文章编号:1000-582X(2010)02-0069-04

自适应 CCV 及等价关系聚类的视频摘要的生成

罗 斌,戴玉名,翟素兰
(安徽大学 计算机学院,合肥 230039)

摘 要:在静态视频摘要生成过程中,提取代表帧是关键。通过分析视频帧底层的颜色特征,根据其内容自动设置连通阈值,提取颜色聚合向量,并以此进行基于等价关系的自适应聚类。在确定整体划分后,根据时序特征进行局部修正。整个过程无需设定任何阈值和人为干预,对变化程度不同的各种视频进行实验,结果比较满意。

关键词:颜色聚合向量;自适应等价关系;视频摘要
中图分类号:TP391.4 文献标志码:A

Generating and reducing video summary with adaptive CCV and equivalent relation clustering

LUO Bin, DAI Yu-ming, ZHAI Su-lan
(School of Computer Science & Technology, Anhui University, Hefei 230039, P. R. China)

Abstract: It is very essential to extract representative frames in the process of generating video summary. A method is proposed which analyses the color features of the video frames, sets the connectivity threshold values automatically according to the contents, extracts the color coherence vectors (CCV) and then performs adaptive clustering based on equivalent relation. After global partition, local partition was revised with time sequential features. The whole process does not need to set any threshold values. The experiments with diverse videos yields effective results.

Key words: color coherence vector; adaptive equivalent relation; video summary

随着多媒体技术的迅速发展、存储设备的广泛使用及网络性能的不断提高,包含丰富内容的视频层出不穷。同时,庞大的数据量也给存储、传输、浏览和检索等带来挑战。如何快速有效地精简视频,生成视频摘要[1-2]是当前研究的热点问题。“代表帧”序列可以简洁地反映整个视频的主要内容。提取代表帧一般有 3 个要求:其一,代表帧应该具有代表性。其二,代表帧的数目需要随着视频内容的变化而变化,即内容丰富的视频其代表帧应该多一些。其三,提取过程应该尽量避免人为干预,实现自动化。

目前,针对视频帧特征的描述,使用颜色直方图[3-5]最为普遍,它运算简单,对摄像机运动不敏感。但它描述的只是颜色统计信息,缺乏其空间信息,对全局颜色不变而局部变化的视频帧比较敏感。Lim 等[6]通过在视频帧序列上计算光流来反映每个像素的运动信息,光流值一般比较小,容易受误差和噪声的影响,且计算复杂。Greg Pass 等[7]使用颜色聚合向量(color coherence vectors),它包含了颜色的统计和空间信息,设定的固定连通阈值却无法适应内容丰富的视频帧。对于代表帧提取技术,Rasheed 等[8]计算当前视频帧和已存在的各聚类中心的距离,通过设定距离阈值进行聚类。Sun[9]等提

收稿日期:2009-10-21
基金项目:国家自然科学基金资助项目(NO. 60772122);安徽省教育厅自然科学研究重点资助项目(NO. KJ2008A033&NO. KJ2007A072);安徽省高校优秀青年人才基金资助项目(NO. 05010118)
作者简介:罗斌(1963-),男,安徽大学教授,主要从事图像处理与模式识别方向研究,(E-mail)daiyuming2007@163.com。

出自动确定聚类数目，但要指定最大代表帧数目和是否成为候选代表帧的参数。Biswal[10]等使用模糊C均值聚类，其目标函数存在大量局部极值点，初始化不当则得不到最优模糊划分，而且需要指定聚类数目。Y. Hadi[11]等使用K-Mediod聚类方法，也要指定聚类数目。

通过自动设置连通阈值提取颜色聚合向量，将结果作为视频帧的特征向量，输入到基于等价关系的自适应聚类算法中进行聚类，并根据时序特征对整体划分后的局部划分作出取舍，在每个保留划分中提取序号居中者作为聚类代表帧生成静态视频摘要。

1 基于自适应阈值的CCV视频帧特征提取算法

针对固定连通阈值提取的CCV无法适应内容丰富的视频帧，得到的CCV特征不能真实反映视频主要内容，提出基于动态阈值的视频帧CCV特征提取算法。

在综合文献[12-13]中描述的搜索标记算法基础上进行像素的搜索标记后，根据视频帧内容动态设定自适应连通阈值，其算法描述如下

Step1：计算每个像素值的标记号的最大值。

$\mathrm{MaxMark}(k) = \max(\mathrm{SearchTag}(i,j))$。

Step2：计算每个像素值每个标记号包含的像素数目。

$$\text{when } \mathrm{SearchTag}(i,j) = k,$$

$\mathrm{NumPerMark}(k) = \mathrm{NumPerMark}(k) + 1$。

Step3：计算在搜索标记的连续标号中可能漏掉的标号的数目。

when $\mathrm{NumPerMark}(k) = 0$, $\mathrm{leak} = \mathrm{leak} + 1$。

Step4：确定连通阈值为平均每个像素值每个标记号包含的像素数目。

$$\mathrm{Threshold} = \frac{\mathrm{width} * \mathrm{height}}{\mathrm{sum}(\mathrm{MaxMark}(k))\text{-leak}}。$$

Step5：计算每个像素值的连通像素数目与非连通像素数目。

$$\begin{cases} \text{when} \quad \mathrm{NumPerMark}(i) >= \mathrm{Threshold}, \\ \qquad \mathrm{ConnectPixel}(i) += \mathrm{NumPerMark}(i); \\ \text{when} \quad \mathrm{NumPerMark}(i) < \mathrm{Threshold}, \\ \qquad \mathrm{DisConnectPixel}(i) += \mathrm{NumPerMark}(i)。 \end{cases}$$

提取CCV特征向量后，视频中第 i 帧图像就可以表示成如下形式

$$f_i = (a_{i_1}, b_{i_1}, a_{i_2}, b_{i_2}, \ldots, a_{i_{n-1}}, b_{i_{n-1}}, a_{i_n}, b_{i_n})。$$

2 基于等价关系自适应聚类的视频摘要生成算法

在对视频内容一无所知的情况下，提供聚类数目等经验阈值是很困难的。基于等价关系[14]，研究算法可以根据视频内容的变化程度，自适应确定聚类数目。而且，考虑视频的时序性，设置时序窗口，不仅可以极大限度地提取同一镜头或场景的代表帧，还可以提取不同镜头或场景间回放的代表帧，如图1所示。

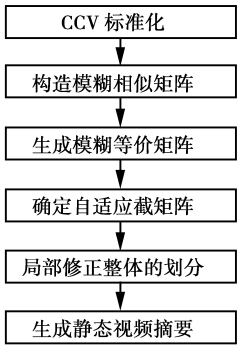


图1 等价关系自适应聚类算法框架图

利用得出的等价矩阵，通过自适应 λ 进行一次整体划分，接着通过时序窗口进一步确定局部划分，最终得出聚类结果，其算法描述如下

Step1：采用极值标准化公式，将CCV特征向量标准化，作为聚类的样本数据 $\bar{\boldsymbol{F}}_{ij}$

$$\bar{\boldsymbol{F}}_{ij} = \frac{F_{ij} - \min(F_j)}{\max(F_j) - \min(F_j)},$$

其中，$\boldsymbol{F}_{ij}$ 表示第 i 帧的CCV特征的第 j 维元素。

Step2：采用Euclid相对距离计算视频帧间的相似度，作为模糊相似矩阵 $\boldsymbol{R}_{ij}$

$$\boldsymbol{R}_{ij} = 1 - \sqrt{\frac{1}{n}\sum_{k=1}^{n}(\bar{F}_{ik} - \bar{F}_{jk})^2}。$$

Step3：通过平方求传递闭包法，将模糊相似矩阵 $\boldsymbol{R}_{ij}$ 转换为模糊等价矩阵 $\boldsymbol{E}_{ij}$

$$\boldsymbol{E}_{ij}^{*} = \max(\min_{k=1}^{n}(\boldsymbol{R}_{ik}, \boldsymbol{R}_{kj}), \boldsymbol{E}_{ij}),$$

其中，用 $\boldsymbol{E}_{ij}^{*}$ 替换 $\boldsymbol{E}_{ij}$ 迭代到 $\boldsymbol{E}_{ij}^{*} = \boldsymbol{R}_{ij}$ 为止。

Step4：构造自适应 λ 截阈值（$0 \leqslant \lambda \leqslant 1$），进行一次整体划分

$$\lambda = \underset{i=1,j\leqslant i}{\overset{n}{\mathrm{mean}}}(\boldsymbol{E}_{ij}{}^{*}) + \underset{i=1,j\leqslant i}{\overset{n}{\varpi r}}(\boldsymbol{E}_{ij}{}^{*})。$$

Step5：在④中的一次整体划分里，先选取帧号最小者作为“准代表帧”，再设置时序窗口，对局部划分做出取舍

$$C_{ij}=\begin{cases}\text{保留，当第 } j \text{ 个局部划分与准关键帧标号连续；}\\ \text{保留，当第 } j \text{ 个局部划分的帧数}>\text{时序窗口帧数；}\\ \text{删除，其他。}\end{cases}$$

其中，C_{ij} 表示第 i 个整体划分的第 j 个局部划分。

Step6：提取每个保留划分的帧号居中者作为代表帧，按照时序关系，生成静态视频摘要。

3　实验结果及分析

为了验证研究的方法，主要进行了 2 方面实验。一是验证动态连通阈值 CCV 视频特征提取的有效性，二是使用基于等价关系自适应聚类的视频摘要结果与 open-video 进行比较。实验数据除部分来自互联网外，主要来自 open-video.

3.1　图像特征提取结果比较

实验中，将测试图像量化成 32 阶灰度图（如图 2），摘录的动态连通阈值与固定连通阈值[7] 提取 CCV 图像特征结果对比如表 1 所示。

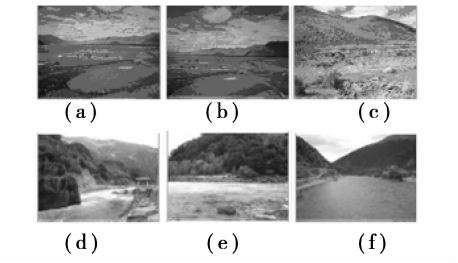


图 2　测试图像

表 1　动态连通阈值与固定阈值 CCV 提取图像特征对比表

	第一组			第二组		
	(a,b)	(a,c)	(b,c)	(d,e)	(d,f)	(e,f)
固定阈值 CCV	0.420	0.300	0.369	0.514	0.301	0.239
动态阈值 CCV	0.422	0.129	0.239	0.397	0.239	0.199

直观来看，(a) 与 (b)、(d) 与 (e) 较相似，且都与 (c)、(f) 差别较大。从表 1 数据来看，自适应连通阈值提取的 CCV 特征具有更好的区分度。而且，在第一组与第二组之间比较，(a) 与 (b) 的相似度应该大于 (d) 与 (e) 的相似度，自适应阈值提取的 CCV 特征反映了这个特性，但固定阈值提取的 CCV 特征却不能反映。

由于通过聚类生成视频摘要，重要的就是要有稳定的能较好反映视频帧的特征数据。因此，本文使用自适应连通阈值提取 CCV 特征的方法比较适合。

3.2　视频摘要生成结果比较

利用不同特点的视频序列做了大量实验（视频序列中每 10 帧提取一帧，将时序窗口设置为 2 帧），现取 3 个典型视频片断说明如下：

①Film 视频片段：其内容变化丰富，包含渐变镜头、曝光不足镜头和大量整体颜色相近镜头等，共 779 帧，每帧 352×240 像素。通过手工和使用本文算法产生的静态视频摘要比较如图 3 所示，漏掉 3 个代表帧，是因为在内容复杂的视频中兼顾整体和局部划分都比较合理的情况下的产生的漏检。总体看来，提取的结果比较合理。

(a) 手工提取的代表帧

(b) 自适应等价关系聚类提取的代表帧

图 3　Film 视频摘要结果对比图

②New Indians Segment 08 视频片段：其背景缓慢变化，前景中速切换，共 706 帧，每帧 320×240 像素。使用 OpenVideo 网站提供的故事板与研究算法生成的静态视频摘要对比如图 4 所示，二者均完全提取了视频中的代表帧，真实反映了视频主要内容，但前者是通过 K-means 聚类产生，需要人为的干预，而后者完全是自适应的。

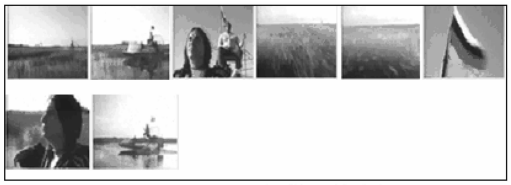
(a)OpenVideo 提供的故事板

(b) 自适应等价关系聚类提取的代表帧

图 4　New Indians Segment08 视频摘要结果对比图

③New Indians Segment 10 视频片段：包含摄像机和目标的快速运动，共 695 帧，每帧 320×240

像素。使用 OpenVideo 网站提供的故事板[2] 与研究算法生成的静态视频摘要对比如图 5 所示，在视频中含有摄像机快速跟踪目标以及回放镜头时，也能很好的筛选出代表帧。与前者相比，整个处理过程无需人为干预，提取的结果比较满意。



(a)OpenVideo 提供的故事板

(b) 自适应等价关系聚类提取的代表帧

图 5 **New Indians Segment10** 视频摘要结果对比图

4 结 论

特征的提取和聚类方法的选择是影响视频摘要结果的主要因素。考虑了视频帧的颜色统计信息和空间信息，自适应的提取 CCV 特征作为视频的特征，输入到一种基于等价关系的自适应聚类中。引入局部划分，修正最终的视频摘要结果。通过实验结果比较，本文的方法可以不经过人为的干预得到较好的视频摘要。

参考文献

[1] TRUONG B T, VENKATESH S. Video abstraction: A systematic review and classification [J]. ACM Transactions on Multimedia Computing, Communications and Applications, 2007,3(1):1-37.

[2] FURINI M, GERACI F, MONTANGER M, et al. VISTO: visual storyboard for web video browsing[C]// Proceedings of the ACM International Conference on Image and Video Retrieval, July 9-11, 2007. Amsterdam, The Netherlands:IEEE,2007: 635-642.

[3] KOTOULAS L, ANDREALAS I. Colour histogram content-based image retrieval and hardware implementation[C]// IEEE Proceedings on Circuits, Devices and Systems. United Kingdom: IEEE, 2003, 150(5):387-393.

[4] FERMAN A M, TEKALP A M, MEHROTRA R. Robust color histogram descriptors for video segment and identification [J]. IEEE Transactions on Image Processing, 2002, 11(5):497-508.

[5] VALDES V, MARTINEZ J M. On-line video skimming based on histogram similarity [C] // Proceedings of the International Workshop on TRECVID Video Summarization, September 28, 2007, Augsburg, Bavaria, Germany:IEEE, 2007: 94-98.

[6] LIM S, APOSTOLOPOULOS J G, GAMAL A E. Optical flow estimation using temporally oversampled video[J]. IEEE Transactions on Image Processing, 2005, 14(8): 1074-1087.

[7] PASS G, ZABIH R, MILLER J. Comparing images using color coherence vectors[C]// Proceedings of the fourth ACM international conference on Multimedia , 1996, Boston, Massachusetts, United States. Boston, Massachusetts: [s. n.], 1997: 65-73.

[8] RASHEED Z, SHAH M. Detection and representation of scenes in videos [J]. IEEE Transanctions on Multimedia, 2005 7(6): 1097-1105.

[9] SUN X, KANKANHALLI M S, ZHU Y, et al. Content -based representative frame extraction for digital video [C] // IEEE Multimedia Computing and Systems, June 28-July 1, 1998. Austin,Texas, USA: IEEE,1998: 347-358.

[10] BISWAL B, DASH P K, PANIGRAHI B K. Power quality disturbance classification using fuzzy C-means algorithm and adaptive particle swarm optimization [J]. IEEE Transactions on Industrial Electronics, 2009, 56(1):212 -220.

[11] HA DI Y, ESSANNOUI F, THAMI R. Video summarization by k-medoid clustering[C]// In Proc. of ACM Symposium on Applied Computing, April 23-27, 2006. Dijon, France:IEEE, 2006,1400-1401.

[12] 左文明. 连通区域提取算法研究[J]. 计算机应用与软件,2006, 23(1): 97-98.
ZUO WEN-MING. Study on connected regions extraction [J]. Computer Application and Software, Jan. 2006,23(1):97-98.

[13] 李仪芳,刘景琳. 基于连通域算法的区域测量[J]. 科学技术与工程,2008,8(9):2492-2494.
LI YI-FANG, LIU JING-LIN. Measurement for area based on connected regions arithmetic [J]. Science Technology and Engineering, 2008,8(9):2492-2494.

[14] 高新波. 模糊聚类分析及其应用[M]. 西安:西安电子科技大学出版社,2004.

(编辑 侯 湘)