}\sum_{t=1}^{T}\sum_{i=1}^{C} r_i^t \quad r \in [1, R] \end{cases} \tag{2}$$

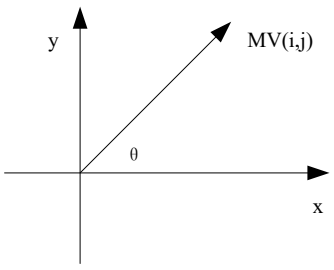
In the formula, q_i^t expresses the quantity of motion vector of No. t frame at quadrant q, r_i^t is the quantity of p quantized to r range in No. t frame coordinate system.

(4) Evaluate motion situation within block with expectation and variance of motion vector within block along x and y directions. That is

$$\begin{cases} \mu_x = \dfrac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{C} C_{x,i}^t \\ \mu_y = \dfrac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{C} C_{y,i}^t \\ \sigma_x^2 = \dfrac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{C} \left(C_{x,i}^t - \mu_x\right)^2 \\ \sigma_y^2 = \dfrac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{C} \left(C_{y,i}^t - \mu_y\right)^2 \end{cases} \tag{3}$$

In the formula, $C_{x,i}^t$ and $C_{y,i}^t$ are weights of motion vectors of No. i block with one frame along x and y directions, μ_x, μ_y, σ_x^2 and σ_y^2 expectations and variances of macro block motion vector along x and y directions.

Picture 1 Diagram of Sports Video Block Classification



Picture 2 Motion vector field mapping

Extract brightness features

Assume frame resolution is M×N, each frame has been divided into k×k blocks and size of each

block is h×v, in which h=M/K, v=N/K, xi is the brightness value of No. i pixel within block and the brightness mean value of each block is $\overline{X(l)}$, l∈[1,k×k], and then:

$$\overline{X(l)} = \sum_{i=1}^{h\times v} x_i / (h \times v) \tag{4}$$

Take y as comparison of code vale of block brightness, and then within the frame, the code value of brightness comparison result between No. m block and No. n block can be expressed by formula (5), in which l≤m≤k×k, 2≤n≤k×k-1.

$$u = \begin{cases} 1 & if \ \ \overline{X(m)} > \overline{X(n)} \\ 0 & otherwise \end{cases} \tag{5}$$

Make “1” and “0” coding for frame based on brightness mean value through formula (5)

Extract color features

Assume the size of frame is M×N, transfer frame into HSV model, divided into k×k blocks, size of each block is h×v, in which h=M/K, v=N/K. Use xi,m,n to express the pixel value of No. i pixel m weight within No. n block in video, in which n∈[1,k×k]， i∈[1,h×v], m∈[H,S,V], and then the features of sports video color can be expressed as following:

$$\mu_{m,n} = \frac{1}{h\times v} \sum_{i=1}^{h\times v} x_{i,m,n} \tag{6}$$

$$\sigma_{m,n} = \sqrt{\frac{1}{h\times v} \sum_{i=1}^{h\times v} \left(x_{i,m,n} - \mu_{m,n} \right)^2} \tag{7}$$

$$S_{m,n} = \sqrt[3]{\frac{1}{h\times v} \sum_{i=1}^{h\times v} \left(x_{i,m,n} - \mu_{m,n} \right)^3} \tag{8}$$

In the formula, $\mu_{m,n}$, $x_{i,m,n}$, $S_{m,n}$ are mean value, variance and third-order moment of m weight within No. n block.

Extract texture feature

Gray level co-occurrence matrix is a method commonly used to describe correlation feature of gray space, therefore, it can be used to express the texture feature of sport video. Set sports video f has L gray levels, G is a gray level co-occurrence matrix, element pij is the times of pixels with gray levels of i and j in f, pij calculation formula is:

$$p_{ij} = N\{(x,y) | f(x,y)=i, f(x+\Delta x, y+\Delta y)=j\} \tag{9}$$

In the formula, f(x,y) is the gray level of pixel at (x,y); Δx and Δy reflect distance d and direction θ between two points.

Five commonest texture features have been adopted as classification characteristics of sports video [13], whose definitions are as following:

$$\begin{cases} f_1 = \sum_{i=1}^{L} \sum_{j=1}^{L} p_{ij} \\ f_2 = \sum_{i=1}^{L} \sum_{j=1}^{L} (i-j)^2 p_{ij} \\ f_3 = \sum_{i=1}^{L} \sum_{j=1}^{L} \log p_{ij} p_{ij} \\ f_4 = \sum_{i=1}^{L} \sum_{j=1}^{L} (p_{ij})^2 \\ f_5 = \sum_{i=1}^{L} \sum_{j=1}^{L} \frac{p_{ij}}{1+|i-j|} \end{cases} \tag{10}$$

D-S PT Theory

Evidence theory is proposed by Dempster and improved by Shafer, so it is also called as D-S PT theory. D_S PT theory integrates belief functions of two or more evidences into a new belief

function through composition rule; the integrated function is taken as basis of decision [14]. It principle is as following:

Set Θ as testing frame, define function $m: 2^{\Theta} \to [0,1]$ meeting: $m(\varnothing) = 0$, ($\varnothing$ is null set), $\sum m(A) = 1 (A \in 2^{\Theta})$, and then $m(A)$ is called as basic probability assignment (PBA) on frame Θ. When $A \neq \Theta$, $m(A)$ is precision trust degree of proposition A, $m(\Theta)$ is the uncertainty of evidence.

The synthesis rule is: set m1，m2，…，mn are the BPA of different evidences on testing frame Θ, and then their orthogonal sum $m = m_1 \oplus m_2 \oplus \cdots \oplus m_n$ can be confirmed as

$$m(A) = \frac{\sum_{B \cap A_i = A} \sum_{j=1} m_j(A_i)}{1 - \sum_{B \cap A_i = \varnothing} \sum_{j=1} m_j(A_i)} \tag{11}$$

Simulation Experiment

Source of Data

The data is attained from videos of football, volleyball, tennis and table tennis etc. intercepted from TV with video capture card. Sample number distribution of various video training sets and testing sets is as shown in table 1. Algorithm is realized through VC++ programming on Intel PIV 3.0GHZ CPU, 2G RAM, Windows XP platform.

Table 1 Sample number of different sports types

Type	Training set	Testing set
Football	500	200
Basketball	1000	300
Volleyball	1200	400
Tennis	400	100
Table tennis	2000	500
Badminton	1500	400

Comparison Algorithm and Evaluation Index

To make D-S PT-LSSVM classification results more convincing, color features +LSSVM(LSSVM1), texture features +LSSVM(LSSVM2), brightness features +LSSVM(LSSVM3), motion vector features +LSSVM(LSSVM4) and traditional combination features +LSSVM(LSSVM5) have been selected for comparison experiment; classification accuracy rate and misclassification rate have been adopted as evaluation index for algorithm performance.

Feature Normalization Treatment

LSSVM is the most sensitive for data at [0,1]. To improve training efficiency, normalization treatment has been made for sports video features, specifically:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{12}$$

In the formula, x is original feature, x' is feature after normalization, xmax and xmin are minimum value and maximum value respectively.

Initial Classification of LSSVM Sports Video

Input feature vectors of color, texture, brightness and motion vector field into LSSVM for study, LSSVM parameter γ=100, σ=0.195. Sports video classification models of LSSVM1, LSSVM2, LSSVM3 and LSSVM4 have been set up respectively. Testing results of test set are as shown in table 2.

Table 2 Accuracy rate of sports video classification of single algorithm

Video Type	LSSVM1	LSSVM2	LSSVM3	LSSVM4
Football	85.8	86.77	87.36	90.41
Basketball	82.13	85.2	84.96	87.48
Volleyball	63.72	62.87	64.1	86.3
Tennis	67.99	60.81	60.36	65.15
Table tennis	85.15	81.19	61.98	66.88
Badminton	62	66.57	61.67	83.3

It can be learnt from table 2 that when making classification with features of motion vector field, table tennis and tennis are with bad classification effect; when making classification with color features, badminton, tennis and volleyball are with bad classification effect; when making classification with brightness features, badminton, tennis, volleyball and table tennis are with bad classification effect; when making classification with texture features, badminton, tennis and volleyball are with bad classification effect, which indicate that sports video classification information can't be described in an accurate and comprehensive way with single features and the classification is with bad reliability and accuracy.

Calculate the value of reliability function

The thresholds of decision rule are ε1=0.50，ε2=0.65，ε3=0.25 respectively; Calculate each feature and reliability function value, specifically as shown in table 3.

Table 3 Reliability value of testing frame

Feature type	$m(A_0)$	$m(A_1)$	$m(\Theta)$
Color	0.445	0.224	0.612
Texture	0.780	0.264	0.239
Brightness	0.699	0.201	0.249
Motion vector field	0.762	0.132	0.250
Integration feature	0.913	0.207	0.344

Results and Analysis

Testing set has been classified with established sports video classification models of LSSVM1, LSSVM2, LSSVM3, LSSVM4, LSSVM5 and D-S PT-LSSVM; accuracy rate (%) of average classification and misclassification rate (%) have been attained as shown in picture 4 and picture 5. It can be learnt from picture 4 and picture 5 that in all sports video algorithms, D-S PT-LSSVM is with the best classification result and following conclusions can be reached:

(1) Affected by various factors, single feature can't reveal sports video classification information; sports video classification with single feature is low in accuracy rate, big in misclassification rate and unreliable in classification results. l

(2) The classification of simple and multi-feature combination algorithm is superior to algorithm with single feature, which explains that combining various features together can provide more information reflecting sports video and improve accuracy rate of classification to some extent, but lower than the accuracy rate of D-S PT-LSSVM. This indicates that redundant information happens easily and can't achieve optimal combination with simple combination of multiple features and the classification performance of sport video needs to be further improved.

(3) The testing results of D-S PT-LSSVM are with good stability, accuracy rate of classification can reach as high as 95.61% and misclassification rate can be decreased greatly, which is mainly caused by the fact that adopting LSSVM classification result to construct reliability assignment of D-S evidence theory, integrating multiple classification results of features from color, texture, brightness and motion vector etc based on evidence combination rule and realizing effective transfer from weak classification decision to strong classification, which can describe sports video classification information in a better way, improve accuracy rate of sports video classification and make the classification results more stable and reliable.

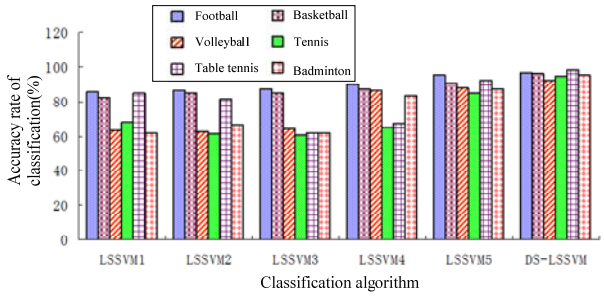


Figure 4 Accuracy rate of classification of different algorithms

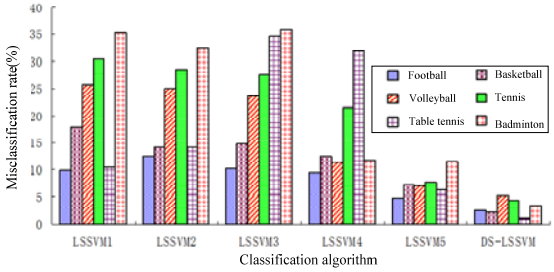


Figure 5 Misclassification rate of different algorithms

Conclusions

Sports video classification algorithm of D-S PT-LSSVM has been proposed for the difficulty of single feature and traditional simple feature in describing sports video classification information accurately and comprehensively. Simulation results have shown that the average classification accuracy rate of D-S PT-LSSVM can reach 95.61% and the accuracy rate and stability of sports video classification have been improved significantly.

Reference

[1] Jinyu Hu and Zhiwei Gao. Distinction immune genes of hepatitis-induced heptatocellular carcinoma[J]. Bioinformatics, 2012, 28(24): 3191-3194.

[2] Jiang, D., Ying, X., Han, Y., & Lv, Z. (2016). Collaborative multi-hop routing in cognitive wireless networks. Wireless personal communications, 86(2), 901-923.

[3] Lv, Z., Tek, A., Da Silva, F., Empereur-Mot, C., Chavent, M., & Baaden, M. (2013). Game on, science-how video game technology may help biologists tackle visualization challenges. PloS one, 8(3), e57990.

[4] Jiang, D., Xu, Z., & Lv, Z. (2015). A multicast delivery approach with minimum energy consumption for wireless multi-hop networks. Telecommunication systems, 1-12.

[5] Lv, Z., Chirivella, J., & Gagliardo, P. (2016). Bigdata Oriented Multimedia Mobile Health Applications. Journal of medical systems, 40(5), 1-10.