沈云涛¹, 郭 雷¹, 任建峰¹, 沈玉利^{1,2}

(1. 西北工业大学 自动化学院, 陕西 西安 710072; 2. 湛江海洋大学 信息学院, 广东 湛江 524025)

摘 要:由于非压缩视频序列具有明显的直观性视觉特征,并在实际应用中仍占据相当大的比重,针对在非压 缩视频的处理中仍然存在对镜头变化的类型、镜头渐变和突变的位置分析不够准确的难点,提出了一种时域分 辨率的非压缩域视频分割算法,在提取序列的基于强边缘块的直方图,比较不同序列之间的差异的基础上,采用 多分辨率的算法进行镜头检测。实验证明,这种方法实现了自动视频分割并较好地克服了上述的缺点。 关键词:视频分割;强边缘块;多分辨率;自适应阈值

中图法分类号: TP391 文献标识码: A 文章编号: 1001-3695(2005)10-0151-02

A Multi-resolution Un-compressed Video Partition Algorithm

SHEN Yun-tao¹, GUO Lei¹, REN Jian-feng¹, SHEN Yu-li^{1, 2}

(1. School of Automation, Northwestern Polytechnic University, Xi'an Shanxi 710072, China; 2. School of Information Technology, Zhanjiang Ocean University, Zhanjiang Guangdong 524025, China)

Abstract: In this paper, a multi-resolution algorithm for video partitioning is provided. Firstly, the histogram from the strong edge blocks in the video sequence are computed, and the difference with ² test is computed. Afterwards multi-resolution steps, coarse step and fine step are adopted to detect the video cuts. And further this method can detect the location of video shot and video gradual transmission. Finally, the experimental results show this method is promising. **Key words:** Video Partition; Strong Edge Blocks; Multi-resolution Steps; Adaptive Threshold

xey words. Video Faithdon, Sciolig Edge Blocks, Midru-Tesordulon Steps, Adapuve miesho

目前视频帧之间的区别可以基于以下几种方法:亮度直方 图、灰度直方图、颜色直方图以及纹理、形状之间的区别^[1~4]。 这些方法试图采用一个全局的阈值来划分不同的镜头。这种 方法不足之处在于:全局的阈值需要人为事先给定;仅仅采用 一个阈值会造成镜头分割的错误相比。

本文提出的时域多分辨的非压缩视频分割算法的主导思 想是:提取序列的基于强边缘块的直方图,比较不同序列之间 的差异的基础上,采用不同尺度的窗口来检测视频的镜头。在 计算视频帧间差别的时候,采用了一个基于边缘直方图的算 法,使得计算时间和存储空间大大减少;在检测视频镜头时,先 采用一个大尺度的窗口,检测到视频镜头的分界点;然后利用 小尺度的窗口具体确定视频镜头的位置,阈值主要是根据视频 帧信号的方差和均值自适应调整;最终实现了比较理想的全自 动视频分割。

1 基于强边缘块的直方图的提取

为了便于比较视频中不同帧之间的差别,首先对视频帧进 行基于强边缘块的直方图提取。在传统的视频镜头检测过程 中,颜色和边缘起到了非常重要的作用。但是基于边缘的方法 的计算复杂性非常高,而且如果视频帧中没有明显的边缘,那 么这种方法就会产生很多的错误结果。颜色直方图因其具有 运动鲁棒性及其简单性而被广泛地采用。 本文同时考虑了颜色直方图和边缘信息,为了克服它们各 自的缺点,利用一个在强边缘块附近统计直方图的算法^[1]。 首先定义强边缘块的概念,认为强边缘块就是那些包含了强边 缘互不重叠的块^[5]。而后在强边缘块中提取颜色直方图就可 以认为总是集中在图像中物体的边缘附近,而不是整个图像范 围内。从而抵消了不相关物体颜色的影响,即边缘周围附近的 颜色信息要比背景中的颜色更加重要^[6];通过提取的这种颜 色直方图来获取图像中物体的边缘变化。

假设我们得到 N_a 个强边缘块, $H_s(f_L, j)$ 表示第 s 个边缘块 周围的颜色直方图, 其中 f_L 表示左边先出现的帧, j为灰度值, f_R 表示右边后出现的帧, 那么 f_L 和 f_R 之间的差值采用²距 离:

$$Diff(f_L, f_R) = \sum_{j=0}^{G} \frac{\sum_{s=1}^{N_a} |H_s(f_L, j) - H_s(f_R, j)|}{\sum_{s=1}^{N_a} H_s(f_R, j)}$$
(1)

2 多分辨率的视频镜头分割算法

多分辨率的视频镜头分割算法首先采用一个低分辨的尺度(窗口较大)来检测是否有可能发生镜头的变化,然后采用 一个高分辨率的尺度(窗口较小)来确定镜头变化的具体位 置,同时标记出镜头是突变还是渐变。

2.1 算法的流程

采用两个不同的分辨率尺度 $Step_{coarse}$ 和 $Step_{tine}$ ($Step_{coarse} > Step_{tine}$)表示 f_L 和 f_R 之间出现序列帧的数目,同时必须满足以下条件:

收稿日期: 2004-11-13; 修返日期: 2004-12-23

基金项目: 国家自然科学基金资助项目(60175001)

(2)

 $Length_{gradual} < Step_{coarse} < Length_{shot}$

 $Step_{fine} < < Length_{gradual}$

其中, Length_{gadual}和 Length_{shot}分别表示视频中发生镜头渐变和 镜头突变的最少帧的数目。

本算法首先基于强边缘块直方图的²距离差值(式 1)计 算视频帧之间的差异,利用低分辨率的尺度 *Step*_{coarse}检测镜头 是否发生变化。如果 *Diff*(f_L , f_R) 小于 T_{shot} /2(T_{shot} 是视频镜头 的阈值),就不断地增加 f_L 与 f_R 之间的距离,一直到 *Diff*(f_L , f_R)大于 T_{shot} /2,那么在 f_L 与 f_R 之间就有可能发生视频镜头的 变换。因此下面我们采用高分辨率的尺度来进一步检测镜头 发生突变和渐变的位置,然后利用一个高分辨率的尺度 *Step*_{fine} 不但 可以检测镜头的渐变和突变,同时还可以检测出镜头变化的位 置。它主要通过 *Diff*(f_L , f_R)来实现,单峰说明出现了镜头突 变,平坦的部分表示镜头渐变。下面是多分辨率的视频分割算 法的详细流程:

开始 //初始化变量 Step_{coarse} = large_step_size $Step_{fine} = small_step_size$ L = 1; $R = L + Step_{coarse}$ //主循环 While(R < Sizeof(Video)) Do{ If (Diff $(f_L, f_R) > T_{shot}/2$ // a suspected shot boundary is detected L1 = L; $R1 = L + step_{fine}$ Clear DiffVector; // move by a small step to find the exact boundary While (R1 < = R) Do{ Push Diff (f_{L1}, f_{R1}) into DiffVector L1 = R1 $R1 = L1 + step_{fine}$ } // determine type of shot boundary If(only one peak is found in DiffVector) Abrupt shot change else if (plateau is found in DiffVector) //gradual transmission Plateau(f_{start}, f_{end}) If $f_{\text{start}} = f_{\text{L}}$ Move backwards to get the actual start point of plateau Else $if_{end} = f_R$ Move forwards to get the actual end point of plateau Else false alarm L = R; $R = L + step_{coarse}$ } 结束

2.2 自适应阈值的选取

在一个镜头内,帧之间的差值比较小,但是如果发生镜头 的变化,那么帧间的差值就比较大,而且视频渐变时帧间的差 值要小于视频突变时帧间的差值。因此需要两种阈值来区分 镜头的突变和渐变。本文提出了一个基于区域统计特性的自 适应阈值选取方法,为了去除噪声的影响,先对帧间差值进行 中值滤波,使得对噪声更加具有抗干扰性。

利用一个大小为 K的窗口用于计算窗内的均值和方差:

$$\mu_D(i) = \frac{1}{K} \sum_{m=i-K}^{i-1} \frac{DiffVector}{K} (f_m, f_{m+Step_{coarse}})$$
(3)

$$D_{D}(i) = \frac{1}{K} \sum_{m=i-K}^{i-1} D_{iff}^{\text{median}} f_{m+Step_{\text{coarse}}}(i) - \mu_{D}(i))^{2}$$
(4)

$$= \frac{1}{K} \sum_{m=i-K}^{i-1} D_{iff}^{\text{median}} f_{m+Step_{\text{coarse}}}(i) - \mu_{D}(i))^{2}$$
(4)

$$= \frac{1}{K} \sum_{m=i-K}^{i-1} D_{if}^{median} f_{m+Step_{\text{coarse}}}(i) - \mu_{D}(i))^{2}$$
(5)

$$= \frac{1}{K} \sum_{m=i-K}^{i-1} D_{if}^{median} f_{m+Step_{\text{coarse}}}(i) - \mu_{D}(i) + c_{1} + c_{2} + D_{if}^{median} f_{m+Step_{\text{coarse}}}(i) - \mu_{D}(i) + c_{2} + D_{if}^{median} f_{m+Step_{\text{coarse}}}(i) - \mu_{D}(i) + c_{2} + D_{if}^{median} f_{m+Step_{\text{coarse}}}(i) - \mu_{D}^{median} f_{m+Step_{coarse}}(i) - \mu_{D}^{median} f_{m+Step_{$$

3 实验结果及其分析

表 1 为实验的视频序列,以及每个序列中的镜头切变和渐 变的帧数目。

表1 实验视频序列的描述

视 频 序 列	帧总数	突变的镜头数目	渐变的镜头数目	镜头的数目
新闻(CCTV-1)	215 400	510	230	740
篮球赛(CCTV-5)	108 000	600	291	891
足球赛(CCTV-5)	67 120	300	30	330

表 2 给出了采用自适应阈值的视频分割算法的实验结果。 表 3,表 4 给出了采用不同固定阈值的多分辨率的视频分割算 法的实验结果。

表 2 基于自适应阈值的视频分割结果

视频序列	正确的镜头	丢失的镜头	误检测的镜头	查全率	查准率
新闻(CCTV-1)	690	50	200	93.2%	77.5%
篮 球 赛 (CCTV - 5)	836	55	180	94.4%	82.3%
足 球 赛(CCTV-5)	308	22	47	93.8%	86.7%

表 3 基于固定阈值为 0.1 的视频分割结果

视频序列	正确的镜头	丢失的镜头	误检测的镜头	查全率	查准率
新闻(CCTV-1)	695	45	450	93.9%	60.6%
篮 球 赛 (CCTV - 5)	837	54	1 245	94.4%	40.2%
足 球 赛(CCTV-5)	319	20	450	94.1%	41.5%

表 4 基于固定阈值为 0.2 的视频分割结果

视 频 序 列	正确的镜头	丢失的镜头	误检测的镜头	查全率	查准率
新闻(CCTV-1)	630	105	360	85.1%	63%
篮 球 赛 (CCTV-5)	792	99	430	88.9%	64.8%
足 球 赛 (CCTV-5)	300	39	225	88.4%	57.4%

从表 2~表 4 可以看出,由于新闻节目没有太多的相机移 动以及物体运动,因此,最优阈值比较小。而体育节目等具有 较多的物体移动,因此必须需要一个较大的阈值才能取得更好 的准确率。虽然采样步长对于查准率影响不大,但是对于检测 速度的影响较大。如果步长较大,检测速度较快;反之,如果步 长较小,检测速度较慢。

4 小结

非压缩视频序列中视觉特征提取较为容易现实,其特征相 对较为直观,而且在实际应用中仍然存在大量的非压缩的视频 数据需要处理。对于当前对非压缩视频序列分析存在对镜头 变化的类型、镜头渐变和突变的位置分析不够准确的难题,本 文提出了一种针对非压缩视频进行镜头分割的算法。

在同时考虑了颜色直方图和边缘信息基础上,利用一个在 强边缘块附近统计直方图的算法克服了它们各自的缺点,并通 过提取这种颜色直方图获取图像中物体的边缘变化。之后,采 用多分辨率的视频镜头分割算法,确定了镜头变化的具体位 置,同时标记出镜头是突变还是渐变。辅以一种基于区域特性 的自适应阈值算法,解决了阈值选择的问题。(下转第161页) 转向 S4,S5 进行幅度细化编码。否则,如果还有采样位没有进行过编码处理,下一状态仍在S3 对下一个采样位进行编码,如果所有采样位进行过编码处理则从S3 转S6。

在 S4 读取 bit_{[n}刚变重要性状态字,同时若 bit_{[n}刚变重 要性状态字为1,将它记下来,待到 S6 时更新为0,同时置该采 样位编码标志为1。S4 的下一个状态是 S5,在 S5 进行幅度细 化编码,编码完后,如果还有采样位没有进行过编码处理,下一 状态仍回到 S3 对下一个采样位进行处理,否则转 S6。在 S6 更新 Justsignificant_ram和 Visited_ram,因为系数的重要性没有 发生改变,所以不用更新 Significant_ram。S6 后下一状态转向 S0, S1, S3 中的哪一个原理和清理通道编码时一样。在 S6 如 果当前位平面幅度细化通道编码结束,则输出 Encode_end = 1, 通知系数位平面编码控制模块此次编码完成。位平面编码控 制模块根据收到的 Encode_end = 1,将 Start 置为0 停止幅度细 化编码。在 S6 还要使输出 End_of_column为1,告诉系数位编 码控制模块,幅度细化通道完成了一扫描列数据的编码,使系 数位平面编码控制模块计算器 Count 加1。模块复位或在其他 状态时 End_of_column为0。

4 电路的模拟与验证

本文采用 VHDL 在 RTL 级对 JPEG 2000 系数位平面编码 进行描述,并以 Xilinx XC2V8000-5FF1152 为基础,在 ISE 5.2 下完成设计综合,综合前后用 Modelsim se5.7e 模拟,在 JPEG 2000 编码系统设计的背景下,最高工作频率达 66MHz。设计 模块占用1 385个 Slice(2%),使用 836 个 Slice 触发器(0%)和 2 430个 4 输入 LUT(2%),138 个 TBUF(0%),使用 63 个约束 IOB(7%),五个 BRAM(2%),一个 GCLK(6%)。综合结果在 经布线后进行后模拟,得到的结果与 C 语言的软件结果及前 仿真结果一致。

因为输入的量化后小波系数大部分为 0, 在这样的情况 下, 变换后完成一幅串行输入的 512 ×512 的 8 位灰度图像系 数位平面编码需要约 100ms, 即每秒钟 10 幅。由于 EBCOT 算 法规则可以实现三个方向的子带并行编码,并且一幅图像可以 分成几个 32 ×32 块同时进行并行系数位平面编码, 因此, 采用

(上接第152页)获得较高的查全率和查准率。实验证明,本文 提出的方法实现了自动视频分割并较好地克服了目前非压缩 视频序列分析存在对镜头变化的类型、镜头渐变和突变的位置 分析不够准确等问题。

下一步是把这种方法推广到压缩视频中。由于一般压缩 视频中,都采用了 DCT 变换,使得获取 I 帧的 DC 图像非常容 易,在此基础上,可以提取基于强边缘块的直方图,只是此时的 DC 图像的分辨率更低。

参考文献:

- Idris, S Panchanathan. A Review of Image and Video Indexing Techniques[J]. Journal of Visual Communication and Image Representation, 1997, 8(7):146-166.
- [2] A Nagasaka, Y Tanaka. Automatic Video Indexing and Full Video Search for Object Appearance[J]. Visual Databases Systems, 1992, 2:113-127.
- [3] HJZhang, A Kan Kanhalli, SW Somilar, et al. Automatic Partitio-

XC2V8000-5FF1152 达到每秒 24 幅的视频速率也是很容易的。 本设计可以完成任何 16 位图像的压缩,该结构若用 ASIC 实现 频率会更高速度会更快。

5 结论

本文以 JPEG 2000 图像压缩算法的硬件实现为研究对象, 提出了系数位平面编码的一种 VLSI 结构设计。该结构提高了 压缩数据的吞吐率,最大程度地减少了内部小波系数的存储 量以及小波系数状态字的访问频率。在 JPEG 2000 协议^[1]中, 清理通道编码是在整个位平面完成幅度细化通道编码后才进 行清理通道编码的,两者串行工作。该文中幅度细化通道编码 和清理通道编码并行工作,该结构比完全按照协议规定的串行 工作方式设计的结构运算速度提高 30% 以上。将一幅图像分 几个 32 ×32 块,采用数个并行的该种 VLSI 结构系数位平面编 码,可以保证数据压缩的实时进行。

参考文献:

- [1] oliek M, Christopoulos C, Majani E. JPEG 2000 Part I Final Draft International Standard [S]. ISO/IEC FCD15444-1, ISO/IEC JTC1/ SC29 WG1, 2000.
- [2] Taubman D. High Performance Scalable Image Compression with EB-COT[C]. Proc. of IEEE Int. Conference on Image Processing, Japan: IEEE, 1999.344-348.
- [3] Taubman D, Zakhor A. High Performance Scalable Image Compression with EBCOT[J]. IEEE Transactions on Image Processing, 2000, 9(7):1158-1170.
- [4] Lian Chung-Jr, Chen Kuan-Fu, et al. Analysis and Architecture Design of Lifting-based DWT and EBCOT for JPEG 2000 [C]. Proceedings of International Symposium on VLSI Technology, Systems, and Applications, Hsinchu, Taiwan: IEEE, 2001. 180-183.
- [5] Michael D Adams. The JPEG-2000 Still Image Compression Standard
 [S]. ISO/ IEC JTC 1/SC 29/WG 1 N 2412, http://www.jpeg.org/ jpeg2000guide/papers/papers. html, 2002-11-25.

作者简介:

杨珂(1968-),男,湖南邵阳人,博士研究生,主要研究领域为 ASIC设 计和图像压缩等;刘明业(1934-),男,辽宁营口人,教授,博士生导师, 主要研究领域包括 EDA 技术和多媒体信息处理技术等。

ning of Full Motion Video [J]. ACM Multimedia Systems, 1993, 1 (1):10-23.

- [4] U Gargi, R Kasturi, S H Strayer. Performance Characterization of Video-Shot-Change Detection Methods [J]. IEEE Transactions on Circuits and System for Video Technology, 2000, (1):1-13.
- [5] Han S Song. Scene Change Detection by Feature Extraction from Strong Edge Blocks[J]. Visual Communications and Image Processing, 2002.784-792.
- [6] F Souvannavong, B Merialdo, B Huet. Latent Semantic Analysis for an Effective Region-based Video Shot Retrieval System [C]. New York: Proc. of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval, 2004. 243-250.

作者简介:

沈 云涛(1974-),男,博士生,主要研究方向为图像处理、模式识别等; 郭 雷(1956-),男,国家杰出青年科学基金获得者,教授,博士生导师, 主要研究方向为图像和视频处理、模式识别等;任建峰(1977-),男,博 士生,主要研究方向为图像处理等;沈玉利(1955-),男,副教授,博士, 主要研究方向为模式识别与智能系统等。