基于双阈值滑动窗口子镜头分割和完全连通图的关键帧提取方法

钟 忺[1] 杨 光[1] 卢炎生[2]
(武汉理工大学计算机科学与技术学院 武汉 430070)[1]
(华中科技大学计算机科学与技术学院 武汉 430074)[2]

摘 要 随着多媒体技术的发展,当今工作和生活中的多媒体信息日渐丰富。如何通过分析海量视频快速有效地检索出有用信息成为一个日益严重的问题。为了解决上述问题,提出了一种基于双阈值滑动窗口子镜头分割和完全连通图的关键帧提取方法。该方法采用基于双阈值的镜头分割算法,通过设置双阈值滑动窗口来判断镜头的突变边界和渐变边界,从而划分镜头;并采用基于滑动窗口的子镜头分割算法,通过给视频帧序列加一个滑动窗口,在窗口的范围内利用帧差来对镜头进行再划分,得到子镜头;此外,利用基于子镜头分割的关键帧提取算法,通过处理顶点为帧、边为帧差的完全连通图的方法来提取关键帧。实验结果表明,与其他方法相比,提出的方法平均精确率较高,并且平均关键帧数目较低,可以很好地提取视频的关键帧。

关键词 子镜头分割,关键帧提取,双阈值滑动窗口,完全连通图
中图法分类号 TP319.41 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.6.057

Method of Key Frames Extraction Based on Double-threshold Values Sliding Window Sub-shot Segmentation and Fully Connected Graph

ZHONG Xian[1] YANG Guang[1] LU Yan-sheng[2]
(School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, China)[1]
(School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China)[2]

Abstract With the development of multimedia technology, multimedia information is becoming more and more common in our life and work. How to efficiently retrieve the useful information in massive amounts of video information is becoming a more and more serious problem. In order to solve the above problems, this paper presented a method of key frames extraction from key frames based on double-threshold sliding window sub shot segmentation and fully connected graph. Firstly, it uses the double-threshold-based shot segmentation method, and gets the mutation boundary and the gradient boundary of a shot by setting double-threshold sliding window in order to divide the shot. Then, it uses the sub-shot segmentation method based on sliding window that adds a sliding window to the video frame sequence, and divides the shot again according to the frame differences at the range of a certain window. Finally, it uses key frame extraction method based on sub-shot segmentation, which regards the sub-shot as a fully connected graph. In this graph, the vertex is treated as a frame, the edge is treated as the frame difference so as to extract the key frames. The experimental results show that the method proposed in this paper has higher average accuracy and a less average number of key frames compared to baselines. Therefore, we can use this method to extract the key frames of video efficiently.

Keywords Sub shot segmentation, Key frame extraction, Double threshold values sliding window, Fully connected graph

1 引言

视频数据是一种综合了文本、图像、音频等多种特征的媒体信息,具有表现力强、信息量大等特点。面对海量的视频信息,针对其非结构化数据的特点,快速且有效地检索已经成为一项重要的研究课题[1,2]。

如图1所示,视频结构通常可以在粒度上由大到小地划分为视频、场景、镜头和帧等几个层次。帧是一幅独立的静态图像,是视频的最小组成单元,由于视觉暂留现象,常见帧率FPS(Frame Per Second)中24Hz/25Hz/30Hz的帧序列会让视觉以为是连贯的。镜头是摄像机不间断拍摄的时间上连续的帧序列,是视频分析的基本单元,通常同一镜头内的帧差别较小,涵盖了少量的语义信息。场景是语义上相关和时间上相近的一系列镜头的集合,涵盖了基于事件的高层语义。

到稿日期:2015-05-21 返修日期:2015-09-07 本文受国家自然科学基金项目(61003130),武汉市创新团队项目(201307020402005)资助。
钟 忺(1985—),男,博士,讲师,主要研究方向为视频语义检索、大数据分析,E-mail:345128484@qq.com;杨 光(1989—),男,硕士,主要研究方向为大数据挖掘、云计算;卢炎生(1950—),男,博士生导师,主要研究方向为数据库系统、大数据分析。

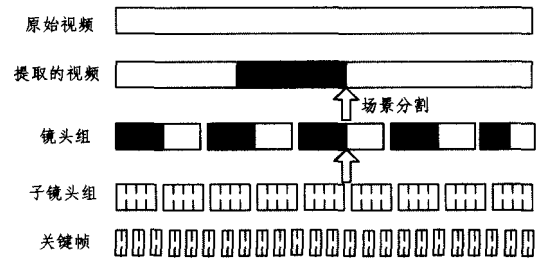


图 1　视频结构示意图

视频帧序列的帧之间存在时间和空间的冗余。关键帧是代表镜头的图像，关键帧的特征在不考虑运动特征的情况下可以代表镜头的特征，方便其进一步的结构化。提取关键帧的原则是宁错勿少，能够较为完整和准确地代表镜头的主要内容是关键帧提取的基本要求[3]。

关键帧提取可以在压缩域或非压缩域进行。前者并不解码视频编码，而是使用诸如 DCT 系数、运动矢量、宏块等信息直接提取关键帧，但是丢失了视频的部分语义信息。后者大致可以分为关键帧选取和关键帧重组两类。关键帧选取是在镜头帧序列中选择一幅或多幅帧直接作为关键帧；关键帧重组则是基于镜头的特征，构造新的图作为关键帧。

最早的关键帧提取方法是简单地选取固定位置的帧作为关键帧。Nagasaka 等人提出了选取镜头的起始帧、中间帧和终止帖作为关键帧[4]。Ardizzone 等人提出了选取镜头每秒的第 R 帧作为关键帧[5]。像素平均法和直方图平均法等方法都是选取与镜头内所有帧的平均值最接近的帧作为关键帧。Zhang 等人提出根据帧差即帧之间的特征距离来选取关键帧[6]。这类方法不带有足够的语义信息。

基于运动分析的关键帧提取方法通常选取镜头内对象运动最强或者最弱的帧作为关键帧[7]。Liu 等人构造了基于运动能量的模型，选取运动速度最快的帧作为关键帧[8]。Ma 等人选取运动加速度最大的帧作为关键帧[9]。Wolf 使用光流法，选取实时运动量局部最小值对应的帧作为关键帧[10]。Divakaran 等人在光流法的基础上，引入了 MPEG-7 的运动行为强度描述子[11]。这类方法计算量较大。

关键帧的提取可以引入图论、曲线分割和聚类等方法，将帧看作特征空间中的点，关键帧提取则转化为在这些点之中选取一个或者几个点，使得其和其他所有的点的距离都小于阈值[12]。Chang 等人将镜头看作完全邻接图的顶点，关键帧提取转化为邻接图的顶点覆盖问题[13]。Zhao 等人将镜头看作特征空间中的点的连线，关键帧提取转化为曲线分割问题[14]。Hanjalic、Yang 和 Zhou 等人通过聚类的方法提取关键帧[15-17]。Lo 等人提出了基于直方图的自适应阈值 C-means 算法[18]。

上述研究中分别提出了一些关键帧的提取算法，这些方法或者计算量较大，或者不太支持语义特征提取和检索。为了有效地解决这些问题，设计了如图 2 所示的基于特征提取的关键帧提取算法流程，该算法可以自顶向下地划分为镜头分割、基于滑动双窗口的子镜头分割和基于完全连通图的关键帧提取 3 个模块，特征提取是这 3 个模块的基础。

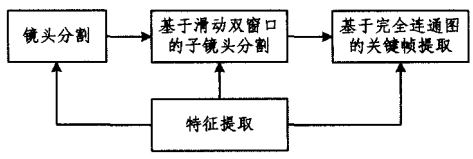


图 2　视频关键帧提取的流程

近年来，学者们围绕视频关键帧提取的课题作了大量的研究工作。面对海量的视频信息，针对其非结构化数据的特点，本文提出基于视频帧图像特征对帧序列分割子镜头，得出了基于双阈值滑动窗口子镜头分割算法 SBDW，并提出基于子镜头的关键帧提取算法 KFES。这种方法能够更加准确、更具代表性地进行关键帧提取。目前，通过子镜头分割和完全连通图的方法来完善关键帧提取已经成为众多学者的研究方向。

2　镜头分割

镜头分割算法已经相对成熟，基于双阈值的方法应用最为广泛，最早由 Zhang 等人提出[19]。

算法设置了两个阈值 T_b、T_s（$T_b>T_s$），从镜头起始帧开始逐次计算相邻帧差，当相邻帧差大于 T_b 时，认为是镜头的突变边界；当相邻帧差大于 T_s 时，认为可能是镜头的渐变边界，从此帧开始计算相隔帧差，相隔帧差一旦累积到大于 T_b，而此时的相邻帧差小于 T_s，则认为是镜头的渐变边界。

3　基于滑动窗口的子镜头分割算法

因为有长镜头这样背景变化较为频繁、内容涵盖较为复杂的镜头的存在，所以将镜头再次割成背景和内容尽量单一的子镜头，可以更加便于基于语义提取关键帧。由于子镜头分割是在镜头内进行的再次分割，使用镜头分割算法无法得到理想的结果。同时，常规的聚类分类算法也不能直接使用，因为它们忽略了镜头的有序性。

子镜头分割同样利用了帧差，与镜头分割计算和基准帧的帧差不同的是，子镜头分割无法避免计算所有帧之间的帧差，时间复杂度是 $O(n^2)$。当镜头内的帧数较多时，$O(n^2)$ 的时间复杂度可能影响到分割的执行效率。本文提出了一种可行的解决方案，给视频帧序列加上一个滑动窗口，在窗口的小范围内找寻可能的子镜头边界，该方案也更能符合流媒体格式的实时需要。

定义 1　在两类问题上，Fisher 线性判别的思想是将多维空间的数据样本投影到一维空间上，使得两个类的类间方差尽量大，而类内方差尽量小。

不妨设两个类分别为 C_1 和 C_2，样本为 x，定义这两个类的均值 m_1 和 m_2，以及方差 s_1 和 s_2。类间方差和类内方差的比值为 J。

$$m_k=\frac{1}{N_k}\sum_{n\in C_k}^{\circ} x_n,k=0,1 \tag{1}$$

$$s_k^2=\sum_{n\in C_k}^{\circ}(x_n-m_k)^2,k=0,1 \tag{2}$$

$$J=\frac{(m_2-m_1)^2}{s_1^2+s_2^2} \tag{3}$$

使得不同类的类间方差尽量大，而类内方差尽量小，即使得两者的比值 J 最大。J 最大时，即在一维空间中将两个类最优划分。

Fisher 准则是类间方差和类内方差的比值，滑动窗口内使得 Fisher 准则函数最大的划分是局部最优解，可以认为是子镜头边界。滑动窗口的大小要满足一个假设，窗口不横跨 3 个子镜头，即窗口内是一个两类问题，这样就把一个多类问

题转化成了多个两类问题，进而降低了复杂度和计算量。

本文的算法不直接使用 Fisher 线性判别在滑动窗口内尝试寻找最佳的划分，而是从正中将窗口天然地划分成两个部分，由于镜头内连续帧之间的帧差相对较小，这两个部分都可以当做类来处理。在理想状态下，子镜头内部的 Fisher 准则较小，趋近于 0，而摄像机或者物体运动导致子镜头边界处的 Fisher 准则较大。实际上，子镜头内部存在因为微小波动导致的局部极大值，应当被视为噪音，解决噪音问题的一个简单的方法是选择高于 Fisher 准则均值的局部极大值作为子镜头边界。

算法 1 基于滑动窗口的子镜头分割算法 SBDW (Sub shot Boundary Detection based on sliding Window)

第 1 步 输入镜头边界，直到镜头结束。

第 2 步 计算第 i+1 帧到第 i+M 帧的均值，第 i+M+1 帧到第 i+2M 帧的均值。

第 3 步 计算第 i+1 帧到第 i+M 帧的方差，第 i+M+1 帧到第 i+2M 帧的方差。

第 4 步 利用 Fisher 准则求出 J_i 的均值。

第 5 步 输出子镜头边界。

SBDW 算法的时间复杂度分析：$T(N)=O(N-2M)*(O(M)+O(M))+O(N-2M)+O(N-2M)=O(2MN+2N-2M^2-4M)$，由于 $2M \ll N$，SBDW 算法的时间复杂度 $T(N)=O(N)$。

4 基于完全连通图的关键帧提取

下面讨论在子镜头分割的基础上提取关键帧，把子镜头看作完全连通图，帧看作图的顶点，帧差则看作连接顶点的边。这样就可以将关键帧提取的问题转化为在完全连通图中求图中心的问题。

定义 2 设 G 是有 n 个顶点的无向连通图，顶点 v_i 和其最远的顶点的距离称为 v_i 的半径，记作 $\rho(v_i)$，即

$$r(v_i)=\max_{1\neq j\, n}\{d(v_i,v_j)\} \tag{4}$$

半径最小的顶点称为图的中心，记作 v_G，无向连通图 G 的半径是图的半径，记作 ρ_G，即

$$r_G=\max_{1\neq j\, n}\{d(v_i,v_j)\}=\min_{1\neq j\, n}\{d(v_i,v_j)\} \tag{5}$$

图中心是邻接图上半径最小的点，对应的帧则是与子镜头内所有其他帧的最大帧差最小的帧，可以认为是代表了子镜头的内容，视为关键帧。

算法 2 基于子镜头分割的关键帧提取算法 KFES (Key Frame Extraction based on Sub shot)

第 1 步 输入子镜头边界，直到镜头结束。

第 2 步 计算 i，j 两帧的帧差。

第 3 步 计算顶点 i 的半径。

第 4 步 输出子镜头边界。

KFES 算法的时间复杂分析：$T(N)=O(N)*O(N)+O(N)*O(N)=O(2N^2)=O(N^2)$，KFES 算法时间复杂 $T(N)=O(N^2)$。

5 实验分析

实验环境：windows7、Visual Studio 2010、C++。

数据来源：OpenCV 示例视频、宜昌市街道监控视频以及优酷等视频网站抓取的军事相关视频。

5.1 提取关键帧

(1)运动视频提取的关键帧

对一个 425 帧的运动视频分为 4 部分提取，每部分提取的关键帧如图 3 所示。

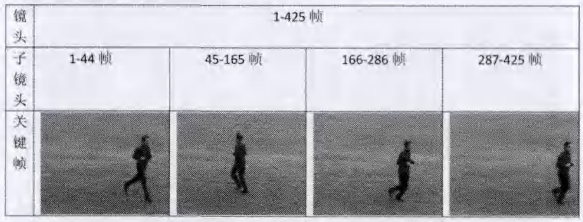


图 3 运动视频提取的关键帧

(2)交通视频提取的关键帧

针对一个长度为 165 帧的视频分为两部分提取关键帧，结果如图 4 所示。

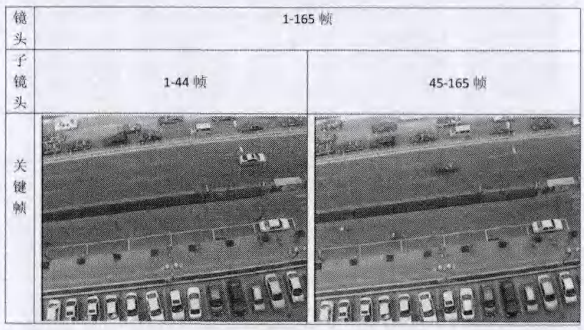


图 4 交通视频提取的关键帧

(3)网络军事视频提取的关键帧

图 5 原视频 41s，3.48MB，1519 帧，提取关键帧 49 帧。图 6 原视频 25s，1.90MB，1275 帧，提取关键帧 68 帧。

图 5 网络军事视频提取的关键帧(1)

图 6　网络军事视频提取的关键帧(2)

以上多组实验结果表明，提取的视频帧具有代表性，本文提出的方法能成功地提取关键帧。

5.2　与其他方法的比较

从人工检测的角度来看，本文所用的关键帧提取算法与其他算法相比，保持了视频的主要内容，没有遗漏主要的场景或者事件，关键帧的内容没有过于冗余，数量也是合适的。在计算方面，本方法较大幅度地减少了计算量，可以为语义特征的提取和检索提供支持。这主要体现在子镜头分割上面，视频具有时序性，不能简单地聚类分类。使用 Fisher 线性判别的投影的思想，保持被分割部分的时序性。再使用滑动窗口技术，不仅可以实时并行处理镜头分割、子镜头分割和关键帧提取，更符合流媒体的需要，而且将多类问题转化成多个两类问题进行处理，简化了计算，使得计算量较其他方法有很大的减小。

而在语义特征提取方面，本方法提取的关键帧更具有视频的代表性，更加能够反映视频的主要内容；再根据本文的方法，能够在下一步对关键帧提取语义特征，为视频语义检索提供技术支持，这一方面是其他算法目前还不能实现的。

数据来源：城市交通视频来源于宜昌市某干道监控视频，隧道监控视频来源武汉市水果湖隧道监控视频，运动视频来源于 OpenCV 示例视频，军事视频来源于优酷等视频网站。

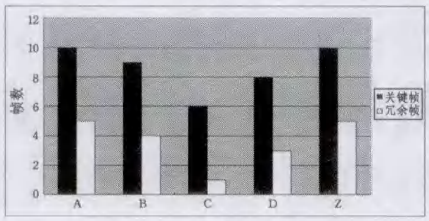
用本文方法(Z)与 DiantingLiu 方法 A[20]、Rong Pan 方法 B(带子镜头分割)[21]、Yuanfen Yan 方法 C(带子镜头分割)[22]及 Lei Pan 方法 D(带子镜头分割)[23]进行比较，结果如表 1 所列。

表 1　视频关键帧提取比较

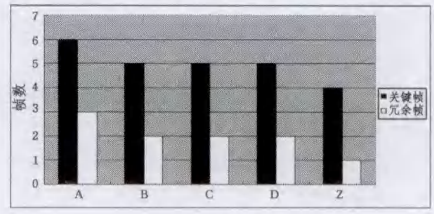
	A			B			C			D			Z		
	A1	A2	A3(%)	B1	B2	B3(%)	C1	C2	C3(%)	D1	D2	D3(%)	Z1	Z2	Z3(%)
城市交通视频	10	5	50	9	4	56	6	1	84	8	3	63	10	5	50
隧道监控视频	6	3	50	5	2	60	5	2	60	5	2	60	4	1	75
运动场视频	11	8	28	6	3	50	9	5	45	4	1	75	4	1	75
网络军事视频	80	42	48	59	21	65	61	23	63	77	29	63	56	18	68

表中，A1、A2、A3 对应 A 方法提取的关键帧数目、冗余关键帧数目及精确率，其他类同。由表 1 可知，本文方法的平均精确率较高，平均关键帧数目较低。可以增加视频中运动对象的数量以及运动幅度分类，这样才能表达更普适的结论，以此表明各个方法在不同情况下的优劣。

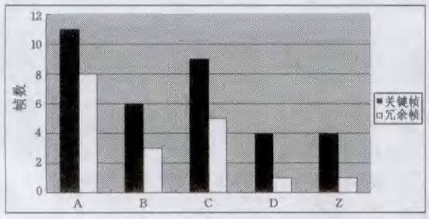
图 7 和图 8 为表 1 的直方图表示，结果更直观。



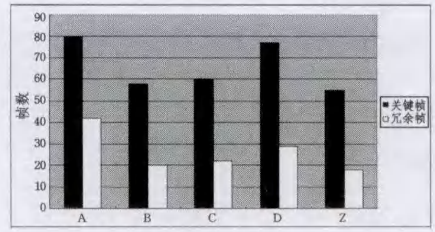
(a)市交通视频直方图



(b)隧道监控视频直方图



(c)运动视频直方图



(d) 网络军事视频直方图

图 7　视频关键帧提取比较直方图

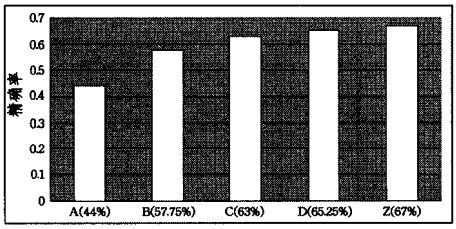


图 8　平均精确率比较直方图

结束语　本文针对视频帧序列的关键帧提取进行了研究，设计了一个基于滑动窗口的子镜头分割算法 SBDW，其可以有效地保持视频帧序列的时序性，并将多类问题转化为多个两类问题，降低了算法时间复杂度并减少了计算量。接着提出了一种基于子镜头分割的关键帧提取算法 KFES，它可将关键帧提取转化为图的顶点覆盖问题，简化了计算，为语义特征的提取和检索提供了支持。整个方法在保持视频的主要内容的同时未遗漏主要的场景或者事件，使关键帧的内容避免了过于冗余的现象，获取的关键帧的数量也比较合适。通过实验验证了该方法的有效性，进一步的对比实验也表明该方法具有较高的平均精确率和较低的平均关键帧数目。今后将结合视频语义检索方法，对提取的关键帧进行进一步的语义分析。

参 考 文 献

[1] Dai Ke-xue, Li Qiang, Li Guo-hui. Video Mining Research [J]. Computer Science, 2010, 37(10): 11-15(in Chinese)
代科学，李强，李国辉. 视频挖掘研究进展[J]. 计算机科学，2010，37(10)：11-15

[2] Wei Wei, You Jing, Liu Feng-yu, et al. Semantic video retrieval Review [J] Computer Science, 2006, 33(2): 1-7(in Chinese)
魏维，游静，刘凤玉，等. 语义视频检索综述[J]. 计算机科学，2006，33(2)：1-7

[3] Qu Zhong, Gao Teng-fei, Zhang Qing-qing. An improved video key frame extraction algorithm [J]. Computer Science, 2012, 39 (8): 300-303(in Chinese)
瞿中，高腾飞，张庆庆. 一种改进的视频关键帧提取算法研究[J]. 计算机科学，2012，39(8)：300-303

[4] Nagasaka A, Tanaka Y. Automatic video indexing and full-video search for object appearances [C] // Proceedings of the IFIP TC2/WG2. 6 2nd Working Conference on Visual Database Systems II. 1992: 113-127

[5] Ardizzone E, La Cascia M. Automatic video database indexing and retrieval [J]. Multimedia Tools and Applications, 1997, 4 (1): 29-56

[6] Zhang H J, Wu J, Zhong D, et al. An integrated system for content-based video retrieval and browsing [J]. Pattern Recognition, 1997, 30(4): 643-658

[7] Zeng Wei, Xue Xiang-yang. Review retrieve video information based on the motion characteristics[J]. Computer Science, 2004, 31 (2): 135-138(in Chinese)
曾玮，薛向阳. 基于运动特征的视频信息检索综述[J]. 计算机科学，2004，31(2)：135-138

[8] Liu T, Zhang H J, Qi F. A novel video key-frame-extraction algorithm based on perceived motion energy model [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(10): 1006-1013

[9] Ma Y, Chang Y, Yuan H. Key-frame extraction based on motion acceleration[J]. Optical Engineering, 2008, 47(9): 957-966

[10] Wolf W. Key frame selection by motion analysis [C]// Conference Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP96). 1996, 2: 1228-1231

[11] Divakaran A, Radhakrishnan R, Peker K A. Video summarization using descriptors of motion activity: A motion activity based approach to key-frame extraction from video shots [J]. Journal of Electronic Imaging, 2001, 10(4): 909-916

[12] Hui Wen, Zhao Hai-ying, Lin Chuang, et al. Video forensics research based on content [J]. Computer Science, 2012, 39 (1): 27-31(in Chinese)
惠雯，赵海英，林闯，等. 基于内容的视频取证研究[J]. 计算机科学，2012，39(1)：27-31

[13] Chang H S, Sull S, Lee S U. Efficient video indexing scheme for content-based retrieval [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1999, 9(8): 1269-1279

[14] Zhao L, Qi W, Li S Z, et al. Key-frame extraction and shot retrieval using nearest feature line (NFL)[C]//Proceedings of the 2000 ACM workshops on Multimedia. 2000: 217-220

[15] Hanjalic A, Zhang H J. An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1999, 9(8): 1280-1289

[16] Yang S, Lin X. Key frame extraction using unsupervised clustering based on a statistical model [J]. Tsinghua Science & Technology, 2005, 10(2): 169-173

[17] Doucet A, Godsill S, Andrieu C. On sequential Monte Carlo sampling methods for Bayesian filtering[J]. Statistics and Computing, 2000, 10(3): 197-208

[18] Lo C C, Wang S J. Video segmentation using a histogram-based fuzzy c-means clustering algorithm[J]. Computer Standards & Interfaces, 2001, 23(5): 429-438

[19] Zhang H J, Kankanhalli A, Smoliar S W. Automatic partitioning of full-motion video [J]. Multimedia Systems, 1993, 1(1): 10-28

[20] Liu D, Shyu M L, Chen C, et al. Within and between shot information utilization in video key frame extraction[J]. Journal of Information & Knowledge Management, 2011, 10(03): 247-259

[21] Pan R, Tian Y, Wang Z. Key-frame extraction based on clustering[C]// IEEE International Conference on Progress in Informatics and Computing (PIC). 2010, 2: 867-871

[22] Yang Y, Cui Z, Wu J, et al. Traffic Video Segmentation and Key Frame Extraction Using Improved Global K-Means Clustering [C] // IEEE International Symposium on Information Science and Engineering (ISISE). 2010: 521-525

[23] Pan L, Wu X, Shu X. Key Frame Extraction Based on Sub-Shot Segmentation and Entropy Computing[C]// IEEE Chinese Conference on Pattern Recognition (CCPR). 2009: 1-5