计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2015.20140701
52(12):2879-2887, 2015

基于支配集的视频关键帧提取方法

聂秀山[1,2] 柴彦娥[1] 滕 聪[3]
[1](山东财经大学计算机科学与技术学院 济南 250014)
[2](山东省金融信息工程技术研究中心 济南 250014)
[3](山东财经大学数学与计量经济学院 济南 250014)
(niexsh@sdufe.edu.cn)

Keyframe Extraction Method Based on Dominating Set

Nie Xiushan[1,2], Chai Yan'e[1], and Teng Cong[3]
[1](*School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan* 250014)
[2](*Shandong Engineering & Technology Center of Financial Informatization, Jinan* 250014)
[3](*School of Mathematic and Quantitative Economics, Shandong University of Finance and Economics, Jinan* 250014)

Abstract Keyframe extraction is one of the important steps in video processing, and it is popularly used in video content analysis. A keyframe extraction method is proposed in this paper for the content-based video summarization. In order to well depict the structure and relation among the frames of a video, we firstly model the video as an undirected weighted graph, where the frames of the video are taken as vertices, and the lines among vertices are taken as edges. The weights of edges are computed using the Hausdorff distances between pairs of speed-up features frame-by-frame which are local and robust features of frames. Subsequently, based on the representation of the keyframe, the process of keyframe extraction is equivalent to the selection of minimum dominating set in a graph, and integral linear programming is used to select the minimum dominating set in the graph. Finally, the keyframes are extracted according to the vertices in the obtained dominating set. We execute the proposed method on different types of videos, and evaluate the performance of the fidelity and compression ratios. Compared with the traditional methods, the proposed method is depended on video content rather than time and video shots. The experimental results show that the keyframes extracted by the proposed method have good representation and discrimination, and they also have high fidelity and compression ratios.

Key words video summarization; keyframe extraction; graph modeling; dominating set; integral programming; speed-up robust feature

摘 要 关键帧提取是视频处理的重要步骤之一,在视频内容分析中有广泛的应用.针对基于内容的视频分析,为获取高效的视频摘要提出一种视频关键帧提取方法.该方法首先以视频帧为顶点,以顶点之间的连线构造边,利用不同帧的加速鲁棒特征点的豪斯多夫(Hausdorff)距离函数计算边权重,把视频建模成一个无向权重图,然后根据图的支配集理论把视频关键帧提取等价为无向权重图的极小支配集选取问题,进而利用整数线性规划选取图支配集,得到视频关键帧.与传统算法相比,该方法提取的关键帧依赖于视频内容,不受时间和视频镜头约束.实验结果显示,该方法能够体现关键帧的代表性和区分性,具有较高的保真度和压缩率.

收稿日期:2014-07-27;**修回日期**:2015-04-16
基金项目:山东省优秀中青年科学家奖励基金项目(BS2013DX013);国家自然科学基金项目(61101162)

关键词 视频摘要;关键帧提取;图建模;支配集;整数规划;加速鲁棒特征

中图法分类号 TP391.41

随着计算机与多媒体技术的发展,以图像、声音和视频为载体的多媒体数据越来越多,特别是视频数据,因其信息量大、表现力强,逐渐成为人们进行信息交流的主要媒体形式.另外,随着软硬件技术以及网络技术的飞速发展,视频数量急剧增加,越来越多的人选择使用计算机或手机等移动设备观看视频.大量的视频数据亟需高效的视频内容管理方式,从而给用户更好的多媒体体验.进一步地,随着微博、人人网等社交媒体形式地普及,视频的分享量更是急剧增加,如何能让用户在快节奏的生活方式下快速地浏览视频内容是一个重要的研究课题,视频摘要就是一种可行的解决途径,视频摘要是视频内容的一种典型且有效的表达方式,是对视频内容的高度概括[1].通过视频摘要,用户可以在浏览少量内容的情况下对视频有基本的了解,视频摘要分为静态视频摘要和动态视频摘要 2 种形式,关键帧就是静态视频摘要的重要形式.因此,研究者一直在努力开展关键帧提取技术的研究.另一方面,由于视频数据的几何式增长,视频检索在多媒体处理领域中越来越重要,传统的视频检索主要是依靠文本标注来实现,这种方法工作量大、效率低,而且主观性较大,因此一种自动、客观、全面的视频检索方式——基于内容的视频检索是近年来的一个研究重点[2].基于内容的视频检索的一个重要步骤就是从视频序列中提取关键帧,并以关键帧为索引对原始内容进行检索.因此,关键帧提取在基于内容的视频检索中有着重要的作用.本文就是基于视频摘要和视频内容检索这 2 个应用背景,提出一种视频关键帧提取的方案.

传统的视频关键帧提取的方法大致分为 2 大类:1)基于采样的关键帧提取方法[3-4].这类方法采用随机或均匀抽样的方式得到关键帧,这类方法虽然简单快捷,但是可能会导致一些重要的视频片段没有选到关键帧,或者是一些片段取到重复的关键帧.2)基于镜头分割的关键帧提取方法[5].这类方法把视频分成若干个视频镜头,然后选取每个镜头的首帧或末帧作为关键帧,此类方法受限于镜头分割的精度,同时,此类方法并不能完全体现视频镜头的内容.近年来,一些基于视频内容和语义特征的关键提取方法陆续被提出.王方石等人[6]利用阈值和聚类的方法选取关键帧;Naveed 等人[7]提出了根据运动和空间显著度,并加权融合提取关键帧的方法;Zhang 等人[8]通过运动能量来实现关键帧的选取;Kin 等人[9]提出了基于时域最大值的关键帧提取算法.文献[10-11]分别结合人类视觉系统的视觉注意模型进行关键帧的提取.但是,现有算法在关键帧提取过程中,大多受时间先后的影响,即相同或相似的视频帧,因在不同的时间点仍然可能同时被选为关键帧,而对于静态视频摘要来说,用户的目的是快速了解视频的基本内容,对事件的时间性要求不是很高,因此,这些方法得到的关键帧对于用户快速浏览视频内容来说还是冗余的.另外,现有的关键帧提取算法大多以镜头的分割或聚类为前提,然后结合其他工具或特征来提取关键帧,但镜头分割不仅计算复杂度高,而且易受噪声、光线变化等参数影响,而聚类方法又易受到聚类阈值的影响,为解决上述问题,本文提出一种基于支配集的关键帧提取方法,该方法以视频内容为依据、以图论理论为工具,完全抛开时间性和视频镜头的约束,选取表达不同内容的帧作为关键帧.

1 基于支配集的关键帧提取方法

本文方法的框架图如图 1 所示,主要包括 2 部分:视频图建模和视频关键帧提取.本方法首先把原始视频等价为无向权重图,其中权重由不同帧之间加速鲁棒特征点(speed-up robust feature, SURF)的豪斯多夫(Hausdorff)距离函数表示;然后利用整数规划在该图上选取极小支配集,进而得到关键帧.具体内容在以下各节描述.

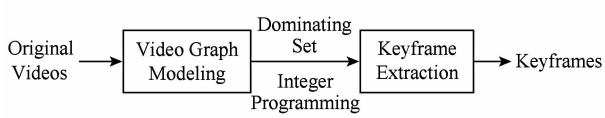


Fig. 1 Framework of keyframe extraction.

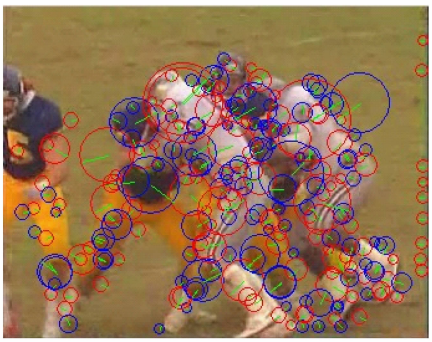
图 1 关键帧提取框架图

1.1 视频图建模

图在表达不同数据点之间的关系和可视化分析方面有天然的优势,而且把视频序列建模成图,一些图论的工具和方法可以很好地应用其中.因此,本文首先对视频进行图建模,视频的每一帧抽象为 1 个顶点,顶点之间连线构成边,则视频可以看作平面里

的 1 个无向权重图 $G=(V,E,W)$,其中 $V=\{v_i\}$,$E=\{e_{ij}\}$ 和 $W=\{w_{ij}\}$ 分别代表图顶点集、边集和边权重的集合,为方便描述,本文把这个图称为相应视频的视频映射图. 其中,映射图的边权重反映了 2 个顶点之间相似的程度,即 2 帧之间的相似程度. 因此,边权重的定义是视频图建模的关键. 众所周知,图像的局部特征能够较好地描述图像的内容,其中 Harris 角点、尺度不变特征点(scale invariant feature transform, SIFT)等都是常用的图像局部特征描述器,但是 Harris 角点对放缩变换的鲁棒性差,SIFT 虽然在描述图像局部特征上具有良好的效果,但缺点是计算复杂度太高,因此,本文采用另外一种特征点——加速鲁棒特征(SURF)——来计算权重. SURF 除了具有良好的性能之外,最大的优点是速度快,这对视频处理是非常重要的. SURF 算法主要包括图像极值点检测和特征表述 2 个部分. 其主要思路是:首先利用 Hessian 矩阵行列式的值来判断角点是否是极值点;其次,为描述局部极值点,SURF 算法在 4×4 的块上计算采样点的 Haar 小波响应,并把水平和垂直方向的小波响应值以及其绝对值累加值形成一个 4 维向量,这样共得到 64 维的特征向量用来描述图像局部特征点. 详细内容参考文献[12]. 图 2 给出了 2 个视频帧的 SURF 点示意图.

(a) SURF of frame "excavator"



(b) SURF of frame "fooball"

the line is the principal direction of the corresponding point, and the circles represent the SURF points with different signs of Laplacian, which can be seen in Ref [12].

Fig. 2 Illustration of SURF.

图 2 SURF 算法提取特征点示意图

视频映射图的边权重体现了 2 帧的相似程度,权重越大,2 帧的相似性越大,反之相似性越小. 显然,每一个视频帧可以看作其 SURF 特征点集合,因此帧的相似性即特征点集合的相似性就可以用 2 个集合的距离来描述,Hausdorff 距离是刻画集合之间距离常用的度量方式,因此本文采用 Hausdorff 距离来刻画 2 帧的相似性. 其定义如下:

设 $A=\{a_1,a_2,\cdots,a_M\}$ 和 $B=\{b_1,b_2,\cdots,b_N\}$ 是 2 个有限点集合,则 A 和 B 的 Hausdorff 距离 $H(A,B)$ 定义为

$$H(A,B)=\max\{h(A,B),h(B,A)\}, \quad (1)$$

其中,$h(A,B)=\max\limits_{a\in A}\min\limits_{b\in B}\|a-b\|$,$h(B,A)=\max\limits_{b\in B}\min\limits_{a\in A}\|b-a\|$.

由定义可知,Hausdorff 距离越大,2 个集合的相似性越小. 为保证视频映射图的边权重和顶点的相似性成正比关系,本文利用核函数思想[13],采用核函数 $e^{\left(-\frac{\|x_i-x_j\|^2}{\sigma}\right)}$ 生成权重,其中,x_i 和 x_j 是数据点,$\|x_i-x_j\|$ 用 Hausdorff 距离表示;σ 为参数,本文 $\sigma=1$,则视频映射图顶点 i 和 j(代表相应顶点对应视频帧的特征点集合)之间的边权重 w_{ij} 为

$$w_{ij}=e^{-H(i,j)}. \quad (2)$$

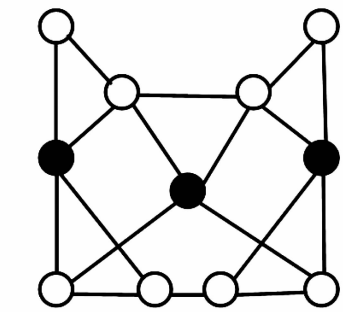
由此,视频映射图变为 1 个无向权重图,为使视频映射图更简洁地反映视频帧之间的关系,本文对映射图进行简化,若边权重小于阈值 $\alpha\times mean(W)$(说明 2 个顶点对应的视频帧的相似性较小),则移除相应边. 其中,$\alpha\in(0,1]$ 是 1 个参数,$mean(W)$ 表示所有边权重的平均值.

1.2 关键帧提取

视频关键帧的提取问题,可近似地等价为视频映射图的支配集选取问题,为说明此等价关系,首先给出支配集的定义.

定义 1. 支配集. 集合 M 是顶点集合 V 的非空子集,若 $\forall x\in V-M$,x 与 M 里的至少一个顶点相连,则称 M 是图 $G=(V,E)$ 的支配集. 若 M 的任何真子集都不是支配集,则称 M 是极小支配集. 若不存在其他支配集 M',使得 $|M'|<|M|$(其中,$|\cdot|$ 表示集合"·"中的元素个数),则称 M 为最小支配集. 图 3 给出 1 个支配集的示例,实心点对应的顶点集合组成了一个极小支配集.

所谓关键帧是指在视频帧中能够代表视频内容的少量帧的集合,换句话说,即选取的关键帧集合要覆盖所有视频帧的内容. 反映到视频映射图上就是



the black solid vertices represent a minimum dominating set

Fig. 3 Illustration of dominating set in a graph.

图 3 支配集示例图

选取较少的顶点覆盖图的所有顶点,这里"覆盖"可理解为"代表",即选取到的这类点可以代表图的所有顶点.反映到图上就是没有被选到的点要至少和一个被选到的点有边相连(边相连意味着 2 点有很强的相似性,一个可以被另一个"代表"),这类点的集合恰恰就是图的支配集,因此,我们使用支配集理论选取关键帧是可行的.显然,最小支配集对应的帧集合是一个理想的关键帧集合,但是图的最小支配集问题是一个 NP-hard 问题,因此,本文利用整数规划来选取视频映射图的一个极小支配集,从而得到视频的关键帧集合.

设 U 是视频映射图的一个极小支配集,N 是视频映射图顶点数目,定义变量 t_i 如下:

$$t_i=\begin{cases}1, & v_i\in U,\\ 0, & v_i\notin U,\end{cases} \tag{3}$$

t_i 可以看作顶点 v_i 的标号,即若顶点 v_i 属于极小支配集集合,其标号为 1,否则其标号为 0.整数规划模型定义如下:

$$\min\sum_{i=1}^{N}t_i,\quad \text{s.t. } \boldsymbol{A}\boldsymbol{t}\geqslant\boldsymbol{B}, \tag{4}$$

其中,$\boldsymbol{t}=(t_1,t_2,\cdots,t_N)^{\mathrm{T}}$ 是所有顶点标号组成的向量,$\boldsymbol{B}$ 是一个全为 1 的列向量,$\boldsymbol{A}$ 的定义如下:

$$\boldsymbol{A}=(a_{ij})_{N\times N}, \tag{5}$$

其中,$a_{ij}=\begin{cases}1, & e_{ij}\in E\\ 0, & \text{otherwise}\end{cases}$,即顶点 v_i 和 v_j 之间有边相连,则 $a_{ij}=1$,否则 $a_{ij}=0$.特别地,$a_{ii}=1$,此整数规划问题保证了选取较少的支配集,对应于关键帧对原始视频的帧压缩特性;其中,约束条件保证了未被选入支配集的顶点至少和支配集中的一个顶点有边相连,保证了所选集合的代表性.

式(4)可以容易地转化为整数规划的标准形式(6)求解.

$$\min\sum_{i=1}^{N}t_i,\quad \text{s.t. } -\boldsymbol{A}\boldsymbol{t}\leqslant-\boldsymbol{B}. \tag{6}$$

1.3 关键帧数目的讨论

视频关键帧的数目是关键帧提取方法的一个重要问题,关键帧提取本质上是选择能够代表视频内容的帧.如果关键帧数目太多,虽然较高程度地体现了视频的内容,但增加了视频检索的计算量,而且也在某种程度上失去了关键帧的意义(选取关键帧的目的是为了简洁地表示视频);如果关键帧数目太少,则不能完全体现视频的内容.本文把关键帧集合等价为图的极小支配集,这就保证了关键帧的内容代表性(因为非关键帧都至少与 1 个关键帧相连),关键帧的数目即为图极小支配集中的顶点数,定量地表示 1 个图极小支配集的顶点数是很难的,但其上界却是可控的.根据定理 1,可以得到图支配集中顶点数的上界.

定理 1[14]. 最小度为 k 的 n-顶点的图有 1 个大小至多为 $n\dfrac{1+\ln(k+1)}{k+1}$ 的支配集.

所谓图顶点的"度"是指与该顶点相连边的数目.由定理 1 可以看出,随着图顶点最小度的增加,支配集中顶点数的上界是逐渐减小的,如图 4 所示.这恰好与实际情况相符,对于 1 个内容变化比较小的视频来说,关键帧的数目应该是较少的,这类视频不同帧之间的相似性较大,反映到视频映射图上,映射图的边就比较多,因此每个顶点的度就会较大,根据定理 1 我们得到的支配集中顶点的数目应该也是较小的.

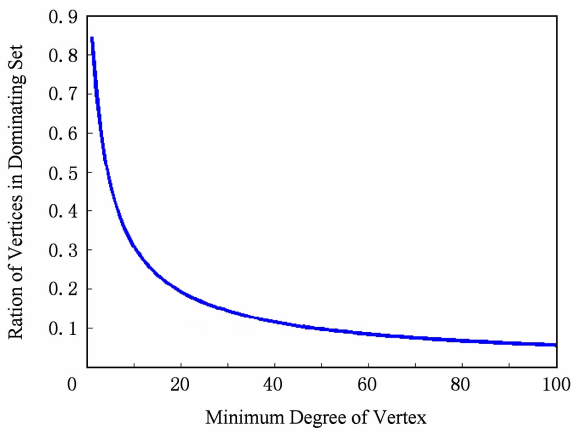


Fig. 4 Relation between the upper bound of dominating set and the degree of vertices.

图 4 支配集顶点数上界与图顶点最小度的关系

定理 1 给出了一个支配集中顶点数的上界,更进一步地,文献[15]给出了最小支配集内顶点数的一个上界:当图顶点的最小度大于 2 时,最小支配集中的顶点数不大于 $0.4n$. 对于视频来说,因为其内容的连续性,所以大多数视频映射图中的顶点度都是大于 2 的. 因此,利用支配集理论得到的关键帧的数目在保证视频内容代表性的基础上,又具有很好的视频压缩性. 但是,在某些视频中,个别帧的内容可能与其他帧的内容差别很大,则此类帧在视频映射图上的对应顶点可能为孤立点(即没有顶点与之相连). 为保证算法的精确性,我们在求极小支配集时首先去除此类点,同时,为了保证关键帧的代表性,最后再把这些点对应的帧加入关键帧集合.

2 实验结果与分析

2.1 实验设定以及评价标准

本文实验的视频来源于公开视频数据库 http://www.open-video.org 和研究组从视频网站上自主下载的视频组合而成,视频的类型包括民俗、体育和新闻等不同类型的视频,视频的帧数在 50～1 000 之间. 如何评价关键帧提取算法性能的优劣是一个开放性的问题,现有的评价方法分为主观评价和客观评价 2 种形式. 所谓主观评价即用户对提取出的关键帧的直观感受,用户的主观感受虽然是一个直观的评价方法,但这容易受到外界因素以及用户主观偏好的影响,这种方法不能实现自动化判别,而且效率比较低,因此对算法定量的客观评判是必要的. 本文采用保真度和压缩率作为性能评价的标准,保真度越高说明关键帧越能代表视频的整体内容;压缩率越高说明关键帧越简洁. 保真度和压缩率的定义如下[16]:

设原始视频的帧集合为 $F=\{f_1,f_2,\cdots,f_N\}$,关键帧集合为 $K=\{k_1,k_2,\cdots,k_M\}$,N 和 M 分别表示原始视频帧的数目和关键帧的数目,则视频帧 f_i 到关键帧集合 K 的距离 $D(f_i,K)$ 定义为

(a) Videos with scene changing fast

(b) Videos with content changing small in a scene

(c) Videos with content changing largely in a scene

Fig. 5 Original videos of different classes.

图 5 不同类别的原始视频示意图

$D(f_i,K)=\min\{d(f_i,k_j)|j=1,2,\cdots,M\}$, (7)

其中,$d(\cdot)$是一种距离度量,本文仍然用 2 个不同帧的 SURF 点集之间的 Hausdorff 距离来表示.

原始视频和关键帧集合之间的距离定义如式(8):

$$D(F,K)=\max\{D(f_i,K)|i=1,2,\cdots,N\}, \quad (8)$$

保真度 P 定义如下:

$$P=e^{-D(F,K)}. \quad (9)$$

采取式(9)定义保真度基于 2 点原因:1)把度量的数值归一化,从而更容易比较不同算法的性能;2)可以使保真度与距离成反比(视频和关键帧集合的距离),更符合人的直观感觉.

压缩率的定义如下:

$$C=1-\frac{M}{N}. \quad (10)$$

为验证本文算法的性能,本文实验设计共分为 2 组:

组 1 是针对不同类型的视频关键帧提取的保真度和压缩率性能实验,主要目的是验证在不同场景或同场景中镜头变化不同的视频情况下本文算法性能的差异性.在本组实验中,取 3 类不同特点的视频作为原始数据:1)场景转换较快的视频片段.图 5(a)给出了随机从类 1 视频中选取的原始视频帧示意图,该视频主要内容是汽车在都市街道上行驶,车内场景和街道高楼、商户场景不断切换;2)同一场景下镜头内容变化很小,如图 5(b)所示,主要内容是一辆摩托车在路边景物相似的公路上行驶;3)同一场景下镜头内容变化较快的视频,如图 5(c)所示,场景是一个闹市,但随着镜头的切换,镜头内容不断改变.

组 2 实验设置为对任意视频关键帧提取性能测试,主要目的是把本文算法和已有算法的性能进行比较.对于视频内容类型不加任何限制,随机选取 10 个视频序列进行实验,分别使用本文算法,文献[7-9]的算法提取关键帧,并对平均实验结果进行比较.文献[7]是一种基于运动和空间显著度的关键帧提取算法(简称 vt-based 方法);文献[8]是一种利用运动能量提取关键帧的方法(简称 motion-based 方法);文献[9]是一个经典的时域关键帧提取算法(简称 TMOF 方法).

2.2 实验结果与分析

在组 1 的实验中,我们对 3 类不同特点的原始视频,运用本算法共进行了 5 次实验,每次实验中 3 类视频各取 10 个,并求出每类中的平均保真率和压缩率,平均保真率和压缩率的柱状图如图 6 所示.由图 6 的保真率和压缩率结果可以看出:类 2 视频的保真率和压缩率性能最好.这是因为,对类 2 视频来说,大部分帧的内容差别很小,因此,由此类帧构成的映射图中支配集的点的数目较少,而且选取的关键帧与其余帧相近,所以压缩率和保真率都较高,这也是和定理 1 的结论是一致的.类 1 和类 3 视频的平均保真率和压缩率总体相当,但是在部分实验视频序列中类 3 视频压缩率稍高于类 1 视频,主要原因是:虽然类 3 视频内容有变化,但其中很多帧因在同一个场景下有部分相似的背景内容,与类 1 视频相比,其视频映射图的支配集中顶点数目较少,因此压缩率较高.

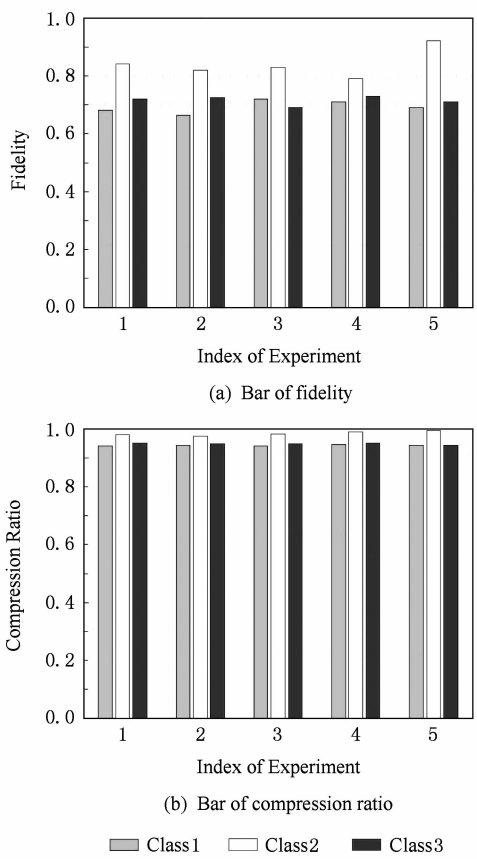


Fig. 6 The bar of fidelity and compression ratio.

图 6 保真度和压缩率柱状图

在组 2 实验的结果中,图 7(a)给出了实验示例视频 BOR08_007.mpg 的部分视频序列;图 7(b)是使用本文算法针对此部分序列提取的关键帧,可以看出,提取的关键帧反映了视频序列表达的内容.因此,从用户的主观角度来看,提取的关键帧是符合用户的主观认知的.对于保真度和压缩率的客观评价,图 8 给出了不同算法的保真度和压缩率比较.由图 8(a)可以看出,与文献[7-9]算法相比,本文的算法

在保真度上有明显地改善.压缩率方面,由图 8(b)可以看出,本文方法和文献[7-9]算法性能相当,虽然个别点的性能低于其他算法,但是 90%以上的压缩率已经能够满足实际应用的需要了.

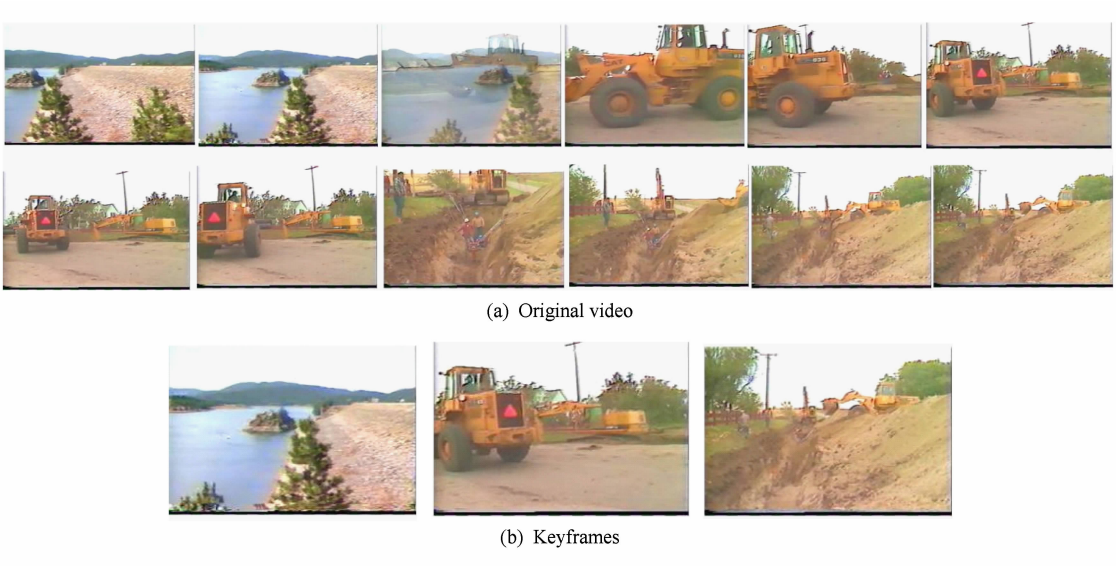


Fig. 7 Video clip and keyframes.

图 7 视频序列及其关键帧

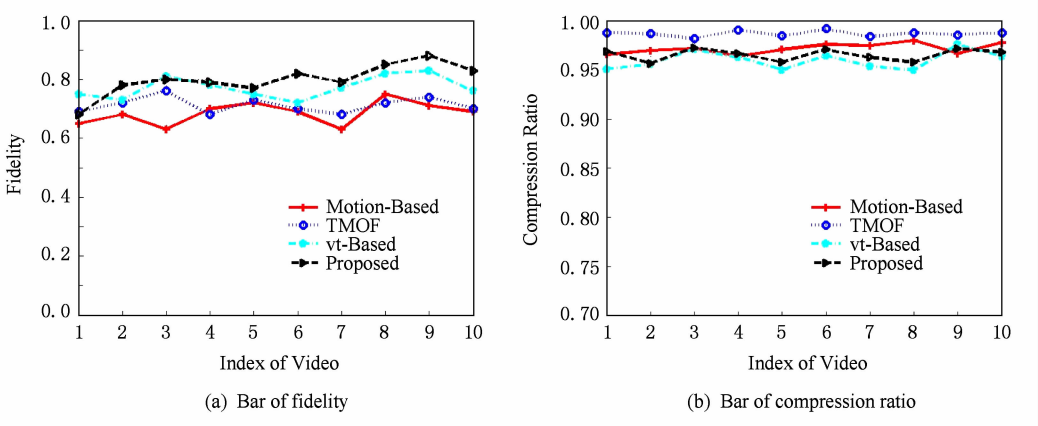


Fig. 8 Curves of fidelity and compression ratio extracted by keyframes.

图 8 关键帧提取的保真度和压缩率曲线

2.3 计算复杂度分析

本文算法的计算复杂度主要体现在视频映射图建立和整数规划的求解上,以下分别就这 2 部分的计算复杂度进行分析.

1) 视频映射图计算复杂度分析

视频映射图建立的复杂度主要体现在边权重的计算上,边权重的计算包括 SURF 特征点提取和 Hausdorff 距离计算.设视频帧的大小为 $W\times H$,对于 SURF 特征点提取来说,主要分为 2 步:①积分图像中极值点检测,对于每个像素来说计算复杂度为 $O(1)$[12],因此对每一帧的计算复杂度是 $O(WH)$;②利用 Haar 小波对特征点描述,每个特征点描述的计算复杂度为 $O(n\log n)$(通常的 SURF 算法中 $n=16$,详见文献[12]),不失一般性,设每帧产生 J 个 SURF 特征点,视频共有 N 帧,则 SURF 特征点提取总的计算复杂度为 $O(NWH+J)$. Hausdorff 距离计算的复杂度为 $O(J^2)$,则视频映射图总的计算复杂度约为 $O(NWH+J)+O(J^2)=O(NWH+J^2)$,可以看出总的计算复杂度是随着视频帧数 N 的增加呈线性增长,而总的计算复杂度是与每帧的特征点数目呈平方级关系.在实际的应用中,每帧的特征点的数目一般是在数百个的级别,如果特征点数目过多,可以舍弃部分对帧主要内容影响很小的细节点来降低计算复杂度.因此,当 N 较大时,视频映射图建立的总体计算复杂度主要来自于 SURF 特征提取的复杂度.

对于大规模视频序列数据,算法采取分层的策略对视频进行预处理,从而降低计算复杂度. 如图 9 所示,对于一个包含较大帧数的视频 V,以时间中点为界,把视频分成 V_1,V_2 这 2 个片段,同样的方法把视频片段 V_1 分为 V_{11} 和 V_{12},片段 V_2 分为 V_{21} 和 V_{22},以此类推,直到每个片段的帧数在 1 000 帧以内(可以根据实际情况进行确定),对于最低层的每个视频片段构造视频映射图,提取每个片段的关键帧,组合在一起构造最终的关键帧待选集合,在待选集合中再去除内容相近的帧,最终得到视频关键帧集合.

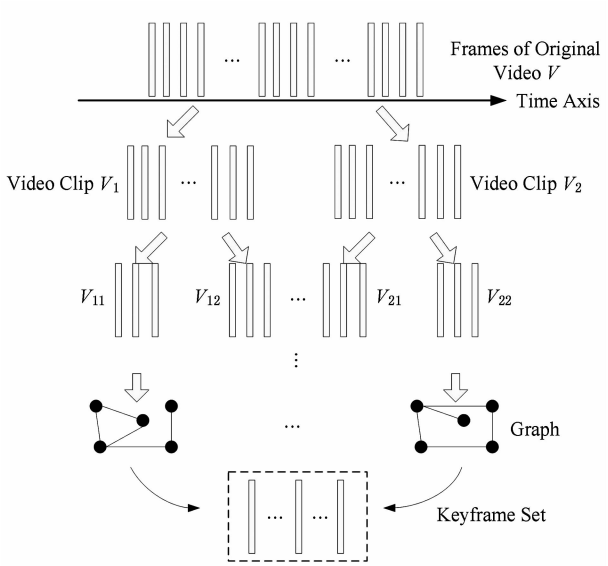


Fig. 9 Illustration of video layer preprocessing.

图 9 视频分层预处理示意图

2) 整数规划计算复杂度分析

在利用整数规划计算视频映射图支配集的过程中,系数矩阵 $\boldsymbol{A}$ 的大小为 $N\times N$,当 N 较大时,会增加整数规划求解和程序存储的复杂度,但是,矩阵 $\boldsymbol{A}$ 是包含 0 或 1 的稀疏矩阵,因此,本文算法在程序设计时利用稀疏矩阵存储的形式来降低计算的复杂度.

3 结 论

本文基于支配集和整数优化理论提出了一种关键帧提取方法,该方法具有如下特点:1)构造视频映射图;2)把关键帧提取等价为视频映射图的极小支配集的选取问题;3)通过整数规划来求出视频映射图的极小支配集. 实验证明,本文方法提取的关键帧在保真度和压缩率方面都具有较好的性能. 与传统方法相比,本文算法更关注于关键帧内容的代表性而非时间的先后顺序,同时,关键帧提取也与视频镜头划分无关. 因此,本文算法对于构造视频的静态摘要是非常有效的. 另外,本文算法把关键帧的提取等价为一个图论的经典问题,因此,本文算法具有良好的理论基础.

关键帧提取在视频信号处理中具有重要的地位,本文对于图论在视频关键帧提取方面的应用进行了初步的探索. 下一步的工作主要集中在如何构造更有效、更合理的视频映射图,从而能够选取更加优化的极小支配集来表示关键帧. 例如,本文算法利用加速鲁棒特征构造视频映射图,体现了视频帧的内容特征,但是,视频的使用者是人,如何加入人对内容的主观感觉是我们下一步要考虑的问题之一.

参 考 文 献

[1] Ma Cuixia, Liu Yongjin, Fu Qiufang, et al. Video summarization method based on sketches interaction and cognitive analysis [J]. SCIENCE CHINA: Information Sciences, 2013, 43(8): 1012-1023 (in Chinese)
(马翠霞, 刘永进, 付秋芳, 等. 基于草图交互的视频摘要方法及认知分析[J]. 中国科学: 信息科学, 2013, 43(8): 1012-1023)

[2] Wang Shen. Research on key frame extraction based on clustering algorithm and multi-feature fusion [D]. Wuhan: Huazhong University of Science & Technology, 2012 (in Chinese)
(王深. 基于聚类算法的多特征融合关键帧提取技术研究[D]. 武汉: 华中科技大学, 2012)

[3] Dimitrova N, Zhang H J, Shahraray B. Applications of video-content analysis and retrieval [J]. IEEE Multimedia, 2002, 9(3): 42-55

[4] Smoliar S W, Zhang H J. Content-based video indexing and retrieval [J]. IEEE Multimedia, 1994, 1(2): 62-72

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

[6] Wang Fangshi, Xu De, Wu Weixin. A cluster algorithm of automatic key frame extraction based on adaptive threshold [J]. Journal of Computer Research and Development, 2005, 42(10): 1752-1757 (in Chinese)
(王方石, 须德, 吴伟鑫. 基于自适应阈值的自动提取关键帧的聚类算法[J]. 计算机研究与发展, 2005, 42(10): 1752-1757)

[7] Naveed E, Irfan M, Sung W. Efficient visual attention based framework for extracting key frames from videos [J]. Signal Processing: Image Communication, 2013, 28(1): 34-44

[8] Zhang Xudong, Liu Tieyan, et al. Dynamic selection and effective compression of key frames for video abstraction [J]. Pattern Recognition Letters, 2002, 24(9/10): 1523-1532

[9] Kin W S, Kin M L, Guoping Q. A new key frame representation for video segment retrieval [J]. IEEE Trans on Circuits and Systems for Video Technology, 2005, 15(9): 1148-1155

[10] Ling Laijie, Yang Yi. Key frame extraction based on visual attention model [J]. Journal of Visual Communication and Image Representation, 2012, 23(1): 114-125

[11] Jiang Peng, Qin Xianlin. Key frame based video summary using visual attention clues [J]. IEEE Trans on Multimedia, 2010, 17(2): 64-73

[12] Herbert B, Andreas E, Tinne T, et al. SURF: Speeded-up robust features [J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359

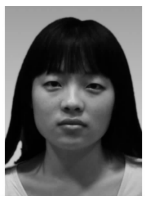
[13] Belkin M, Niyogi P, Sindhwani V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples [J]. Journal of Machine Learning Research, 2006, 7(11): 2399-2434

[14] West D B. Introduction to Graph Theory [M]. 2nd ed. Delhi, India: Pearson Education (Singapore), Indian Branch, 2002: 117
(韦斯特 D B. 图论导引[M]. 李建中, 骆吉洲, 译. 2 版, 北京: 机械工业出版社, 2006: 90-91)

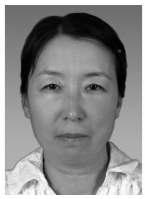
[15] Sanchis L A. Bounds related to domination in graphs with minimum degree two [J]. Journal of Graph Theory, 1997, 25(2): 139-152

[16] Hyun S C, Sanghoon S, Sang U L. Efficient video indexing scheme for content-based retrieval [J]. IEEE Trans on Circuits and Systems for Video Technology, 1999, 9(8): 1269-1279

Nie Xiushan, born in 1981. PhD and professor. Member of China Computer Federation. His research interests include multimedia processing and multimedia security.



Chai Yan'e, born in 1990. Master candidate of Shandong University of Finance and Economics. Her research interests include multimedia processing and multimedia security.



Teng Cong, born in 1968. PhD and associate professor. Her research interests include graph and large-scale computing.