基于聚类的视频镜头分割和关键帧提取 *

潘 磊,吴小俊,尤媛媛

(江苏科技大学 电子信息学院,江苏 镇江 212003)

摘要:镜头分割是基于内容的视频检索和浏览首先要解决的关键技术。视频分割为镜头后,下一步的工作就是进行关键帧提取,用以描述镜头的主要内容。提出了一种改进的基于聚类的镜头分割和关键帧提取算法,在无监督聚类中引入一个参考变量,解决了利用无监督聚类进行镜头分割和关键帧提取时可能产生的帧序不连续或分割错误的问题。在关键帧提取阶段,将镜头分割为子镜头后,引入图像熵的概念提取关键帧。实验结果表明了改进算法在镜头分割和关键帧提取方面的有效性。

关键词:镜头分割; 关键帧; 聚类; 图像熵

中图分类号:TP391 **文献标识码**:A **文章编号**:1007-2276(2005)03-0341-04

Video shot segmentation and key frame extraction based on clustering*

PAN Lei , WU Xiao-jun , YOU Yuan-yuan

(Department of Electronics and Information,Jiangsu University of Science and Technology , Zhenjiang 212003, China)

Abstract: Shot segmentation is a vital technology that must be resolved firstly in video retrieval and browse. Then key frame extraction will be carried out after shot segmentation to describe main content of shot. An improved algorithm for shot segmentation and key frame extraction based on clustering is proposed, a referenced variable is used in unsupervised clustering to resolve the frame sequence's incontinuity or false segmentation problems which can be caused probably by unsupervised clustering. During key frame extraction, the concept of image entropy is used after shot being segmented into sub-shots. Experimental results demonstrate the efficiency of the improved algorithm in shot segmentation and key frame extraction.

Key words: Shot segmentation; Key frame; Clustering; Image entropy

0 引 言

目前研究的基于内容的视频检索系统,一般都是先进行镜头分割,即在时域上将视频序列按照一定的标准分割为镜头,然后提取镜头的关键帧。因此,如何将视频序列正确地分割为镜头是视频检索首先要解决的问题,镜头分割的好坏,直接影响到视频检索系统性能的优劣。而关键帧的提取,对于描述镜头的内

收稿日期:2004-08-01; **修订日期**:2004-08-20

* **基金项目**:江苏省自然科学基金资助项目(苏科基 2002-006)

作者简介:潘磊(1980-),男,江苏镇江人,硕士生,主要研究方向为模式识别、基于内容的视频检索。

容起到决定性的作用,是对视频节目建立索引的基础。

镜头是视频的基本物理单元,由一个摄像机连续拍摄的时间上连续的若干帧图像组成。镜头之间的变换包括两种:切变和渐变。切变是指一个镜头不采用任何编辑效果直接变换到另一个镜头;渐变是指一个镜头通过某种编辑手段,如淡入、淡出、叠化等,缓慢地变换到另一个镜头[1,2]。关键帧是用来描述一个镜头内部主要内容的某帧或某几帧图像,通过镜头分割后对每个镜头提取关键帧,就可在此基础上对视频建立索引,为视频检索和浏览提供了快捷简便的手段,并且极大降低了视频检索系统的处理时间,使得视频检索系统的实时性得到很大提高。

1 基于聚类的镜头分割算法

聚类技术在信息科学领域得到了广泛应用,其基本思想是从一个初始化的聚类出发,将一个样本集 $X=(X_1,X_2,\cdots,X_n)$ 中的每个元素分配给某个聚类,以达到系统或用户的要求。典型的基于聚类的镜头分割算法可参阅参考文献[3~5]。实验中发现,通常采用的聚类算法,可能导致镜头出现帧序号不连续以及镜头错误分割的问题。如采用参考文献[4]中的聚类算法,图 1 中的三幅图像应该属于图 2 中的第五类镜头,结果被错误地划分到图 3 的第二类镜头中,导致第二类镜头的帧序号不连续,而第五类镜头分割错误。

图 1 错误分类的帧

Fig.1 False classified frames

图 2 第五类镜头的几个代表帧

Fig.2 Some representative frames in the 5th shot

研究发现,导致此问题的关键在于聚类算法存在缺陷。由于参考文献[4]中算法每次将帧和各个已知镜头之间进行聚类比较,取相似性最大的镜头作为帧

图 3 第二类镜头的几个代表帧

Fig.3 Some representative frames in the 2nd shot

所属镜头,因此很容易出现上述问题。对此参考文献[4]提出了一种后处理方法。本文则对聚类算法进行了改进,无须进行后处理:当出现新的镜头时,前面已经分割完毕的镜头不再参加聚类。为此引入参考变量 *visit*,*visit* 为"1"时表示镜头尚未分割完毕,可继续进行聚类;为"0"时表示镜头已分割完毕,不再进行聚类。

对于视频序列 $V=\{f_1,f_2,\cdots,f_n\}$,将其投影到 *HSV* 颜色空间,空间分割采用参考文献[4]的 *HSV*(12×5×5)制,即 *H* 分量等分为 12 块,*S*、*V* 分量各自等分为 5 块。建立 *H* 分量的直方图为:

$$H(i)=\frac{H\_follow(i)}{M\times N} \tag{1}$$

式中 *H_ follow*(*i*)为 *H* 分量像素值落入第 *i* 段的像素个数;*M*,*N* 是图像两个方向的像素个数。类似地可建立 *S*、*V* 分量的直方图为:

$$S(j)=\frac{S\_follow(j)}{M\times N} \tag{2}$$

$$V(k)=\frac{V\_follow(k)}{M\times N} \tag{3}$$

式中 $i\in[1,12]$;$j\in[1,5]$;$k\in[1,5]$,则 *HSV* 空间的直方图 $H(i,j,k)$ 为一个三维数组,分别对应 *H*、*S*、*V* 三个分量的直方图。定义帧与镜头在 *H* 分量上的相似性为:

$$S_H(f,Shot)=\sum_{i=1}^{12}\min(H(i),Shot\_H(i)) \tag{4}$$

式中 $H(i)$ 是帧 *H* 分量的直方图;*Shot_H*(*i*)是镜头类内中心 *H* 分量的直方图,类似地可建立帧与镜头在 *S*、*V* 分量上的相似性为:

$$S_S(f,Shot)=\sum_{j=1}^{5}\min(S(j),Shot\_S(j)) \tag{5}$$

$$S_V(f,Shot)=\sum_{k=1}^{5}\min(V(k),Shot\_V(k)) \tag{6}$$

采用改进的聚类算法将其分割为 C 类，即分割后的镜头集为 $S=(S_1,S_2,\cdots,S_C)$。

(1) 令第一帧 f_1 为第一个镜头，其本身即为类内中心，该镜头 $Shot.visit=1$。

(2) 抽取下一帧 f_{next}，利用公式(4)，(5)，(6)得到 f_{next} 与镜头的相似性为：

$$S(f_{next},Shot)=\frac{\alpha\times S_H(f_{next},Shot)+\beta\times S_S(f_{next},Shot)+\gamma\times S_V(f_{next},Shot)}{3} \quad (7)$$

式中 α、β、γ 分别是 H、S、V 三个分量的加权系数，因为一般情况下人眼对 H 分量比较敏感，因此有 $\alpha\geqslant\beta$，$\alpha\geqslant\gamma$。此时进行聚类的镜头必须满足 $Shot.visit\neq 0$ 。

(3) 如果 $S(f_{next},Shot)$ 大于镜头分割阈值 T，则认为 f_{next} 属于镜头 $Shot$，将 f_{next} 放入镜头内，重新计算镜头的类内中心为：

$$Shot=\frac{f_{next}+\sum_{i=1}^{Shot.length} f_i}{Shot.length+1};\ Shot.length=Shot.length+1 \quad (8)$$

式中 f_i 是镜头内部已有的帧。如果 $S(f_{next},Shot)$ 小于镜头分割阈值 T，则认为 f_{next} 不属于镜头 $Shot$，重新建立一个新镜头，此时该镜头只包含 f_{next}，类内中心也为 f_{next}，$Shot.visit=1$。对上一个镜头做如下操作：

$$Shot.visit=0 \quad (9)$$

(4) 若视频未处理完毕转(2)，否则结束。

可以看出，改进后的聚类算法可以避免镜头出现不连续帧以及镜头错误分割的问题，实验表明该算法取得了较好的实验效果。

2 基于聚类的关键帧提取

视频分割为镜头后，下一步工作是从镜头中提取关键帧，用来描述镜头的主要内容。参考文献[5]提出了一种在镜头内部聚类形成子镜头集合以提取关键帧的算法，聚类算法与参考文献[4]类似。与对参考文献[4]的改进一样，本文对参考文献[5]的聚类算法做同样的改进。当镜头内部各帧相差不大时，该镜头即为其子镜头，只提取一个关键帧；当镜头内部帧差较大时，该镜头被聚类为若干子镜头，提取若干关键帧。

在子镜头集合聚类完毕后，引入图像熵的概念提取关键帧。参考文献[6]提出了一种图像熵的定义，在此引用该定义，加以一定的约束条件后用于关键帧提取。图像 f_i 的图像熵可定义为：

$$E(f_i)=-\sum_{k=1}^{n} h_k\log h_k \quad (10)$$

式中 h_k 代表像素值为 k 的像素占图像像素总数的比例，容易证明图像熵总是大于“0”的。可以看出，当 $h_k=0$ 时，图像熵无意义，因此加入一个约束条件如下：

$$\text{如果}(h_k==0),\ \text{那么}\ \log h_k=0 \quad (11)$$

采用公式(10)，(11)分别对图像 f_i 的 H、S、V 三个分量进行计算得到图像熵为：

$$E(f_i)=\alpha\times E(f_i)\_H+\beta\times E(f_i)\_S+\gamma\times E(f_i)\_V \quad (12)$$

式中 $E(f_i)\_H$，$E(f_i)\_S$，$E(f_i)\_V$ 分别是图像在 H、S、V 三个分量上的图像熵，由于人眼对 H 分量比较敏感，因此有 $\alpha\geqslant\beta$ ，$\alpha\geqslant\gamma$。当镜头内部差异不大时，提取图像熵最大的帧作为关键帧；当镜头内部差异较大而被聚类为若干子镜头时，提取各子镜头内部图像熵最大的帧作为关键帧。

镜头被分割为子镜头集合后，各子镜头分别代表了镜头内部的不同内容，子镜头间内容差异相对较大，而子镜头内部内容相对一致。图像熵反映了图像颜色分布的均匀程度，代表了图像包含信息量的大小，图像熵越大，图像颜色分布越不均匀，图像包含的信息量也就越大。因此，在子镜头中提取图像熵最大的帧作为关键帧，可有效反映镜头的主要信息，实验表明该方法提取的关键帧能够准确刻画镜头的主要内容。

3 实验结果及分析

本文采用 12 帧/s 的三星手机广告和 30 帧/s 的篮球比赛的部分片断进行实验。实验视频中包括切变、渐变等镜头变换，镜头中存在较大的物体运动和摄像机自身运动。H、S、V 三个分量的加权系数分别为 $\alpha=1.005$，$\beta=\gamma=0.995$，手机广告镜头分割阈值 $T_1=0.85$，子镜头分割阈值 $T_2=0.97$，篮球比赛镜头分割阈

值 T_3 =0.81 ,子镜头分割阈值 T_4 =0.91。表 1 是算法检测出的部分镜头集和关键帧。图 4 是部分镜头分割的结果图,图 5 展示了相对于图 4 的关键帧。

表 1 算法检测的镜头集

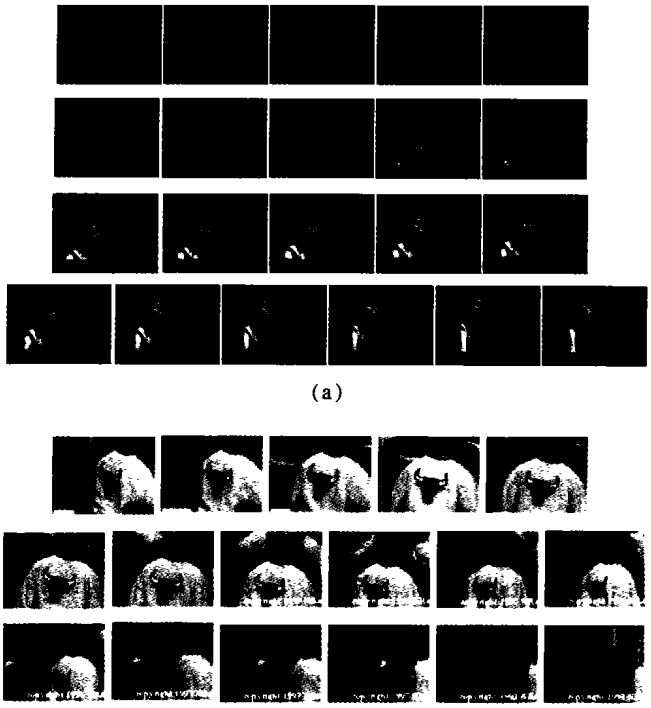
Tab.1 Shot sets detected by the algorithm

Advertisement of mobile	Detected shot and key frame		Actual shot
1	[1~94]	61, 83	[1~94]
2	[95~176]	118, 167	[95~176]
3	[177~208]	187, 203	[177~208]
4	[209~230]	226	[209~230]
5	[231~240]	239	[231~240]
Basketball match	**Detected shot and key frame**		**Actual shot**
1	[1~87]	82	[1~87]
2	[88~95]	93	[88~95]
3	[96~109]	96	[96~109]
4	[110~123]	118	[110~123]
5	[124~140]	125,137,140	[124~140]

可以看出,图 4 中镜头(a)虽然相对较长,但只表明了目标缓慢出现这么一个动作, 因此抽取的图 5(a)两幅关键帧已经足够描述该镜头;而图 4 镜头(b)虽然相对较短,但包含了目标转身、步行和停止三个动作,因此针对不同的动作抽取了图 5(b)三幅关键帧。所以,该算法提取出的关键帧能够有效刻画镜头的主要内容,尤其是运动的内容。

4 结束语

提出了一种改进的基于聚类的镜头分割和关键帧提取算法, 并在关键帧提取中引入图像熵的概念。从实验结果看,基于聚类的方法在镜头分割和关键帧提取中有比较广泛的应用前景,而图像熵在关键帧提取则是一种新的尝试。实验中视频各种镜头变换比较明显,阈值容易人为确定,分割效果比较理想。而对于复杂且不规律的镜头变换, 该算法仍有一定的缺陷,阈值的选取比较困难,分割容易出现误差。因此,自动且鲁棒的阈值确定将是今后研究的一个重点。



(a)

(b)

图 4 几个镜头集合

Fig.4 Some shot sets

(a) (b)

图 5 关键帧

Fig.5 Key frames

参考文献:

[1] 章毓晋.基于内容的视觉信息检索[M].北京:科学出版社,2003. 221-223.

[2] 陆海斌,章毓晋.一种高效视频切变检测算法[J].中国图像图形学报,1999,4(10):805-810.

[3] 刘政凯,汤晓鸥. 视频检索中镜头分割方法综述[J].计算机工程与应用,2002,38(23):84-87.

[4] 金红,周源华,梅承力.用非监督式聚类进行视频镜头分割[J].红外与激光工程,2000, 29(5):42-46.

[5] Zhang Yue-ting, Rui Yong, Thomas Shuang, et al. Adaptive key frame extraction using unsupervised clustering[A]. Proceedings of International Conference on Image Processing[C].1998,1.886-870.

[6] 张明,沈祖治,王志坚. 一个基于内容图像检索系统的设计与实现[J].水电能源科学,2003,21(3):46-48.