

基于谱残差法显著图的视频镜头边界检测

况盛坤, 王晓红

(上海理工大学, 上海 200093)

摘要: **目的** 针对传统视频镜头边界检测算法精度低和较复杂等缺点, 提出一种基于谱残差显著图和分块灰度直方图相结合计算似然比的算法。 **方法** 为了提取有效结构信息作为特征来比较帧间差异, 采用谱残差方法得到各帧图片的显著图。将显著图分块并计算各块的灰度直方图数据, 进一步提高检测的精度。计算各帧似然率后用文中提出的帧间差指标计算比较帧间差异。 **结果** 所提方法的综合检测精度能达到93.82%, 对常见影响因素下相邻未变化帧的相似性保持在99%以上。 **结论** 实验结果表明, 文中算法的检测精度高于传统方法, 过程相对简单, 对常见影响因素具有较强的鲁棒性。

关键词: 镜头边界检测; 谱残差; 显著图; 分块直方图; 似然比

中图分类号: TS807 **文献标识码:** A **文章编号:** 1001-3563(2016)07-0136-05

Video Shot Boundary Detection Based on Spectral Residual Saliency Map

KUANG Sheng-kun, WANG Xiao-hong

(University of Shanghai for Science and Technology, Shanghai 200093, China)

ABSTRACT: In order to overcome shortcomings such as the low accuracy and the high complexity of the traditional approach of video shot boundary detection, an algorithm using spectral residual saliency map combined with sub-block gray histogram to calculate likelihood ratio was proposed in this paper. First, to extract effective structure information as the feature to compare differences between frames, the spectral residual approach was used to obtain the structure saliency map of frames. Second, the frames were segmented into necessary subblocks to calculate all the histogram data which further improved the precision. The likelihood ratio of each frame was calculated then the frame-to-frame difference index proposed in this paper was computed to calculate and compare frame-to-frame differences. The detection precision of the proposed method can reach 93.82%, and the similarity of unchanged neighboring frames maintained as more than 99%. Experimental results showed that the method was more effective which had a higher precision and relatively simpler process, and it had a strong robustness against common influencing factors.

KEY WORDS: shot boundary detection; spectral residual; saliency map; sub-block histogram; likelihood ratio

视频镜头是由一台摄像机持续采集内容得到在时间和空间上具有连续性的一系列单帧图片组成^[1], 一个视频镜头(下文中简称“镜头”)是视频内容的基本单元。镜头作为视频的重要组成部分, 其边界检测(Shot boundary detection, SBD)对视频检索、视频压缩、视频质量评价等视频应用技术具有非常重要的应用意义。镜头变化一般分为镜头的切变和渐变2种类

型, 切变是发生在相邻两帧图片之间, 然由一个镜头转变到另一个镜头; 而渐变是后期采用的视频剪辑处理导致的镜头发生转变, 其过程往往需要过渡连续几帧图片。镜头边界检测根据判断相邻2帧图片之间的相似性来确定, 其检测步骤一般包括: 特征选取、确定帧间相似性的算法、判定阈值的选取以及窗口大小的确定^[2]。文中首次将谱残差法提取的显著图作为检测

收稿日期: 2015-07-10

作者简介: 况盛坤(1991—), 男, 湖北人, 上海理工大学硕士生, 主攻颜色科学技术。

通讯作者: 王晓红(1971—), 女, 陕西人, 上海理工大学教授, 主要研究方向为印刷技术、颜色科学及应用。

特征,并提出一个新的检测指标用于SBD中。

SBD算法中频繁被采用的基础特征有像素强度、颜色直方图、边缘信息、运动特性和关键点描述等^[3]。由于计算的低功耗,颜色直方图是常被用来计算帧间相似性的指标。边缘信息则表现了每一帧图片的细节和结构信息,然而直接用边缘的相似性来度量镜头是否突变在计算上的效率较低,而且精度不够高^[4]。Zahib等^[4]提出一种基于边缘变换率(ECR)的检测方法,该方法主要的不足是计算耗时并且有较高的噪声敏感性。Bruno等^[5]将光流法计算图片帧间运动矢量的方法用于SBD中,但运动估计的复杂性通常较高,不便于实时性的检测。Sinnu Susan等^[6]提出的基于镜头内背景连续的理论提出的SBD方法比较新颖,不过其分块数过多,整体过程也较复杂,导致其检测精度不够高。Lienhart等^[7]比较了基于不同特征的SBD方法的优劣,一些较复杂的特征,如边缘信息,其表现并不比基本的颜色直方图更优秀,但是,有些复杂的方法在某些情况下也能获得较为良好的效果。FU等^[8]提出的基于HSV色空间的SBD算法中直接统计通道中分块后的差异值求平均,然后将三通道的值相乘。该算法较复杂而且精度不够高。在计算获得帧间相似性之后,相似性阈值的选取通常有全局阈值^[9]和多重阈值^[10]。算法过程中计算窗口大小的确定直接对计算的速度有着直接的影响^[11],因此,为了获得更广阔的适用性,融合多种不同特征的SBD方法在发展趋势上的必然性就显得尤为明显。

既然同一镜头内相邻帧的结构信息是极其相似的,文中首先采用谱残差提取的显著区域作为特征来进行相似性比较,有效地提取了各帧的结构显著信息;然后结合分块直方图,利用了其在SBD方法中一贯表现出的高精度;最后采用似然率值计算帧间相似性,并在此基础上提出一个新的帧间突变检测指标。文中提出的多特征相结合的SBD方法整体过程比较简单,并且对常见因素的影响具有较强的鲁棒性。

1 谱残差法计算显著区域

侯晓迪等^[12]于2007年提出一种简单的基于谱残差方法计算视觉显著图的模型。从信息理论的角度出发,1幅图像的信息由新颖部分(图像的显著部分)和先验知识部分(编码系统中应当去掉的残余信息)组成。自然图像统计学中有一条 $1/f$ 法则表明自然图像并不是随机的,而是处于一种高度可预测的分布状态,即幅度谱的期望与频率成反比,见式(1)。

$$E\{A(f)\} \propto 1/f \quad (1)$$

式中: f 表示图像在频域中的频率; $A(f)$ 表示自然图像经过傅里叶变换后得到的幅度谱; $E\{A(f)\}$ 表示幅度谱的期望。通过去除图像的统计残余部分得到新颖部分就会突出图像的显著区域。

在对图像的幅度谱取对数后得到图像的对数谱,统计比较大量图像的对数谱图后发现不同自然图像的对数谱都有相似的趋势。同时将多幅图像的对数谱取平均后发现非常平滑,即具有局部线性的性质。因而得出以下理论:一幅图像的对数谱图中波动的部分就是显著部分,而平滑线性的区域是冗余部分。故将一幅图像的对数谱减去区均值滤波后的对数谱图即可计算出显著图。具体步骤如下:

首先对一幅图像计算二维傅里叶变换,将其从空间域转换到频率域,同时可得到幅值谱。接着对幅值谱取对数得到对数谱图,该过程见式(2)~(4)。

$$A(f) = \partial(F(\text{Img})) \quad (2)$$

$$P(f) = \angle(F(\text{Img})) \quad (3)$$

$$L(f) = \log(A(f)) \quad (4)$$

式中: F 代表二维傅里叶变换; ∂ 和 $\angle$ 分别代表幅值和相位; $P(f)$ 和 $L(f)$ 分别表示相位谱和对数幅度谱(下文简称“对数谱”)。接着采用 3×3 的均值滤波器 $h_n(f)$ 将对数谱做平滑处理,见式(5)。

$$V(f) = L(f) * h_n(f) \quad (5)$$

接着将对数谱和平滑处理后的对数谱做差得到谱残差 $R(f)$,见式(6)。

$$R(f) = L(f) - V(f) \quad (6)$$

为了凸显显著区域,而不仅仅只是一个轮廓,原算法中是以 $R(f)$ 作为实部,本文则将 $V(f)$ 作为实部,原相位谱 $P(f)$ 作为虚部进行二维傅里叶逆变换,结果取平方,最后再用高斯模糊滤波器 $g(x)$ 处理得到显著区域,计算公式如式(7)。

$$S(x) = g(x) * F^{-1}[\exp(V(f) + P(f))]^2 \quad (7)$$

该过程以及下文算法主要步骤:(1)依次提取相邻的两帧图片信息;(2)分别提取其显著区域;(3)对显著图分块并提取其直方图数据;(4)计算对应子块间的似然率值,求取平均得到帧间似然率;(5)判定是否为镜头边界。

2 显著图分块及突变检测值的计算

2.1 显著图分块原理

显著图提取成功之后,并不是直接用来计算其相

似性,而是进一步提取显著图的灰度直方图,然后采用似然率指标来计算显著灰度直方图之间的差异。尽管不同的2幅自然图像拥有相同直方图的概率不大,但采用分块直方图^[13-14]则完全可以避免该问题。其原理如下,假设图1a和图1b是2帧连续的图片,未分块之前的两帧图具有相同的直方图数据,但将图分为4块后,各自对应的每一块却都不一样。因此采用分块计算的方法避免了由于巧合导致的2帧图像的直方图数据一致或者近似一致,从而保留了数据的在原图上的空间特性。统计所有对应子块的相似性能够直观有效的检测出镜头突变结果。

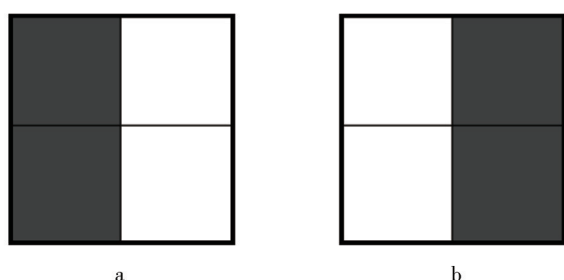


图1 分块统计帧间相似性

Fig.1 Diagram of statistical frames' similarity in each subblock

2.2 似然率值的计算

如果分别用 $\sigma_{k,i}$ 和 $\mu_{k,i}$ 代表第 i 帧第 k 子块直方图数据的方差和均值,那么显著图子块间似然率(likelihood ratio)^[15-16]的计算公式见式(8):

$$\tau_k = \frac{\left[\frac{\sigma_{k,i} + \sigma_{k,i+1}}{2} + \left(\frac{\mu_{k,i} - \mu_{k,i+1}}{2} \right)^2 \right]^2}{\sigma_{k,i} \cdot \sigma_{k,i+1}} \quad (8)$$

将各子块的似然率值相加得到一帧图片与前一帧图片相比较得到的似然率值:

$$\gamma = \sum_{k=1}^B \tau_k \quad (9)$$

式中: τ_k 为第 k 块的似然率值; B 为分块数。整个视频的平均似然率值 D 的计算公式则如下式。

$$D = \frac{1}{N-1} \sum_{i=2}^N \gamma_i \quad (10)$$

2.3 判定阈值的确定

通常情况下,同一镜头内的任意相邻2帧的变化及其微小,其似然率值一般都稳定在平均值左右的范围内。而一旦镜头发生转变时,变化帧的似然率值会突然增大,相邻2帧图片的相似性会降低。对于渐变镜头,连续多帧图片的似然率值虽然大于平均值,但

会在一定的范围内变动。而对于切变镜头,该值跃升的幅度则会相当大(尤其相对于前一帧和后一帧),因此对于切变镜头的检测,文中提出公式(12)来突出切变帧的帧间差值。

$$C_i = \frac{\gamma_i - BS}{(\gamma_{i-1} + \gamma_{i+1})/2 - BS} \quad (11)$$

$$C = \frac{1}{N-1} \sum_{i=2}^N C_i \quad (12)$$

式中: S 为最佳似然率值(2幅相同图片的似然率), $S=1$; γ_i 为第 i 帧的似然率值; C_i 为第 i 帧最终的切变检测值; C 为整个视频的切变检测均值。切变检测阈值如式(13)。

$$T_1 = C \alpha \quad (13)$$

当某一帧的帧间差大于 TH_1 时,可判定为切变镜头,式中 α 为切变系数。

将检测完之后的切变帧对应的似然率值归零。由式(10)可得到视频的平均似然率值,而当连续几帧的似然率值大于 TH_2 时,可判定为渐变镜头,见式(14)。式中 β 为渐变系数。

$$T_2 = D \beta \quad (14)$$

3 试验结果及分析

为了保证文中算法的有效性,实验采用4段MPEG格式的视频,详细信息见表1。TRECvid是由美国国家标准技术研究所(National Institute of Standards and Technology, NIST)于2003年发展成的视频检索项目。该项目每年都发布一系列用于不同目的的标准视频,TRECvid中部分标准视频即用于视频边界检测。

表1 实验所用4段视频的信息

Tab.1 Information of the four experimental videos

视频	帧数	镜头边界数	切变数	渐变数
视频1	38 313	261	248	13
视频2	49 815	147	139	8
视频3	44 991	104	95	9
视频4	35 802	63	58	5

3.1 试验结果对比

NIST有一套标准的镜头边界检测评估方法,以镜头变换的查全率、查准率和一个综合参数来进行评估^[17],参数定义见式(15)~(17):

$$R = \frac{N_c}{N_c + N_m} \times 100\% \quad (15)$$

$$P = \frac{N_c}{N_c + N_f} \times 100\%$$
 (16)

$$F(R,P) = \frac{2RP}{R + P}$$
 (17)

式中: N_c 和 N_m 以及 N_f 分别为检测的正确数、漏检数以及错检数; R,P,F 分别为查全率、查准率以及这2个参数的综合比较参数。

文中算法和文献[6]的算法用于视频边界检测结果见表2和表3。

表2 指标R和P结果对比

Tab.2 Comparison of experimental results of indexes R and P

%

视频	文献[6]算法		文中算法	
	R	P	R	P
视频1	91.95	85.71	97.7	92.72
视频2	91.83	81.81	96.59	91.61
视频3	89.42	84.54	96.15	90.09
视频4	88.88	82.35	96.82	92.53
切变检测	91.48	84.44	97.59	91.97
渐变检测	85.71	78.94	91.42	91.42

表3 指标F结果对比

Tab.3 Comparison of experimental results of index F

%

视频	文献[6]算法		文中算法	
	R	P	R	P
视频1	91.95	85.71	97.7	92.72
视频2	91.83	81.81	96.59	91.61
视频3	89.42	84.54	96.15	90.09
视频4	88.88	82.35	96.82	92.53
切变检测	91.48	84.44	97.59	91.97
渐变检测	85.71	78.94	91.42	91.42

从表2和表3可以看出文中算法在查全率、查准率以及综合比较指标3方面都优于文献[6]算法。对于切变镜头和渐变镜头的检测精度分别能达到94.7%和91.42%。综合比较指标的整体平均值也能达到93.82%。由此可以看出文中提出的算法要优于文献[6]算法,有一定的实际应用意义。

3.2 鲁棒性验证

一个好的帧间相似性算法不仅要低功耗、高精度,同样应该具有不受外界因素影响的特性,最常见的一些因素有:光强度变化、颜色突变、扭曲变形、噪声等。即同一个镜头内连续帧的高度相似性在以上

因素的影响下能否继续保持也是极其重要的。为了验证文中算法对以上因素的影响是否具有鲁棒性,文中随机选取了40组同一镜头内相邻2帧图片进行相似性抗干扰鲁棒性验证实验,其中环境亮度变化中变暗处理和变亮处理是各20组。对每一测试组的第二帧分别进行如图2所示的干扰,然后用文中的算法进行检测。



图2 常见影响因素干扰

Fig.2 Interference with common influencing factors

经过对文中算法检测,文中算法对于常见因素影响下的鲁棒性还是较高的,大部分情况下对于相似性的检测能保持在99%以上。亮度变化影响下检测的平均相似度能够达到99.72%,颜色变化影响下能达到99.95%,内容变形下能达到99.65%,但是对于噪声影响的鲁棒性稍微偏低,不过检测的平均相似性也能达到97.87%,并不影响镜头边界的检测结果。

4 结语

文中采用显著特性和分块直方图相结合进行边界镜头检测,谱残差法提取显著区域保留了原图的细节和重要信息,分块进一步考虑了视频内容在空间上的分布特性,而计算灰度直方图数据简化了算法过程。在似然率值的基础上改进的视频镜头突变检测指标表现出了较高的精度,并且文中算法对于常见的影响因素有较高的鲁棒性,不足的是对于噪声的鲁棒性偏低,这也是下一步工作需要改进的地方。

参考文献:

[1] DAVENPORT G, SMITH T A, PINCEVER N. Cinematic Primitives for Multimedia[J]. IEEE Computer Graphics Applications, 1991, 11(4): 67—74.

[2] GUDER M, CICEKLI N K. Dichotomic Decision Cascading

- for Video Shot Boundary Detection[J]. IEEE International Symposium on Multimedia, 2013, 13(1): 227—230
- [3] LIENHART R. Reliable Transition Detection in Videos; A Survey and Practitioner's Guide[J]. International Journal of Image and Graphics (IJIG), 2001, 1(3): 469—486.
- [4] ZABIH R, MILLER J, MAI K. A Feature-based Algorithm for Detecting and Classification Production Effects[J]. Multimedia Systems, 1999, 7: 119—128.
- [5] BRUNO E, PELLERIN D. Video Shot Detection Based on Linear Prediction of Motion[C]// Multimedia and Expo (IC-ME), 2002: 289—292.
- [6] THOMAS S, GUPTA S, VENKATESH K S. An Energy Minimization Approach for Automatic Video Shot and Scene Boundary Detection[C]// Proceedings of 2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Kitakyushu, Japan, 2014, 14(5): 297—300.
- [7] LIENHART R W. Comparison of Automatic Shot Boundary Detection Algorithms[C]// Proc SPIE Image Video Process VII, 1999, 3656: 290—301.
- [8] FU Qi-Fan, ZHANG Yi-Chun, XU Li-Hong, et al. A Method of Shot-boundary Detection Based on HSV Space[C]// Proceedings of 2013 Ninth International Conference on Computational Intelligence and Security, 2013, 13(3): 219—223.
- [9] QÛENOT G, MORARU D, BESACIER L. CLIPS at TRECvid: Shot Boundary Detection and Feature Detection[J]. Proceedings of TRECvid, 2003: 35—40.
- [10] CERNEKOVA Z, PITAS I, NIKOU C. Information Theory-based Shot Cut/Fade Detection and Video Summarization [J]. IEEE Transactions on Circuits & Systems for Video Technology, 2006, 16(1): 82—91.
- [11] MOHANTA P P, SAHA S K, CHANDA B. A Model-Based Shot Boundary Detection Technique Using Frame Transition Parameters[J]. IEEE Transactions on Multimedia, 2012, 14(1): 223—233.
- [12] HOU Xiao-Di, ZHANG Li-qing. Saliency Detection: A Spectral Residual Approach. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[M]. Minneapolis: IEEE Press, 2007.
- [13] 巢娟. 基于内容的视频帧间篡改被动检测算法研究[D]. 上海: 上海交通大学, 2012.
- CAO Juan. Research on Digital Video Inter-frame Forgery Passive Detection Algorithm based on Visual Content[D]. Shanghai: Shanghai Jiao Tong University, 2012.
- [14] 刘帅师, 田彦涛, 万川. 基于 Gabor 多方向特征融合与分块直方图的人脸表情识别方法[J]. 自动化学报, 2011, 37(12): 1455—1463.
- LIU Shuai-shi, TIAN Yan-tao, WAN Chuan. Facial Expression Recognition Method Based on Gabor Multi-orientation Features Fusion and Block Histogram[J]. Acta Automatica Sinica, 2011, 37(12): 1455—1463.
- [15] 章毓晋. 基于内容的视觉信息检索[M]. 北京: 科学出版社, 2003.
- ZHANG Yu-Jin. Visual Information Retrieval Based on Content[M]. Beijing: Science Press, 2003.
- [16] 贾伟. 基于颜色直方图的视频突变镜头边界检测[J]. 机械设计与制造工程, 2014, 43(5): 67—70.
- JIA Wei. The Video Shot Boundary Cut Detection Based on Color Histogram[J]. Machine Design and Manufacturing Engineering, 2014, 43(5): 67—70.
- [17] PAN C, CHUANG Y, HSU W H. Ntu TRECVID-2007 Fast Rushes Summarization System[C]// Proceedings of the International Workshop on TRECVID Video Summarization, 2007: 74—78.