ISSN : 0974 - 7435 Volume 10 Issue 4

BioTechnology
An Indian Journal

FULL PAPER

BTAIJ, 10(4), 2014 [870-875]

Video semantic analysis applied research in football match special effects shots

Yezhi He

Institute of Physical Education, Shandong Normal University, Jinan 250014, (CHINA)

ABSTRACT

With football development, football each direction researches are also rapidly moving forward, especially for football video correlation analysis and research are more focused by broad scholars, which have become a hot topic in football research field. The paper just based on such thought, extracts football video features, integrates playback shots component detection model with multiple modals, intercepts highlights to test and makes personalized plans, which provides impetus for football video correlation fields' development. © 2014 Trade Science Inc. - INDIA

KEYWORDS

Match video;
Semantic analysis;
Football technique;
Special effects shots.

PREFACE

Since 1990s, it has already started researches about video contents aspect, from which analyses bout video could roughly divide into two kind of contents, one is correlation analysis of one kind of videos, it includes football video, news video and so on; and another kind is videos correlation analysis and processing with any contents, but the way due to " arbitrariness", its handling and analysis are very difficult to implement, on the contrary, its adaptation range also accordingly gets improved, the former one, due to its processing objects are very clear, it is very easy to handle with.

Regarding football match videos aspects research, lots of people have made efforts and got achievements, which provides beneficial conditions for scholars in all sectors of society researching on it, and provides impetus for its scientific development. Such as: Wu Ying-Ming analyzed on football video content aspect, he made conclusions according to two ways, proposed that one way was key techniques used in analysis methods, another one was during football video analysis records, made conclusions such as: billiards, basketball, tennis, football and others, and put forward using each kind of advanced techniques in videos such as: video application, analysis, modal features fusion technique, video features' extracting and others to make analysis and research; *Mihajlovic* has ever used pattern matching method to identify results and extract match's highlights, except for that, the scholar also adopted level difference method to position video letters' correlation regions; Zhang Xu and others has ever carried out correlation researches on baseball video's scoring events method and provided their own research results.

The paper on the basis of previous research results, it carries out correlation analysis and research on football match video semantics, and combines with examples to state the method implementation and application and show the method validity and rationality.

FULL PAPER

FOOTBALL MATCH VIDEO SEMANTIC EXTRACTING AND SELECTING

In general, football video event in football match process is syntheses of context clues' semantic paragraphs, it makes conversion from lens to scenes, so video basic unit is lens, its structure is as Figure 1 show:

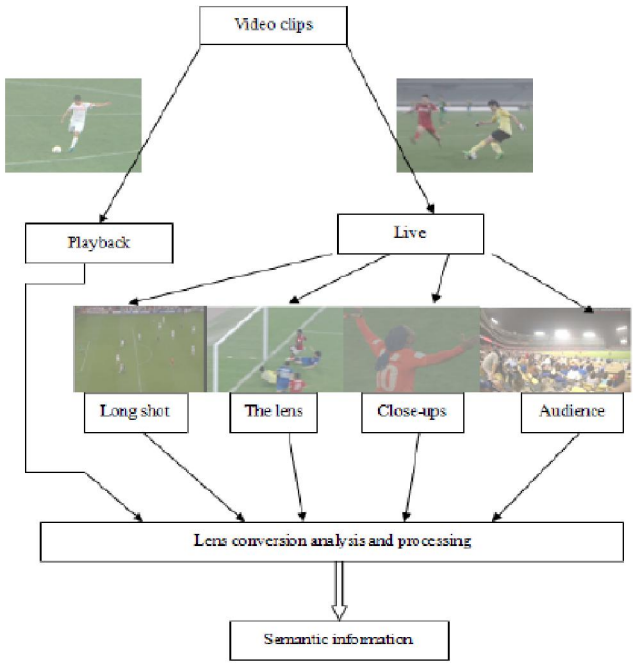

Figure 1: Hierarchical structure of the lens

In the following, use lens frame and their corresponding classifications that include technical analysis, highlights playback and other for semantic analyzing, which includes audience, close-up, medium, field lens, slow motion and so on, except for that it also includes other lens, due to individual lens occurrence probability is very little, so it hasn't been taken into consideration.

EXTRACT PICTURE'S CORRESPONDING FEATURES

For semantic analysis several lens, it proceeds with picture features' extracting, firstly it should extract main colors, and then adopts segmenting to extract image colors.

Color classification

Nowadays main colors space forms have $CIEL*ab$, $CIEL*u*v$, YUV、HSV、RGB these forms, from which most convenient and common one is RGB, and each letters from it represented colors are respectively blue, green and red, but the kind of colors space visual perception cannot better reflect, and HSV is different from it that is composed of brightness, component and tone three kinds, but the two can make mutual transformation, their transformation equation is:

$$r = R/(R+B+G) \quad (1)$$

$$g = G/(R+B+G) \quad (2)$$

$$b = B/(R+B+G) \quad (3)$$

$$u = \max(r,g,b),\ V = \min(r,g,b) \quad (4)$$

$$S = [u-v]/u \quad (5)$$

$$h = \begin{cases} 5+b' & if \quad r=u \quad and \quad g=v \\ 1-g' & if \quad r=u \quad and \quad g\neq v \\ 1-r' & if \quad g=u \quad and \quad b=v \\ 3-b' & if \quad g=u \quad and \quad b\neq v \\ 3+g' & if \quad b=u \quad and \quad r=v \\ 5-r' & otherwise \end{cases}$$

$$H = 60*h$$

$$r' = (u-r)/[u-v]$$

$$g' = (u-g)/[u-v]$$

$$b' = (u-b)/[u-v] \quad (6)$$

Among them, $V,S \in [0,1]; H \in [0,360];\ g,b,r \in (0,1)$, adopt cylinder distance measuring HSV color space, if there are two points $i,\ j$, corresponding colors are (V_i, H_i, S_i) and (V_j, H_j, S_j), color distance measure is:

$$d_{hue}(i,j) = \left|H_i - H_j\right| \quad (7)$$

$$d_{chroma}(i,j) = \sqrt{(S_i)^2 + (S_j)^2 - 2S_iS_j\cos(d_{hue}(i,j))} \quad (8)$$

$$d_{value}(i,j) = \left|V_i - V_j\right| \quad (9)$$

$$d_{hsv}(i,j) = \sqrt{d_{chroma}(i,j)^2 + (d_{value}(i,j))^2} \quad (10)$$

When $d_{value} = 0$, then it only need to consider $HandS$ components.

Extract colors' corresponding plan

In color histogram, it has corresponding peak-peak,

FULL PAPER

take several small intervals nearby the value, and in the intervals, it includes corresponding color values, if let histogram to be $h[\square]$, then, k is accommodation coefficient, i_{peak} is peak value maximum corresponding color value, $[i_{\min,} i_{\max}]$ is i_{peak} small interval as Figure 2 show:

In above Figure 2, its interval left side boundary is histogram's proportion that less than peak value, and its maximum color value is following formula:

$$h[i_{\min}] \geq k \cdot h[i_{peak}] \quad and \quad h[i_{\min} - 1] < k \cdot h[i_{peak}]$$

$$and \quad i_{\min} \leq i_{peak} \tag{11}$$

In above Figure 2, its interval right side boundary is histogram's proportion that less than peak value, and its minimum color value is following formula:

$$h[i_{\max}] \geq k \cdot h[i_{peak}] \quad and \quad h[i_{\max} - 1] < k \cdot h[i_{peak}]$$

$$and \quad i_{\max} \geq i_{peak} \tag{12}$$

After that, calculate $[i_{\min,} i_{\max}]$ histogram's weighted value is main color, as formula:

$$h_{mean} = \sum_{i_{\min}}^{i_{\max}} h[i] \cdot i \,/ \sum_{i_{\min}}^{i_{\max}} h[i] \tag{13}$$

So it can calculate football video main color value range $< H_m, S_m >$, and the main color selection is obtained by short time statistics before match, if the value is always unchanged, then it can be used as main color

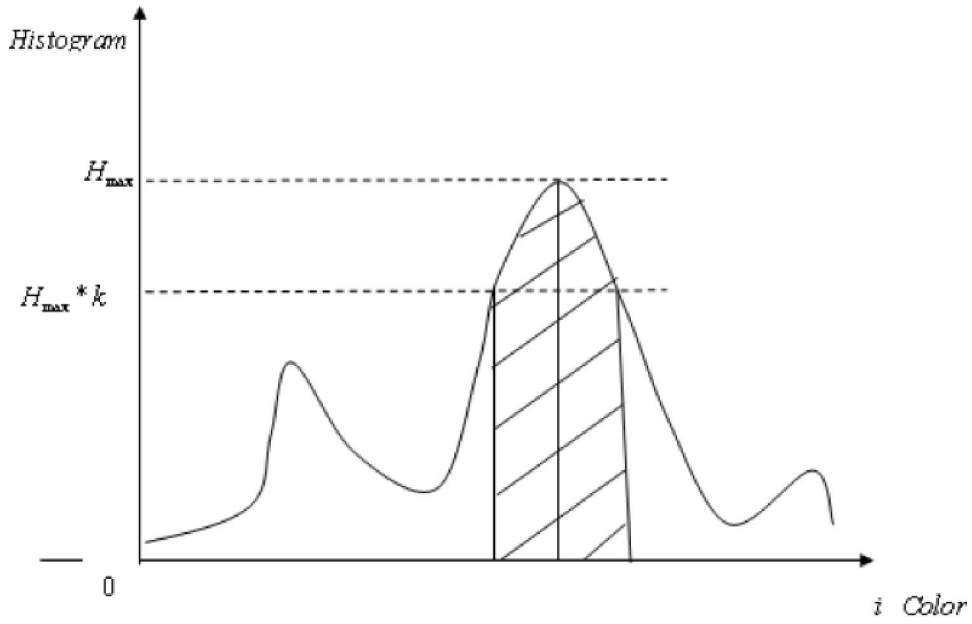

Figure 2 : Extracted main color values from color histogram

selection value, and can be used to extract and segment in field, assume in HSV image, it takes one point j, corresponding color is $< V_j, H_j, S_j >$, $< H_m, S_m >$ is field color, then its distance to j color is:

$$d_{hue}(i, j) = \left| H_j - H_m \right| \tag{14}$$

$$d_{hsv} = \sqrt{(S_j)^2 + (S_m)^2 - 2S_j S_m \cos(d_{hue})} \tag{15}$$

If it meets:

$$d_{hsv} < T_{hsv} \tag{16}$$

Among them, define T_{hsv} as threshold value, j is field color value

ATHLETE FACIAL DETECTION

In general, detection model about athlete face can be divided into two steps to proceed. Firstly is detecting complexion, corresponding probability function is:

$$p(x|\text{skin}) = \frac{1}{2\pi \left\| \sum_s \right\|^{1/2}} \exp$$

$$[-\frac{1}{2}(x - \mu_s)^T \sum_s^{-1} (x - \mu_s)] \tag{17}$$

In above formula $\sum_s$, μ_s represents covariance

matrix and complexion pixel's mean vector, and in normalized color space, input pixel chroma $r = R/(R+B+G)$, $g = G/(R+B+G)$ is using $x = [r, g]^T$ to express, if non-complexion and complexion have the same probability, if it has $p(x|\text{skin}) > T_S$, in this case it is complexion, then it has:

$$p(\text{skin}|x) = \frac{p(x|\text{skin})p(\text{skin})}{p(x|\text{skin})p(\text{skin}) + p(x|\text{skin})p(\text{skin})} = \frac{p(x|\text{skin})}{p(x|\text{skin}) + p(x|\text{skin})} \propto p(x|\text{skin}) \quad (18)$$

To a qualified athlete facial region, it should meet:

1) Let minimum enclosing rectangle aspect ratio parameter constraint (*Orientation*) to be:

$$Orientation = D_y / D_x \quad (19)$$

2) Regional size's height cannot be too small.
3) Minimum enclosing rectangle area and regional area derivative value is defined as compactness.

PICTURE EVALUATED RESULT

Use handwork to select from match plentiful videos that are respectively: two thousand audience graphs, two thousand field features, two thousand lens images and others, detection results mixed matrix after normalization is as TABLE 1 show:

By above TABLE 1 picture detection result, we can get that average accuracy rate is 90.3%, due to field area ratio's threshold is not easy to define, so *Medium* classification effect is the worst, and due to effective complexion detector and simple background, so the CloseF classification effect is best.

SLOW MOTION DETECTION AND PLAY-BACK ALGORITHMS

For slow motion's playback problems, it puts forward Re *play* algorithms to do corresponding detection, meanwhile it also combines with context and *Logo* advantages to detect, it totally includes three ways:

Logo – transition detecting algorithm

In football videos, *Logo* conversion process has plateau mode features, its conversion is nearly 0.5and 0.8 seconds, in the whole process, *Logo* icon lighten, after that it will gradually disappear, so we adopt detection steps are:

*step*1: In partial windows, it carries out mid-value smoothing and calculates by *MSD*

*step*2: To surpass T_i number c_d, it makes statistics, if it detects that *Logo* makes conversion, then T_i number is less than c_d, on the contrary it changes to zero and moves to next step.

TABLE 1: Classification result table

	Medium	CloseF	CloseNF	Audience	Prec.	F-mecu.	Long	Recall
CloseNF	0	0.0065	0.921	0.071	0.9579	0.9393	0.001	0.9215
Audience	0.0255	0	0.0405	0.898	0.9258	0.9117	0.036	0.898
Long	0.0797	0	0	0	0.9385	0.9296	0.9203	0.9203
Medium	0.8485	0.006	0	0.001	0.7074	0.7716	0.1445	0.8465
CloseF	0.0865	0.9135	0	0	0.9865	0.9486	0	0.9135

*step*3: For *Logo* template extracting process, it extracts n pieces of *Logo*, and record class group as: $f_1, f_2, f_3, \cdots, f_n$.

*step*4: Let c to be clustering index value, $d(f_i, f_j)$ is f_i and f_j images distance, calculate clustering center:

$$c = \arg\min\{\sum\nolimits_j d(f_i, f_j)\} \quad i = 1, 2, \cdots, n \quad (20)$$

*step*4: Calculate far from clustering center gray level average value, let $m = 0, 1, \cdots, M-1$, $n = 0, 1, \cdots, N-1$ the two respectively represent image sizes, average value is *Logo*

FULL PAPER

image lt:

$$lt(m,n)=\frac{1}{k}\sum_{k=1}^{K} f_k(m,n) \quad d(f_c, f_k) \tag{21}$$

Due to in match process, *Logo* is always unchanged, so the initial detection can be used as whole process result.

*step*5: Assume $\{f_i, i=1,2,\cdots,N\}$ is frame sequence, $R_i = r(f_{lr}, f_i)$ is image f_i and *Logo* normalize processing, obtained gray average relative coefficient is:

$$R_i=\frac{\sum_x\sum_y[f_{lt}(x,y)-\overline{f}_{lt}(x,y)][f_i(x,y)-\overline{f}_{lt}(x,y)]}{\{\sum_x\sum_y[f_{lt}(x,y)-\overline{f}_{lt}(x,y)]^2 * \sum_x\sum_y[f_{lt}(x,y)-\overline{f}_{lt}(x,y)]^2\}^{1/2}} \tag{22}$$

In order to look for *Logo* image frame, it needs to meet steps:

*step*1: Take R_i maximum window value that is: $R_i > R_j$, $j = l-\omega,\cdots,l-1,l,l+1,\cdots,l+\omega$

*step*2: R_i above one limit value

*step*3: *Logo* and f_i difference can use d to express, that:

$$d=\frac{\left|\vartheta' - \vartheta^{lt}\right|}{\vartheta^{lt}} \tag{23}$$

Re *play* **detection result**

For one basketball match (Portugal— Japan) and three football matches (Cameroon—Germany, England—Sweden, America— Portugal), they take *Logo* detection, and select proper *Logo* conversion number n, and test the value from 1to 5, assume $n=3$, its result is as TABLE 2 show:

TABLE 2 : Classification result table

	Number of leak detection	Number of false detection	*Logo* totals	Detected number
Cameroon—Germany	1	0	82	81
England—Sweden	2	1	90	88
America— Portugal	0	1	132	132

By above TABLE 2, it can get that due to partial *Logo* is lost in video recording process, subsequently Re *ply* detection occurs false, in the three matches, England—Sweden has two missing, Cameroon—Germany, Portugal— Japan each has one missing.

REGARDING FOOTBALL LENS CLASSIFI-

TABLE 3: Lens classification result

Match	Type of lens	Number of false detection	Number of detection	Totals	*prec.*(%)	Re*call*(%)
England—Sweden	Re *play*	5	43	44	97.5	95.6
	Long	24	276	294	97.6	93.5
	Medium	1	127	138	78.4	91.3
	Close – up	5	195	215	98.5	90.8
	Audience	4	14	14	75.6	100.0
Cameroon—Germany	Re *play*	2	37	42	95.1	90.3
	Long	10	276	294	96.2	95.6
	Medium	2	108	128	82.5	83.4
	Close – up	1	180	191	94.2	95.8
	Audience	2	8	8	81.3	100.0

CATION EXPERIMENTAL RESULT ANALYSIS

Input data into above formula, it can get by calculating that Cameroon—Germany the two countries and England—Sweden two countries' match classification results, as following TABLE 3 show:

By above TABLE 3, it can get *Audience* and *Medium* has certain quantities false detection status, which mainly due to field target leak detection caused *CloseF* and *Long* false and due to inaccurate estimate on athlete face leads to two false classifications.

CONCLUSIONS

By analyzing semantics about football videos as slow motion playback, semantic lens classification and so on, it constructs football videos frame structure, let the algorithm to be clear and easy understanding. If just by single videos' vision, the obtained football information is not enough to achieve required accuracy, so the paper combines with auditory sense, vision and other football relative information to make comprehensive analysis, so that it let multiple fuzzy data mutual integrate, the method is a kind of important processing method for analyzing and preprocessing with football videos. For football video event detection, it gets such way is a kind of complicated work, the paper applies semantic information to carry out events analysis and processing and get experiment results, which proves the method is effective.

REFERENCES

[1] Y.Ariki, S.Kubota, M.Kumano; Automatic Production System of Soccer Sports Video by Digital Camera Work Based on Situation Recognition[C]//IEEE Interna-tional Symposium on Multimedia. San Diego: IEEE, 851–860 **(2006)**.

[2] J.Wang, E.Chng, C.Xu et al.; Generation of Personalized Music Sports Video Using Multimodal Cues[J]. IEEE Trans. on Multimedia, **9(3)**, 576–588 **(2006)**.

[3] X.Yu, H.Leong, J.Lim et al.; Team Possession Analysis for Broadcast Soccer Video Based on Ball Trajectory[C]//International Conference on Information, Com-munication and Signal Processing. Singapore: Springer, **3**, 1811–1815 **(2003)**.

[4] J.Wang, N.Parame Swaran; Analyzing Tennis Tactics from Broadcasting Tennis Video Clips[C]//International Conference on Multimedia Modeling. Melbourne: Springer, 102–106 **(2005)**.