

基于 HSV 空间模型的视频场景切分

黄东军,安琪(卢旺达)

(中南大学 信息科学与工程学院,湖南 长沙 410083)

摘要:针对 RGB 彩色空间模型下统计直方图用于切分场景时,存在对内容变化不敏感、计算量大等问题,提出了一种基于 HSV 空间模型的视频场景切分方法。该方法将图像帧的 RGB 转换为相对应的 HSV 值,然后对 H、S、V 进行非等间隔量化,再合成为一维特征矢量。计算视频前后两帧图像的一维特征矢量直方图,求出它们的相似度,从而判断出是否发生了场景切换。实验证明,该方法能够比较准确地对视频进行场景切分,为视频摘要的关键帧提取提供了必要的条件。

关键词:场景切分;HSV 颜色空间模型;视频摘要

中图分类号:TP391;TN919.85

文献标识码:A

文章编号:1673-629X(2009)09-0009-04

Video Scene Cutting Based on HSV Space Model

HUANG Dong-jun, Ingabire Marie Ange(Rwanda)

(School of Information Science and Engineering, Central South University, Changsha 410083, China)

Abstract: To overcome the problem in RGB color model used to deal with video shot detection, proposes a method based HSV - color space model of scene cutting in video abstraction. In the method, the first step is to transfer the RGB values of the picture frame to the corresponding HSV values. After doing non - equal interval quantizing to the three HSV weights, the next is to compute the quantizing weights to one - dimensional eigenvector. Calculating the one - dimensional eigenvector histograms of the adjacent two frames can get the similarity. According to the similarity can judge if the video scene has converted. As the experiment proves, the method can cut the scenes of video more accurately. It provides a basic technique for extracting key frames in video abstraction.

Key words: scene cutting; HSV - color space model; video abstraction

0 引言

随着网络技术的飞速发展和多媒体技术的广泛应用,大量的数字视频迅速涌现。这种爆发式增长引发了许多新的技术,包括视频存档、编目、索引以及存取等。而视频摘要技术是目前研究的热点,为视频存档和索引等提供了技术支持。在视频摘要的提取过程中,最主要的一个步骤是对视频文件进行场景切分。场景,是基于镜头的一段视频。场景的切分,就是根据镜头与镜头之间的背景、人物、事件的改变,来提取不同的故事情节,或不同的地点事件等。

1 研究现状

一般来说,在整个视频中,当从一个场景变换到另一个场景的时候,相邻的图像帧之间的背景、人物、

环境等都会发生明显变化,具体体现在图像的颜色分布上^[1,2]。因此,要判断两幅图像之间的相似性,可以根据图像的颜色属性来实现。目前基于颜色属性的场景切分,已经提出了多种方法:基于统计、基于变换、基于特征和基于直方图等^[3]。

在文献[4]叙述的方法中,采用了基于交互信息量的技术来对镜头检测,进行场景切分。对连续的两帧,分别独立计算出其三个 RGB 分量的交互信息量和联合熵,再计算总的交互信息量,然后对每个帧间交互量,计算其滑动窗口内交互信息量的均值,再与给定的阈值进行比较,从而判断是否在此帧处检测到了场景边界。

文献[5]的场景切分镜头检测中,采用了基于离散余弦变换系数的方法。由于频域中的变换系数是与像素紧密相关的,因此,DCT 系数可以用于压缩视频序列中的镜头边界检测,从而进行场景切分。具体做法是对从压缩视频文件中选取出的 I 帧,做出时面各个 8×8 块中的第一 DCT 系数的直方图 $H(I_k, j)$,其中 I_k 代表第 k 个 I 帧, j 代表该帧共 N 个块中的第 j 个块。采用 χ^2 检验法比较相连 I 帧的 DCT 系数直方图的各个

收稿日期:2008-12-19;修回日期:2009-03-27

基金项目:国家自然科学基金(60873188)

作者简介:黄东军(1960-),男,教授,研究方向为网络多媒体及数字图像处理;安琪,卢旺达来华留学生,研究方向为视频信息处理。

统计值,如果测度 D 超过了给定的阈值就认为发生了场景切换。这种方法主要是针对压缩视频的,对于非压缩视频则不能采用此方法^[6]。

我们知道,即使两张图片在视觉上相差很大,但它们的颜色直方图还是有可能相似的,所以仅仅依靠颜色直方图还不能完全准确地判断出两幅图像之间的差异性。基于此,考虑一种改进的方法,即将一幅图片分割成一小块一小块的图像,分别计算各小块图像的颜色直方图,再比较两幅图像的对应块颜色直方图,从而可以比较精确地得出这两幅图像的差异性。在对视频文件进行场景切分的时候,就可以采取颜色直方图的方法来处。具体的做法是,先将视频中的每一帧分割成 $a \times b$ 块,再分别计算每一块的颜色直方图,然后再分别比较相邻两帧图像对应块之间的直方图,假如这两帧图像的直方图超过设定的阈值,则可认为在这一帧处发生了场景变换,因此在这里对场景进行切分,切割成一个子视频。

但是,在基于 RGB 颜色模型直方图的场景切分算法中,还存在着一些缺点,比如上面提到的两幅不同图像的颜色直方图却有可能相同,而采用将图像分割成多个小块,分别计算各个小块的直方图,再计算两幅图像的直方图这种方法的计算量太大,严重影响了算法的效率^[7]。因此,在文中的方法中,提出了一种基于 HSV 颜色模型的场景切分算法。

2 场景切分算法分析

HSV 颜色模型是一种基于感知的颜色模型,它把彩色信号表示为三种属性:色调(Hue)、饱和度(Saturation)和亮度(Value),用来确定颜色的一种方法。这种颜色模型用 Munsell 三维空间坐标系统表示^[8]。Munsell 颜色空间具有坐标间的心理感知独立性以及线性伸缩性,有较强的感知度,更为直观,可以使人立即了解是哪种颜色,色彩有多纯,亮度有多高,因此应用 HSV 颜色模型更适合人的肉眼判断。在场景切分算法中,将采用 HSV 颜色模型。

HSV 色彩属性模式中的色调(H)是色彩的基本属性,就是平常所说的颜色名称,如红色、蓝色等,取 $0 \sim 360$ 度的数值。饱和度(S)是指色彩的纯度,饱和度越高色彩纯度也越高,低则逐渐变灰,可以取 $0 \sim 100\%$ 的数值。而亮度(V)是颜色的明暗程度,取值也是 $0 \sim 100\%$ 之间。

从图像中直接得到的一般都是 RGB 三色值,因此要把 RGB 值转换为 HSV 值。给定 RGB 颜色空间的值 (r, g, b) , $r, g, b \in [0 \dots 255]$, 则可以通过计算将其转换成 HSV 颜色空间的 h, s, v 值。

$$\max = \text{Max}(r, g, b) \quad (1)$$

$$\min = \text{Min}(r, g, b) \quad (2)$$

$$v = \max/255; \quad (3)$$

$$s = \begin{cases} (\max - \min)/\max, & \text{if } \max \neq 0 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

$$h = \begin{cases} (g - b)/(\max - \min), & \text{if } r = \max \\ 2 + (b - r)/(\max - \min), & \text{if } g = \max \\ 4 + (r - g)/(\max - \min), & \text{otherwise} \end{cases} \quad (5)$$

若 $h < 0$, 那么

$$h = h + 2\pi \quad (6)$$

相对于在 RGB 颜色空间模型下,HSV 空间模型更能体现出一幅图像的位置信息。比如法国国旗和俄罗斯国旗,它们都是由蓝、白、红三种颜色组成的,只是位置不一样。它们的 RGB 颜色特征基本一致,RGB 直方图是相同的,因此无法分辨出这两幅图像。假如将 RGB 转换为 HSV 之后,则它们的直方图就不一样了,这样即可以区分它们。所以在 HSV 空间模型下对视频场景进行切分,将会取得更好的效果^[9]。而一幅图像的颜色一般会比较多样,尤其是真彩色图像,因此直方图的维数也会比较多,如果对 HSV 空间进行适当的量化后再计算直方图,则计算量要少得多。量化的方法如下:

按照人的颜色感知能力和视觉分辨能力,对 H, S, V 三个分量进行非等间隔的量化,把色调 H 空间分为 8 份,饱和度 S 分为 3 份,而亮度 V 也分为 3 份。然后再根据色彩的不同范围采用如下公式进行量化。

$$h' = 180 \times h/\pi \quad (7)$$

$$H = \begin{cases} 0, & \text{if } h' \in [316, 20) \\ 1, & \text{if } h' \in [20, 40) \\ 2, & \text{if } h' \in [40, 75) \\ 3, & \text{if } h' \in [75, 155) \\ 4, & \text{if } h' \in [155, 190) \\ 5, & \text{if } h' \in [190, 270) \\ 6, & \text{if } h' \in [270, 295) \\ 7, & \text{if } h' \in [295, 316) \end{cases} \quad (8)$$

$$S = \begin{cases} 0, & \text{if } s \in [0, 0.2] \\ 1, & \text{if } s \in (0.2, 0.7] \\ 2, & \text{if } s \in (0.7, 1] \end{cases} \quad (9)$$

$$V = \begin{cases} 0, & \text{if } v \in [0, 0.2] \\ 1, & \text{if } v \in (0.2, 0.7] \\ 2, & \text{if } v \in (0.7, 1] \end{cases} \quad (10)$$

得出 H, S, V 的量化值之后,就可以把三个颜色分量合成一维的特征矢量。

$$L = HQ_SQ_V + SQ_V + V \quad (11)$$

其中, Q_S, Q_V 分别是分量 S 和 V 的量化级数,取值为 $Q_S = 3, Q_V = 3$ 。因此,上式可以表示为:

$$L = 9H + 3S + V \quad (12)$$

从而, h, s, v 三个分量即可以分布在一维矢量上。根据上面的式子可以知道, L 的取值范围为 $[0, 1, \dots, 71]$, 计算图像的 L 值就可以获得图像的一维直方图。

得出基于 HSV 颜色空间模型的直方图之后, 就可以根据这个直方图来判断两幅图像之间的相似性了。假设图像 A , 其直方图的特征矢量为 $H_A(a_1, a_2, \dots, a_L)$, 图像 B , 其直方图的特征矢量为 $H_B(b_1, b_2, \dots, b_L)$ 。其中, $a_i \in [0, 1], b_j \in [0, 1]$ 为归一化的比例值, L 为一维直方图矢量的维数。归一化公式为:

$$a_i = A_i / \sum_{k=1}^L A_k, A_i \text{ 为特征值为 } L_i \text{ 的像素总数} \quad (13)$$

利用直方图相交算法进行图像相似性度量的方法如下:

$$\text{Similar}(A, B) = \sum_{i=1}^L \min(a_i, b_i) \quad (14)$$

按照上式, 如果两幅图像的颜色分布相同, 那么它们的相似度为 1, 否则为 0 到 1 之间的某一个数, 从而得出两幅图像的相似性。

3 算法描述

基于 HSV 颜色空间模型的场景切分算法描述可表示为下面的步骤。

- 1) 分别获取一帧图像中每个像素的 RGB 颜色值 (r, g, b) ;
- 2) 利用公式(1) ~ (6) 将其转换为 HSV 空间模型下的 HSV 值 (h, s, v) ;
- 3) 再利用公式(7) ~ (10) 将 (h, s, v) 进行量化得到量化值 (H, S, V) ;
- 4) 然后利用公式(12) 计算一维特征矢量值 L ;
- 5) 计算图像的直方图;
- 6) 重复以上 1) ~ 5) 五个步骤得出后一帧图像的直方图;
- 7) 根据公式(14) 计算这两帧图像的相似性;
- 8) 根据给定的阈值 m , 判断视频文件是否在此帧处发生了场景切分。若 $\text{Similar} > m$, 没有发生场景切分, 若 $\text{Similar} < m$, 则发生了场景切分, 应该在此处将视频文件分割成子场景视频。

此算法的关键之一在于最后阈值 m 大小的确定。 m 值取得太大了, 场景切分得太细, 也许同一个场景的视频也会被错误地分为几个子视频; 而 m 值取得太小, 则有可能错误地将不同场景的视频片段划分到同一个子视频中。因此, 要获取恰当的场景切分视频, m 的取值不能太大也不能太小。而对于不同类型的 AVI

视频文件来说, 镜头之间的切换在视觉上也有很大的不同, 有些很平缓, 如新闻类、体育类视频; 有些却很急剧, 难以进行统一, 如电影类视频等。在文中, 采用了实验的方法来确定阈值 m 的大小。

4 实验结果分析

为了检测此场景切分算法的效果, 选取了几个视频片段进行试验。在实验过程中, 利用查全率和查准率来衡量算法的场景切分效果。查全率和查准率的定义分别为:

$$\text{查全率} = \text{正确检测数} / (\text{正确检测数} + \text{漏检数})$$

$$\text{查准率} = \text{正确检测数} / (\text{正确检测数} + \text{误检数})$$

查全率和查准率越高说明检测的效果越好。

首先是确定阈值 m 。选取五段视频(樱桃小丸子、男儿本色、超人特工队、变相怪杰和长沙七号), 分别在不同的 m 值下对其进行场景切分。根据实验结果, 分析如图 1 所示。

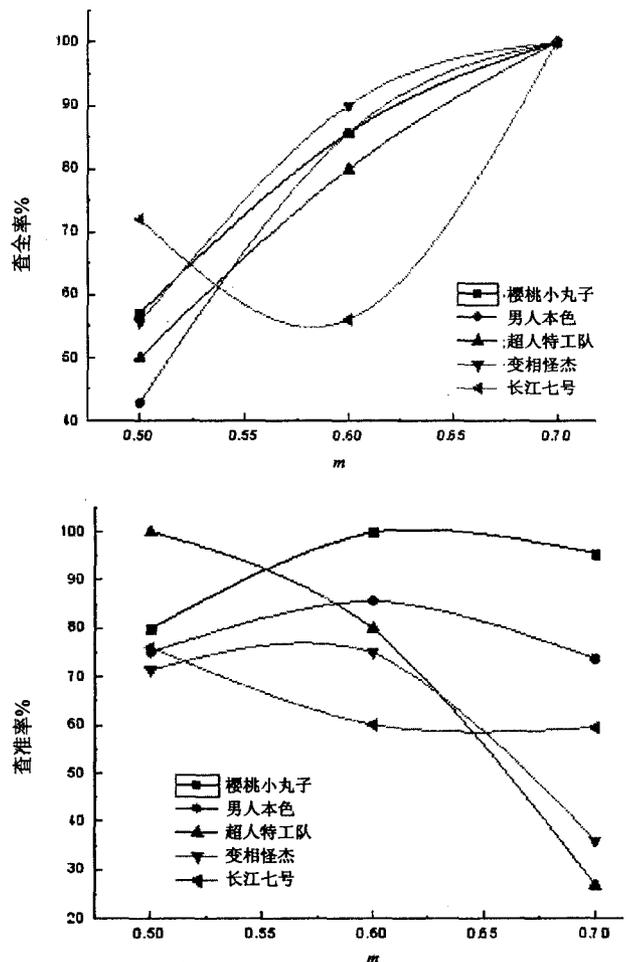


图 1 查全率与查准率与 m 值关系图

从图中可知, 随着 m 的增大, 查全率也越高, 在 $m = 0.7$ 以后, 查全率几乎达到了 100%。然而, 查准率在 $m = 0.6$ 时是最好的, m 越大, 查准率反而会下降。因

此,针对电影视频来说 $m = 0.6$ 时效果较好。

表 1 实验结果数据

实验视频	总场景数	总帧数	时长	正确检出数	漏检数	误检数	查全率	查准率
樱	21	2879	2min	18	3	0	85.7%	100%
男	14	2161	1min30s	12	2	2	85.7%	85.7%
超	4	2495	1min43s	4	1	1	80%	80%
变	9	2471	1min42s	9	1	3	90%	75%
长	25	3215	2min13s	34	1	10	56%	60%

注:樱:樱桃小丸子;男:男儿本色;超:超人特工队;变:变相怪杰;长:长江七号。

从表 1 中可以看出,查全率和查准率都超过了 80%,这说明了基于 HSV 颜色空间模型的场景切分算法取得了很不错的效果。对于不同的 AVI 视频文件,查全率和查准率可能会相差比较大,原因就是阈值 m 的取值影响了场景切分的准确性与完全性。

图 2 的两幅图像是 NBA 球赛视频中相邻的两帧,而在本实验中被切分在不同的子视频里。图 3 的两帧图像在《生化危机》视频中属于不同的场景,却没有被切分出来。出现这样的情况就是因为 m 的取值造成的。进一步实验证明,当 $m = 0.7$ 的时候,图 3 中的两帧图像就被切分到了不同的子视频中。

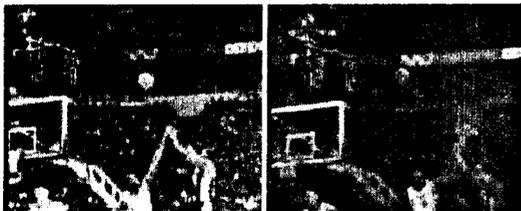


图 2 NBA 比赛视频



图 3 生化危机视频截图

5 结束语

基于 HSV 空间颜色模型的场景切分算法,改进了以往的基于 RGB 颜色模型的灰度直方图的场景切分

算法,首先将图像的 RGB 值转换为 HSV 值,然后量化得出一维特征矢量值,再计算其直方图,通过统计比较计算得出两幅图像的相似性,最后根据其相似性与实验给定的阈值进行比较,以判断是否发生了场景切换。这种算法避免了两幅不同图像之间因 RGB 颜色直方图却有可能相同而无法区分的缺点,从而更加准确地对视频进行场景切分。经过实验证明,这种基于 HSV 空间颜色模型的场景切分算法在对 AVI 视频文件进行场景切分取得了很好的效果,能够比较准确地分割出合适的子视频,达到了预期的目的,为后部分的视频关键帧的提取提供了基本的技术条件。后续的工作将是对渐变镜头进行检测,使得场景的切分更加准确。同时,由于不同场景之间的音频也会不同,因此,检测音频的切换也可以检测出场景是否发生了切分。这也是后续的工作重点。

参考文献:

- [1] Lienhart R, Pfeiffer S, Effelsberg W. Video Abstracting[J]. Communications of ACM, 1997, 40(12): 54-62.
- [2] Yeung M, Yeo B L, Liu B. Extracting Story Units form Long Programs for Video Browsing and Navigation. Proc [C]// IEEE Multimedia Computing & Systems 1996. Hiroshima, Japan: [s. n.], 1996: 296-305.
- [3] Winnemoller H, Olsen S C, Gooch B. Real-Time Video Abstraction[J]. ACM Transactions on Graphics (TOG), 2006, 25(3): 1221-1226.
- [4] 高俊杰. 一种基于交互信息量的视频摘要生成方法[J]. 微电子学与计算机, 2007, 24(2): 128-131.
- [5] 姚洪英, 范铁生. 基于压缩域的视频摘要技术的研究[J]. 鞍山师范学院学报, 2004, 6(2): 67-68.
- [6] 苏新宁. 视频信息索引技术研究进展[J]. 情报学报, 2004, 23(4): 440-446.
- [7] 程文刚, 须德. 一种层次视频摘要生成方法[J]. 中国图象图形学报, 2004, 9(1): 118-123.
- [8] Miyahara M, Yasuhida Y. Mathematical transform of (R, G, B) color data to Munsell(H, V, C) color data[C]//In: Visual Communications and Image Processing '88. Cambridge: [s. n.], 1988: 650-657.
- [9] 王峰, 刘文波. 基于频域的指纹图像增强算法[J]. 计算机技术与发展, 2008, 18(7): 12-16.

(上接第 8 页)

- [7] 张素文, 付薇, 刘明兰. 嵌入式视频图像传输系统的设计与实现[J]. 微计算机信息, 2007(3-2): 60-62.
- [8] Matthew N, Stones R. Linux 程序设计[M]. 第 2 版. 杨晓云, 王建桥, 杨涛, 高文雅等, 译. 北京: 机械工业出版社,

2002.

- [9] Li Jing, Hao Weidong. Research and Design of Embedded Network Video Monitoring System Based on Linux[J]. Computer Science and Software Engineering, 2008(9): 1310-1313.