

一种改进的视频关键帧提取算法研究

瞿 中^{1,2} 高腾飞¹ 张庆庆¹

(重庆邮电大学计算机科学与技术学院 重庆 400065)¹ (重庆邮电大学移通学院 重庆 401520)²

摘 要 视频镜头分割和关键帧提取是基于内容的视频检索的核心问题。提出了一种改进的关键帧提取算法,其为视频检索奠定了基础,镜头分割部分采用改进直方图方法及基于像素方法的综合方法。首先,通过结合直方图交集及非均匀分块加权的改进直方图方法,根据视频内容将视频分割为镜头;然后,利用基于像素的帧差法,对得到的检测镜头进行二次检测,优化检测结果;最后,在 HSV 颜色空间的基础上,计算每个镜头内每帧的图像熵,从而确定关键帧序列。实验结果表明,提出的改进算法所得到的关键帧结构紧凑且分布均匀。

关键词 视频镜头检测,关键帧提取,帧差法,图像熵,视频检索

中图法分类号 TP391.41 **文献标识码** A

Study on an Improved Algorithm of Video Keyframe Extraction

QU Zhong^{1,2} GAO Teng-fei¹ ZHANG Qing-qing¹

(College of Computer Science & Technology, Chongqing University of Posts & Telecommunications, Chongqing 400065, China)¹

(College of Mobile Telecommunications, Chongqing University of Posts & Telecommunications, Chongqing 401520, China)²

Abstract Video segmentation and keyframe extraction are the core issues of Content-based Video Retrieval(CBVR). We proposed an improved approach of key-frame extraction for video retrieval, adopting the synthesis method based on histogram-based method and pixel-based method in video shot segmentation. In the method, videos are firstly segmented into shots according to video content, by our improved histogram-based method, with the use of histogram intersection and nonuniform partitioning and weighting. Then, the obtained results are secondly detected by pixel-based method to optimize the results. At last, within each shot, key-frames are determined with the calculation of image entropy of every frame in HSV color space. Our simulation results prove that the extracted key frames with our method are compact and uniformly distributed.

Keywords Video shot detection, Keyframe extraction, Pixel-based method, Image entropy, Video retrieval

1 引言

随着 Internet 和数字多媒体技术的飞速发展, Internet 上的数字视频资源与日俱增。如何帮助用户高效地浏览及检索其所需或感兴趣的视频,具有重要的理论意义和实用价值。作为基于内容的视频检索^[1,2]的基础,如何将视频数据组织成更为紧凑的关键帧是本文所要解决的首要问题。

基于关键帧的视频摘要是用于视频索引和视频检索的有效手段,将选定的关键帧作为视频摘要^[3]呈现给用户,以帮助用户高效地认知和理解视频的基本内容。与原始视频相比,关键帧序列由于只是部分视频帧,因此所需的存储空间要小很多,从而减少了视频摘要的网络传输时间,提高了视频检索的效率。

此前,研究人员已提出多种用于提取关键帧的方法^[4],如 Nagasaka A., H. J. Zhang, Yeung M., H. Zhang 及 Hanjalic A., Wolf W., H. S. Chang 提出了不同的算法,但其都存在一

定的局限性。

2 视频镜头分割

镜头分割是视频结构体系的基础,是指精确检测连续镜头间存在的不同类型转换,并区分镜头内运动,以排除其对于视频镜头检测的干扰。根据镜头切换的不同编辑特性,镜头间的变换方式可被划分为切变和渐变^[6]两类。现有的镜头切变检测方法已比较成熟,而渐变检测则非常困难。

颜色可用 RGB 空间的 3 个颜色分量表示,不过 RGB 3 原色间存在很大的相关性,且 RGB 颜色表征与人类主观视觉有着很大的不同。因此,为了使提出的算法中的颜色距离更好地适应人类的视觉特性,引入了 HSV 空间(Hue: H, Saturation: S and Value: V)。与 RGB 空间相比,HSV 空间可更好地代表图像的位置信息。在本文的处理中,首先要将 RGB 颜色空间的各像素转换到 HSV 颜色空间内。

本文采取非均匀量化^[5] HSV 空间的手段,将色调空间 H

到稿日期:2011-10-07 返修日期:2011-12-27 本文受重庆市科委自然科学基金计划项目(CSTC,2010BB2399),重庆市教委科学技术研究项目(KJ110528)资助。

瞿 中(1972—),男,博士,副教授,主要研究方向为数字图像处理、数字媒体技术、云计算等,E-mail:quzhong@cqupt.edu.cn;高腾飞(1988—),男,硕士生,主要研究方向为图像处理、视频处理、仿真建模;张庆庆(1983—),男,硕士生,主要研究方向为图像处理、视频处理、仿真建模。

分为 8 份,而将饱和度空间 S 和亮度空间 V 各分为 3 份,并引入了一维特征矢量以简化颜色特征,最后形成 72bin 的一维直方图。

直方图有很多种应用方式,本文采用了改进的方式——直方图交集^[3]。由于颜色直方图不考虑已知像素的位置信息,完全不同的图像也有可能拥有相似的直方图,因此,本文通过非均匀分块及加权的预处理过程来改进基于直方图的方法,以突出中心部分对于帧间差值的贡献,同时大大降低了镜头内小范围运动对于镜头检测的影响。与传统的全局直方图方法相比,其所得结果更接近于人类的视觉认知。

定义相邻两帧间 H 分量的相似度如下:

$$S_h(t, t-1) = \sum_{i=1}^N \frac{\min(h_t(i), h_{t-1}(i))}{\max(h_t(i), h_{t-1}(i))}$$

式中, $h_t(i), h_{t-1}(i)$ 分别代表第 t 和 $t-1$ 帧 H 分量的直方图, N 表示图像灰度或颜色量化级别, x, y 则分别代表图像帧的行和列像素数。同理, S, V 分量的直方图相似度分别为 $S_s(t, t-1)$ 和 $S_v(t, t-1)$ 。

HSV 空间下,第 t 和 $t-1$ 帧的直方图相似度为:

$$S(t, t-1) = \frac{m_h \times S_h(t, t-1)}{3} + \frac{m_s \times S_s(t, t-1)}{3} + \frac{m_v \times S_v(t, t-1)}{3}$$

人眼对 H 分量有最为敏感的视觉特性。根据 H, S, V 分量的加权比例,在得到 H, S, V 的量化值后,可得到 H, S, V 分量的系数比为 $Q_s Q_v : Q_h : 1$,其中, Q_s, Q_v 分别为 S 和 V 分量的量化级别,在本文中,系数比设定为 $9:3:1$ 。

为了更多地体现 S 及 V 分量对于直方图相似度的贡献,将 m_h, m_s, m_v 3 个分量的加权系数比设定为 $0.9:0.3:0.1$ 。

为了进一步提高镜头检测的检出率和准确率^[7],在利用改进直方图方法检测出镜头后,利用帧差法对所得结果进行过滤,从而构成了用于镜头检测的结合直方图法和帧差法的综合法,它有效地减少了可能由基于直方图的方法所带来的漏检及误检情况。

此外,针对剧烈光照变化的情况,尤其是闪光,由于其持续帧数较少,且人类对于视觉媒体,如动漫、电影等存在视觉暂留作用(其精确值为 24fbps),因此本文将帧数小于 20 的镜头重新划归到上一个镜头中,使之适合于人类视觉特性。

3 关键帧提取算法

视频镜头分割之后,就可以从分割出的镜头中提取用以描述镜头主要内容的的关键帧。

在用平均值法求取最接近镜头平均值关键帧的基础上,本文将图像熵引入关键帧提取中^[8],选取镜头中的图像熵最大的帧作为关键帧,这样得到的关键帧就可以有效地反映镜头的主要内容。

对一幅图像来说,其直方图可被认为是一种概率密度函数。假定 h_k 表示整幅图像中像素值为 k 的像素所占的比例,考虑到当 $h_k=0$ 的实际情况,应加上约束条件:当 $h_k=0$,则 $\log h_k=0$ 。因此,图像熵可定义为 $E(f_i) = -\sum_{k=1}^n h_k \log h_k$ 。

由于 H, S, V 这 3 个分量的加权系数比定为 $\alpha: \beta: \gamma = 0.9:0.3:0.1$,因此求得综合图像熵:

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

式中, $E(f_i)_H, E(f_i)_S, E(f_i)_V$ 分别表示 H, S, V 分量上的图像熵。

综上所述,本文提出的镜头检测和关键帧提取改进算法为:

(1) 计算相邻两帧直方图的交集,并与设定阈值进行比较,以判断是否发生镜头变化。设定直方图相似度阈值的取值范围为 $0.75 \sim 0.95$,并由大量对比实验得知当阈值设为 0.9 时,该方法综合效果最优。

(2) 对于(1)所得结果,利用帧间灰度/颜色差值进行镜头边界的二次检测。结合非均匀分块加权(分为 9 块,中心部分所占比例最大,且权值和为 1)的方法,分别计算每块像素差值并与预设的分块帧差阈值(取值范围为 $10 \sim 30$)作比较以进行标记;而后对每块的标记变量加权求和,并与设定的分块加权阈值(取值范围为 $0.0 \sim 0.4$)进行比较,以判断是否发生镜头变化。

(3) 考虑到强烈光照变化的情况,尤其是闪光,本文将帧数小于 20 的镜头重新划归到上一个镜头中。

(4) 计算镜头内每帧的图像熵值,从每个镜头中提取综合图像熵最大的帧作为代表该镜头的关键帧。

4 仿真实验及结果分析

本文在检测样本的选择上,考虑到视频选材的普遍性,选择了 4 种类型的视频,包括广告(innisfree_01、Sasaki-nozomi)、新闻(新闻联播)、节目预告(11)及 MV(Doushite Suki Nan Darou_clip),共 5 个视频。并利用检出率(Recall)和准确率(Precision)来衡量视频镜头检测算法的检测效果。

$$\text{检出率 } R = N_c / (N_c + N_m) \times 100\%$$

$$\text{准确率 } P = N_c / (N_c + N_f) \times 100\%$$

式中, N_c, N_m, N_f 分别为镜头的正确检出数、漏检数和误检数。

通过最小函数法计算两帧直方图的交集以衡量其相似度,并与设定阈值 T 进行比较,从而判断是否存在场景切换。相邻两帧直方图的相似度定义为:

$$\text{Sim} = \frac{1}{xy} \sum_{v=0}^V \frac{\min(H(I_t, v), H(I_{t-1}, v))}{\max(H(I_t, v), H(I_{t-1}, v))}$$

考虑到传统的帧差法对于视频中物体/相机的运动十分敏感,容易导致误检测的缺点,本文所述的帧差法结合了非均匀分块加权的思想,分别对每块逐点地计算像素差值,并与预设的分块帧差阈值进行比较,以进一步标记;而后对每块的标记变量进行加权求和,并与设定的分块加权阈值进行比较以判断是否存在镜头切变。相邻两帧之间的对应块的帧差定义为:

$$Fd = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N |f_n(i, j) - f_{n+1}(i, j)|$$

为了定量评估本文的镜头分割算法与直方图法及帧差法的比较,分别对这些算法进行了测试,其实验结果如表 1 所示。

从表 1 可以看出,由综合法得出的镜头检测准确率比两种传统方法要高,不过其镜头的检出率却受制于分别由这两种方法得到的结果。以表中最后一段 MV 视频“Doushite Suki

Nan Darou_clip”为例,由于其中存在大量的快速切变、渐变、镜头内主体运动以及一定的镜头内光照变化,因此在准确率方面,两种传统方法得到的结果都比较低,尤其是直方图法对颜色变化过于敏感,从而造成了大量误检。而本文提出的综合法综合了两种算法的优点,在准确率方面有了较大提高。

表1 视频镜头检测结果

视频名称	视频镜头检测方法	镜头数	检出数	误检数	检出率	准确率
[广告] innisfree_01	综合法	12	13	1	100.0%	92.9%
	直方图法	12	13	1	100.0%	92.9%
	帧差法	12	14	2	100.0%	85.7%
[广告] Sasaki-nozomi	综合法	9	6	0	66.7%	100.0%
	直方图法	9	15	6	100.0%	60.0%
	帧差法	9	6	0	66.7%	100.0%
[新闻] 新闻联播	综合法	13	15	2	100.0%	86.7%
	直方图法	13	15	2	100.0%	86.7%
	帧差法	13	18	5	100.0%	72.2%
[预告] 11	综合法	4	4	0	100.0%	100.0%
	直方图法	4	4	0	100.0%	100.0%
	帧差法	4	5	1	100.0%	80.0%
[MV] Doushite Suki Nan Darou_clip	综合法	26	27	4	88.5%	85.2%
	直方图法	26	57	31	100.0%	45.6%
	帧差法	26	33	9	85.7%	72.7%

此外,本文还对连续拍摄情况的视频进行了实验。连续拍摄属于镜头渐变的一种,一般认为只存在一个镜头。采用综合法来检测这类视频,其结果符合要求。

为了检测用于关键帧提取的熵值法的结果,本文对测试视频中的每个镜头分别提取其首帧、尾帧、中间帧作为关键帧,并将其与由熵值法得到的关键帧,分别形成关键帧索引并进行直观比较。

本文从“Doushite Suki Nan Daro”MV 视频中裁剪出一段作为测试视频,以准确率最高的方法——综合法进行镜头检测。

该视频共有 2705 帧,分辨率为 480 * 320,时长 90.26s,包括视频镜头检测和关键帧提取在内的总处理时间为 47.56s,可得每帧处理时间为 17.583ms。

从图 1—图 4 可以看出,熵值法提取的关键帧可以更好地代表视频内容,且关键帧分布也最为均匀。本文提出的综合法对直方图法中帧间差异的计算进行了一定的改进,结合了非均匀分块加权的思想,利用帧差法对所得结果进行了二次检测,并考虑到可能出现的闪光对镜头检测的影响,在没有显著增加时间及计算复杂度的同时,使镜头的检出率和准确率都有了一定程度的改进,从而提升了关键帧提取的精确性。



图1 熵值法所得关键帧



图2 首帧作为关键帧



图3 中间帧作为关键帧



图4 尾帧作为关键帧

结束语 本文对基于内容的视频检索技术中的关键帧提取技术进行了研究,提出了一种改进的关键帧提取算法,实验结果表明该方法有以下几个优点:(1)计算和时间复杂度较低;(2)有效提高了镜头检测的准确率;(3)所得关键帧能够更好地反映镜头内容,且分布较为均匀,从而为顺利进行后续的视频摘要及视频检索工作奠定了良好的基础。

参考文献

- [1] Collomosse J P, McNeill G, Qian Y. Storyboard sketches for Content Based Video Retrieval[C]// IEEE 12th International Conference on Computer Vision, 2009: 245-252
- [2] Huang Zi, Li Yi-jun, Shao Jie, et al. Content-Based Video Search: Is there a Need, and Is it Possible[C]// International Workshop on Information-Explosion and Next Generation Search, 2008: 12-19

[3] Eldib M Y, Zaid B, Zawbaa H M, et al. Soccer video summarization using enhanced logo detection[C]//16th IEEE International Conference on Image Processing. 2009;4345-4348

[4] Kelm P, Schmiedeke S, Sikora T. Feature-based video key frame extraction for low quality video sequences[C]//10th Workshop on Image Analysis for Multimedia Interactive Services. 2009;25-28

[5] Spahiu C S. A multimedia database server for information storage and querying[C]//International Multiconference on Computer Science and Information Technology. 2009;517-523

[6] Gunal E S, Canbek S, Adar N. Gradual shot change detection in

soccer videos via fractals[C]//International Conference on Electrical and Electronics Engineering. 2009,2;125-128

[7] Tapu R, Zaharia T, Preteux F. A scale-space filtering-based shot detection algorithm[C]//2010 IEEE 26th Convention of Electrical and Electronics Engineers in Israel. 2010;919-923

[8] Ren Li-ping, Qu Zhi-yi, Niu Wei-qin, et al. Key frame extraction based on information entropy and edge matching rate[C]//2010 2nd International Conference on Future Computer and Communication. 2010,3;91-94

[9] 姚杰, 谭建明, 唐超, 等. 基于运动估计的视频质量评价[J]. 重庆理工大学学报:自然科学版, 2012, 26(5):74-78

(上接第 280 页)

从结果中知道,改进后的算法相比于经典水平集算法,得到的分割结果更完整。这是因为经典水平集算法只利用了图像的边缘梯度信息,对噪声敏感,不易得到理想的分割结果。改进后的算法同时利用了边缘梯度信息和区域信息,保持了全局优化的特点,能有效地克服噪声的干扰,得到的分割结果更接近原始图像。

通过两组对比实验结果可以看出,融入了区域信息的改进拉普拉斯水平集方法得到的结果比只包含边缘梯度信息的形状检测水平集分割方法得到的效果更好,更接近完整的原始图像。这是因为改进后的拉普拉斯水平集方法融入了区域信息,涉及图像函数的数据是全图像定义域范围,具有全局性;而且在一些图像边缘比较模糊的地方,改进后的拉普拉斯水平集算法也能够得到比较理想的分割结果。

结束语 本文在 ITK 中拉普拉斯水平集算法的基础上,结合图像的区域信息,重新定义了驱动水平集表面演化的速度函数,提出了一种新的医学图像分割方法,该方法充分利用了图像的区域信息和边缘梯度信息。从实验结果可以看出,融入了区域信息的拉普拉斯水平集算法相对于只利用边缘梯度的算法,可以更准确地分割图像,得到的分割结果更连续;而且,在本算法的计算中没有自由参数,降低了人工分割的难度。下一步需要进行的工作是,提高算法的运行速度并将其推广至三维体数据的分割,将算法与 ITK、VTK 平台结合,开发出准确、科学的医学影像处理及分析系统。

参 考 文 献

[1] 田捷. 集成化医学影像算法平台[M]. 北京:清华大学出版社, 2004

[2] Osher S, Sethian J A. Fronts propagating with curvature dependent speed; algorithms based on Hamilton-Jacobi formulation [J]. Journal of Computer Physics, 1988, 79(1):12-49

[3] Chan T F, Vese L A. Active contours without edges [J]. IEEE Transactions on Image Processing, 2001, 10(2):266-277

[4] Mumford D, Shan J. Optimal approximations by piecewise smooth functions and associated variational problems[J]. Communication of Pure Applied Mathematics, 1989, 42(5):67-79

[5] 李俊, 杨新, 施鹏飞. 对水平集方法鲁棒初始化的双向快速步进法[J]. 系统仿真学报, 2001, 13:164-167

[6] 李传龙, 李颖, 兰国新. 一种新的窄带快速区域水平集 C-V 模型

图像分割方法[J]. 计算机科学, 2011, 38(12):17-35

[7] Rafael C G. Digital image processing(Second Edition) [M]. BEIJING: Publishing House of Electronics Industry, 2003

[8] 李俊, 杨新, 施鹏飞. 基于 Mumford-Shah 模型的快速水平集图像分割方法[J]. 计算机学报, 2002, 25(11):1176-1183

[9] The ITK software guide[EB/OL]. <http://www.itk.org>, November 2005

[10] Wang Bin, Gap Xin-bo, Tao Da-cheng, et al. A unified tensor level set for image segmentation [J]. IEEE Transactions on Systems, man, and Cybernetics, Part B; Cybernetics, 2010, 40(3):857-867

[11] Zhang Kai-hua, Song Hui-hui, Zang Lei. Active contours driven by local image fitting energy [J]. Pattern Recognition, 2010, 43(4):1199-1206

[12] Ianniello S, Mascio A D. A self-adaptive oriented particles Level-Set method for tracking interfaces [J]. Journal of Computational Physics, 2010, 229(4):1353-1380

[13] 谢意, 杨玲. 自适应模型的水平集图像分割方法[J]. 计算机工程与应用, 2011, 47(27):221-224

[14] Esdoglu S, Smereka P. A variational formulation for a level set representation of multiphase flow and area preserving curvature flow [J]. Communication in Mathematical Sciences, 2008, 6(1):125-148

[15] Li C M, Kao C Y, Gore J C, et al. Minimization of region-scalable fitting energy for image segmentation [J]. IEEE Transactions on Image Processing, 2008, 17(10):1940-1949

[16] Chen Yao-tien. A level set method based on the Bayesian risk for medical image segmentation [J]. Pattern Recognition, 2010, 43(11):3699-3711

[17] Mohamed B S, Amar M, Ismail B A. Effective level set image segmentation with a kernel induced data term [J]. IEEE Transactions on Image Processing, 2010, 19(1):220-232

[18] Zhou Hui, Yuan Yuan, Lin Fa-quan. Level set image segmentation with Bayesian analysis [J]. Neurocomputing, 2008, 71(10):1994-2000

[19] Manolya E, David B. Technical Section: Interactive free-form level-set surface-editing operators [J]. Computers and Graphics, 2010, 34(5):621-638

[20] Tom G, Xavier B, Stanley O. Geometric applications of the Split Bregman method; segmentation and surface reconstruction [J]. Journal of Scientific Computing, 2010, 45(1):272-293