

基于 ORB 的镜头边界检测算法

唐剑琪¹, 谢林江², 袁庆生¹, 张冬明³, 包秀国¹, 郭薇⁴

(1. 国家计算机网络应急技术处理协调中心, 北京 100029; 2. 北京邮电大学 信息与通信工程学院, 北京 100876;
3. 中国科学院 计算技术研究所, 北京 100190; 4. 工业和信息化部电信研究院, 北京 100191)

摘要: 场景中摄像机移动、对象运动、光亮突变等现象的存在导致现有镜头边界检测算法的顽健性不足, 为了有效应对这些变化, 提出一种基于 ORB 的镜头边界检测算法, 算法通过计算图像帧之间 ORB 描述子的匹配程度进行镜头边界检测。在检测过程中使用特征点匹配数和匹配率对镜头边界的特点进行描述, 并以相似度曲线描述图像帧之间的关系。实验结果表明, 该算法可以有效解决上述问题造成的镜头误检、漏检, 同时处理速度也获得了明显提升。

关键词: 视频; 镜头边界检测; 局部特征; ORB

中图分类号: TP391.41

文献标识码: A

文章编号: 1000-436X(2013)11-0184-07

Shot boundary detection algorithm based on ORB

TANG Jiang-qi¹, XIE Lin-jiang², YUAN Qing-sheng¹, ZHANG Dong-ming³, BAO Xiu-guo¹, Guo Wei⁴

(1. National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing 100029, China;

2. School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China;

3. Institute Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;

4. China Academy of Telecommunication Research of MIT, Beijing 100191, China)

Abstract: The existing algorithms of SBD show low robustness when there exist camera or object movements, light changes in the scene. An SBD algorithm based on ORB was proposed in order to effectively deal with these changes, in which shot boundary was located by computing the matching degree of ORB descriptor between adjacent images. In the detection process, the matching number and matching rate of the feature points were used to describe the characteristics of shot boundary, and the similarity degree curve was used to represent the relationship between image frames. Experiment results show that the proposed algorithm is effective to solve false and miss detection caused by the above problems, with a sharp rise in procession speed.

Key words: video; shot boundary detection; local features; ORB

1 引言

镜头边界检测(SBD, shot boundary detection)是视频内容分析的重要基础之一, 人们对其早有研究, 当前已得到广泛应用。比如, 在一些基于内容的视频检索分类系统中, 需要通过镜头边界检测方法将视频分割为一些独立的镜头, 再进行高层内容分析、分类、索引或查询。

现有镜头边界检测算法主要包括像素差法, 直方图比较法, 边缘变化率法等。例如, 文献[1]通过

检测颜色直方图等图像底层特征的变化来确定边界位置。文献[2]通过度量运动向量进行镜头边界检测。文献[3]在压缩域检测 DCT 系数, 进行快速边界检测。上述这些方法在遇到场景中存在摄像机移动、对象运动、光亮突变等情况时, 算法性能会快速下降。还有一些学者应用机器学习的方法进行镜头边界检测研究。比如, 文献[4]利用压缩域特征结合神经网络进行边界检测。文献[5, 6]使用像素域特征, 采用有限状态机结合支持向量机进行镜头边界检测。这些基于机器学习的镜头检测算法遇到场景

收稿日期: 2013-03-19; 修回日期: 2013-08-25

基金项目: 国家自然科学基金资助项目(61273247, 61271428)

Foundation Item: The National Natural Science Foundation of China(61273247, 61271428)

中存在光亮变化以及对象运动等情形时顽健性仍然不足。

以上算法主要使用图像的全局特征，而图像局部特征包括角点特征、斑块特征和区域特征等。好的局部特征具有可重复性、独特性、局部性、数量性、准确性、高效性等特点，可以很好地描述图像所蕴含信息。近来，有关研究学者开始引入局部特征进行镜头边界检测研究。文献[7]提出了一种基于 SIFT^[8]特征的镜头边界检测算法，通过匹配相邻图帧进行镜头边界检测。文献[9]使用 CCH(contrast context histogram)^[10]提取图像局部特征，进行镜头边界检测，文献报告称该特征检测子与 SIFT 相比，不但具有更高的精度，而且所耗时间也有所降低。由于该算法需要对视频中所有图像帧进行处理，要降低耗时或者过量的计算，其算法准确率也会降低。文献[11]则提出了一种基于 SURF^[12]特征点匹配的镜头边界检测算法，虽然算法与文献[7,9]相比，检测速度有所提升，但限于 SURF 算子检测能力，检测一些存在对象运动的镜头时，还存在不少误判和漏判。综上，现有算法主要存在两方面问题，一是检测速度较慢，二是当镜头场景中存在摄像机移动、对象运动、光亮突变等内容时算法顽健性和性能下降。

本文提出一种基于 ORB(oriented fast and rotated BRIEF)^[13]的镜头边界检测算法，以应对镜头场景中存在摄像机移动、对象运动、光亮突变等变化情况，算法通过计算图像帧之间 ORB 描述子的匹配程度进行镜头边界检测。

2 算法描述

2.1 ORB 特征

图像全局特征需用使用图像中所有像素运算获得，若当同一镜头中含有部分遮挡、对象运动等变化，或者镜头中出现淡入、淡出、溶解、擦除等渐变情况时，全局特征易导致同一镜头被错误分割为多个短镜头。

相比较而言，局部特征则在图像中存在旋转、尺度、光照等变化时，保持较好的不变性。目前，主要局部特征描述子包括 SIFT、SURF、CCH、BRIEF^[14]、ORB 等。SIFT 对旋转、尺度缩放、亮度变化保持不变性，对视角变化、噪声等也有一定程度的稳定性。SURF 在 SIFT 基础上进行了改进，综合性能优于 SIFT。CCH 由 Huang 等人提出，文献[10]报告称该特征检测子与 SIFT 相比，不但具有

更高的精度，而且所耗时间也有所降低。BRIEF 的优点在于速度，但不具备旋转不变性，对噪声比较敏感。ORB 对 BRIEF 算子进行了改进，Rublee 等人通过实验[13]证明了 ORB 具备旋转不变性，同时噪声顽健性也有良好表现，如图 1 和图 2 所示。

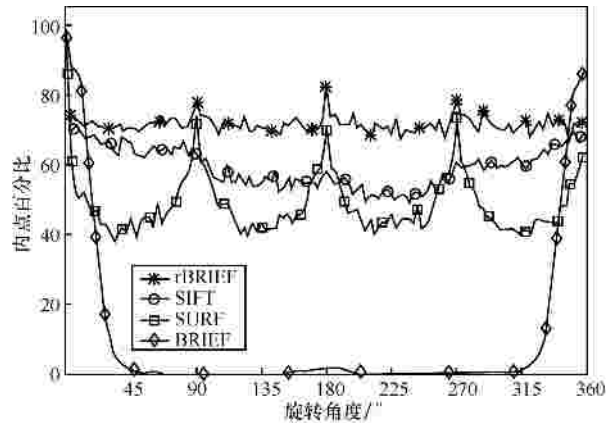


图 1 ORB(rBRIEF)、SIFT、SURF、BRIEF 在图像旋转角度下匹配表现

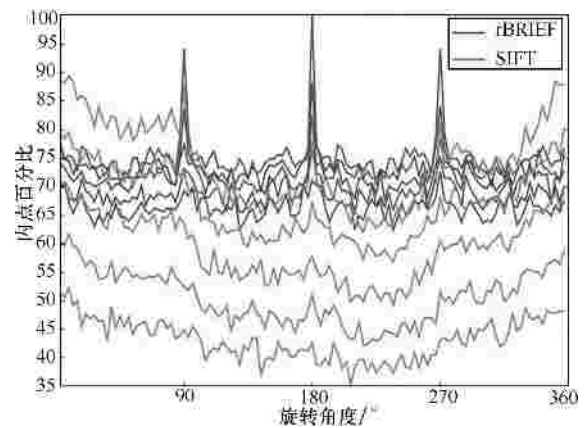


图 2 ORB(rBRIEF)、SIFT 在不同噪声强度下匹配表现

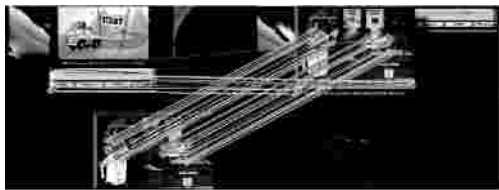
图 1 中，测试集高斯噪声水平为 10，标准 BRIEF 算子在图像旋转到 10° 时，性能就急剧下降；SIFT 表现好于 SURF，后者在图像旋转 45°、135°、225°、315° 时，性能明显下降；ORB(oFAST+rBRIEF)表现最好^[13]。

图 2 中，高斯噪声水平分别为 0、5、10、15、20 和 25。SIFT 受噪声影响，性能下降较快；ORB(rBRIEF)有所下降，但下降幅度较小^[13]。从图 1 与图 2 还可以看出，ORB 在 BRIEF 基础上解决了其不能应对旋转变化的问题，在图像旋转时，检测性能优于 SIFT 以及 SURF。在噪声顽健性方面，ORB 虽然受噪声影响检测性能有所下降，但与 SIFT 相比，下降幅度较小，对噪声的顽健性优于 SIFT 算子。

镜头是视频中基本物理单元和语义单元。一般而言，同一个镜头中的图像帧是比较接近的，特征不会发生剧烈变化。因此，若相邻帧发生了非常明显的特征变化，则可以认为是发生了镜头的转换。局部特征用于镜头边界检测的基本思想在于，如果图像之间特征点匹配数目较多，说明图像相似度高，可能为同一镜头场景，反之，则说明图像相似度低，可能存在镜头边界。为进一步验证 ORB 用于镜头检测的实际性能，选择某个视频中同一镜头中的 2 个不同帧，分别采用 SURF、ORB 以及 CCH 算子对图像进行局部特征，并检测匹配结果，如图 3 所示。



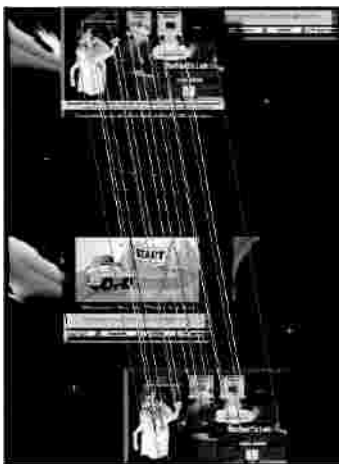
(a) 原始图像对



(b) ORB 算子检测匹配结果



(c) SURF 算子检测匹配结果



(d) CCH 算子检测匹配结果

图 3 ORB、SURF、CCH 检测性能对比

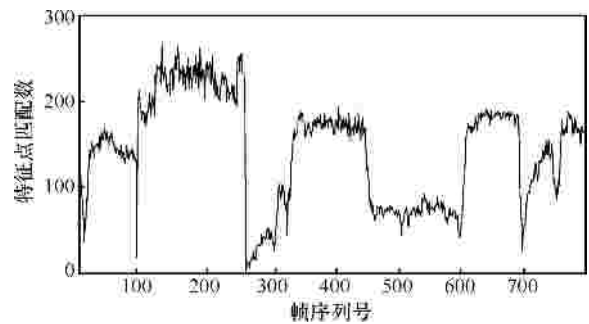
图 3 中，检测前先采用图片测试集将各个算子参数调整到最佳，其中，图 3(d)的检测结果由 CCH 算子检测程序^[15]生成，上述图像的特征点匹配数目分别为 86、78、26。从实际检测结果可以看出，在描述图像局部特征方面，ORB 表现最稳定。

因此，总体而言，ORB 在特征描述方面综合性能最好，本文选择 ORB 算子进行图像局部特征描述。

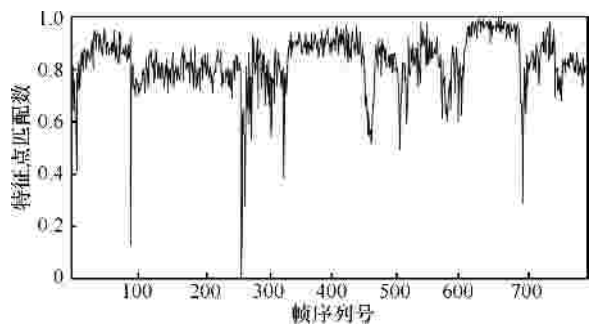
2.2 算法原理及流程

本文算法通过计算图像帧之间 ORB 描述子的匹配程度进行镜头边界检测，下面将从镜头边界特点的描述、算法原理及流程等方面进行阐述。

对某视频片段进行处理，在获得视频片段 ORB 特征点后，对相邻图像帧之间的 ORB 特征点匹配数进行统计分析，来描述镜头边界特点。从图 4(a)中可以看出，不论镜头边界是切变还是渐变形成，在镜头边界附近，特征点匹配数都会快速下降，其最低点是一个很小的数值，但实验结果表明仅使用特征点匹配数来描述镜头边界的特点存在局限性，不能很好地体现出镜头渐变的特点。为此，定义一个新的指标——特征点匹配率，来描述镜头边界的特点，其中特征点匹配率定义为两帧之间特征点匹配数与其中前一帧特征数的比值，如果前一帧没有检测出局部特征，则将该指标置为 0。具体地，该片段特征点匹配率曲线如图 4(b)所示。



(a) 特征匹配数曲线



(b) 特征点匹配率曲线

图 4 视频特征匹配数与匹配率曲线

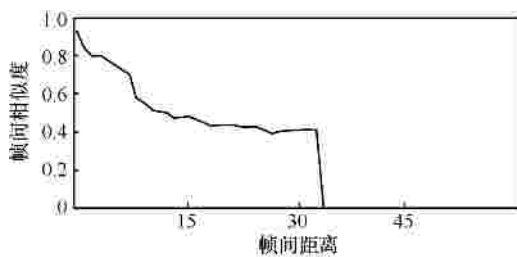
从图 4 可以看出，不论镜头边界是由切变形成，还是由渐变形成，它们在镜头边界处表现出两点特性：1) 该处特征点匹配数和匹配率较低；2) 曲线变化幅度比较大，而在其余地方曲线变化幅度较为平缓。而特征点匹配率曲线可以认为是对匹配数曲线滤波之后的效果，凸显了镜头转换时曲线变化的幅度，弥补了特征点匹配数在描述镜头渐变时的不足。

通过实验知道镜头边界附近的特征点匹配数或匹配率最低点不一定为 0，而是一个相对较小的数值，同时检测算子本身参数的改变将引起图像局部特征描述结果的变化，即视频的局部特征点匹配数以及匹配率曲线的移动。为了有效描述曲线变化的幅度，定义参数 β 描述区间曲线变化的幅度，计算公式为

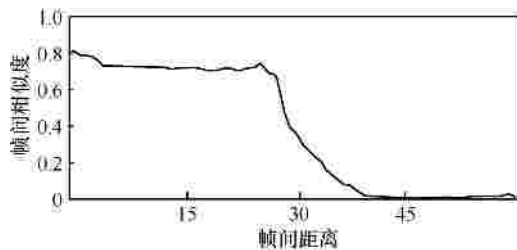
$$b = \frac{f(x_1) - f(x_2)}{f(x_1)} \quad (1)$$

其中， $f(x_1)$ 、 $f(x_2)$ 分别为区间最大值与最小值， $0 < \beta < 1$ 。实验结果表明，参数 β 描述区间曲线变化的幅度是合理的，同时也突现了检测算子本身参数与曲线变化幅度的关系。

而视频片段中两帧间距 N 与其相似度之间也有密切的关系，如图 5 所示。



(a) 切变情形



(b) 渐变情形

图 5 帧间相似度与间距变化关系

图 5 以视频中的一帧图像为参照物，计算远离该帧的图像与该参照帧的相似度，相似度采用特征点匹配率来描述。从图中可以看出，不论切变或者

渐变情形，随着帧间距离 N 的增大，镜头中的两帧相似度逐渐减小。若远离的图像帧跨过了镜头边界，则它们相似度会迅速降低。另外，镜头切换存在一定的间隔，不会在很短时间(比如 1 s 之内)切换多次。可以看出，使用 ORB 特征点匹配率可以很好地描述帧的相似程度，因此本文的算法通过同时度量匹配数和匹配率，共同进行镜头边界检测。简要流程如图 6 所示。其中， Frm 为所获取视频总帧数，匹配对提纯采用 Brute-Force 匹配算法实现， TN 为特征点匹配数阈值， TR 匹配率阈值， β 为曲线变化幅度阈值。

```

提取第一个图像帧 ( $t=0$ ) 的 ORB 描述向量；
For( $t=0; t+N \quad Frm, t=t+N$ )
{
    提取图像帧  $t+N$  的 ORB 描述向量；
    对帧  $t$  与  $t+N$  的 ORB 描述向量进行匹配对提纯；
    计算帧  $t$  与  $t+N$  的 ORB 局部特征点匹配数  $m$  与匹配率  $tr$ ；
    If( $(m < TN \wedge tr < TR)$ )
    {
        计算视频相似度  $R(t', t'+t)$ ,  $t' \in [t, t+N-t]$ ；
        找出该相似度曲线的最大值  $Hi$  与最小值  $Lo$ ；
        若  $(Hi - Lo)/Hi > \beta$ , 满足  $a < Hi, 0 < Lo, b, 0 < \beta < 1$ , 则该区间存在候选边界，且边界中心为  $Lo$  对应的帧序列号；
        若边界中心与相邻边界帧的间隔大于  $d$ , 则记录边界信息。
    }
}

```

图 6 算法流程描述

上述算法中 a 和 b 的主要作用要限制相似度曲线波动的范围，即若该区间相似度的最大值偏小或者最小值偏大时，则认为该区间不存在镜头边界；检测步长 N 不宜设置过大，实验中 N 设置为检测视频的帧率， t 和 d 分别设置为 1 和 15。

3 实验建立及分析

3.1 实验平台及数据

实验程序由 C++和 OpenCv 开源库编程实现，编译环境为 VS2010。测试视频选自 The Open Video Project^[16]，视频中 Volcano Eruptions 有飞机飞行、火山喷发等画面，含有物体运动、镜头伸缩等镜头场景，Measurements Ratios And Graphing 有人物运

动、光亮改变等镜头场景, Estimate And Measure 中有动画特效画面, 含有对象运动、光亮改变等镜头场景, GPS 有摄像机移动和伸缩等镜头场景, Meteors 和 Models 中有物体运动等存在光亮改变或者摄像机移动等镜头场景。所有视频帧率 F 均为 30 frame/s, 视频宽高分别为 352 和 240, 其他相关信息如表 1 所示。

表 1 测试视频信息

编号	视频名称	帧数	边界数
1	Volcano Eruptions	3 332	27
2	Measurements Ratios And Graphing	3 775	17
3	Estimate And Measure	9 300	34
4	GPS	3 857	29
5	Meteors	3 900	24
6	Models	3 893	31

3.2 实验结果和分析

为了验证算法性能, 本文算法与文献[9]和文献[11]中的算法进行了对比实验, 选取它们表现最好的结果作为比较对象, 其中文献[9]算法中特征点检测匹配这部分由文献[15]提供的 CCH Console Mode 实现。本文将实验参数空间设为 $\theta = \{(TN, TR, \beta, a, b)\}$, 实验前采用一组测试视频确定参数的最佳值区间, 针对上述实验数据, 设置参数 $\theta = (40, 0.35, 0.6, 0.7, 0.2)$, 若测试视频中镜头持续时间非常短(不到 0.5 s), 实验时直接排除该情况。

评价镜头边界检测算法的性能, 一般采用召回率 h 和准确率 m 以及它们的乘积 $Q^{[17]}$ 值, 其中, Q 值是一种把召回率和准确率综合考虑的评价方法, 当 Q 值较高时说明实验效果比较理想。

$$h = \frac{hit}{hit + miss} \tag{2}$$

$$m = \frac{hit}{hit + false} \tag{3}$$

$$Q = h \times m \tag{4}$$

其中, hit 为正确检出镜头边界个数, $miss$ 为漏检个数, $false$ 为误检个数。最终, 实验结果如表 2 所示。

从表 2 可以看出, 本文算法的平均召回率介于文献[9]和文献[11]之间, 平均准确率高于文献[9]和文献[11]。从 Q 值来看, 本文算法的平均 Q 值最高, 总体表现优于文献[9]和文献[11], 获得了较好的检测效果。本文还对算法执行的平均速度进行了统计(文献[9]只统计了检测图像局部特征点的耗时), 结果如表 3 所示。

表 2 算法检测结果对比

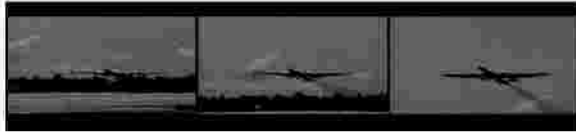
编号	文献[9]的方法		文献[11]的方法		本文算法	
	准确率	召回率	准确率	召回率	准确率	召回率
1	82.76%	88.89%	91.30%	77.78%	100%	85.19%
2	89.47%	100%	81.25%	76.47%	84.21%	94.12%
3	72.73%	94.12%	86.11%	91.18%	91.18%	91.18%
4	96.67%	100%	100%	82.76%	100%	89.66%
5	96%	100%	100%	91.67%	100%	91.67%
6	82.86%	93.55%	87.1%	87.1%	100%	90.32%
平均	85.16%	95.68%	90.79%	85.19%	96.05%	90.12%
Q 值	81.48%		77.34%		86.56%	

表 3 算法检测耗时分析(单位: 秒)

编号	视频时间	文献[9]的方法	文献[11]的方法	本文算法
1	111	1 112.63	176.07	31.42
2	125	1 406.28	168.28	29.06
3	310	3 364.42	417.91	60.03
4	128	1 557.34	215.65	36.53
5	130	1 446.64	166.74	29.28
6	129	1 130.88	215.52	41.96
平均值	155.5	1 669.7	226.7	38.04

从表 3 可以看出, 本文算法平均检测速度比文献[9]和文献[11]分别提高了近 30 倍和 5 倍, 这得益于 ORB 算子快速的检测速度。图 7 展示了算法检测的一些实例。图 7(a)为物体运动场景, 是飞机起飞的画面; 图 7(b)为擦除镜头渐变方式, 视频在该处进行了镜头的转换, 第 2 个镜头的首帧逐渐取代上一个镜头最后帧的像素; 图 7(c)带有特效的渐变

方式,持续过程约为 10 帧;图 7(d)为溶解渐变方式,上一个镜头最后一帧慢慢变暗,第 2 个镜头的首帧慢慢变亮;图 7(e)为摄像机移动镜头场景,摄像机从画面左边转动到画面的右边;图 7(f)为特效镜头场景,画面中物体迅速放大同时有物体(流星)从画面中快速掠过。图 7(g)为误检实例,主要原因是该镜头客观上包含了 2 个场景。



(a) 物体运动场景



(b) 擦除镜头渐变方式



(c) 带有特效的渐变方式



(d) 溶解渐变方式



(e) 摄像机移动镜头场景



(f) 特效镜头场景



(g) 误检实例

图 7 检测实例

4 结束语

本文提出一种基于 ORB 的镜头边界检测算法,算法通过计算图像帧之间 ORB 描述子的匹配程度进行镜头边界检测。实验结果表明,本文算法在一定程度上解决了镜头场景中由于光亮改变、摄像机移动、对象运动造成的误检,对视频中镜头切变以及一些渐变具有良好的顽健性,总体检测效果较好,而且算法计算复杂度低,易于实现,检测速度明显提升,在实际应用方面优势明显。不过,算法的阈值与视频类型或者镜头转换方式有关,需要根据具体数据学习调整,下一步将对算法阈值的自适应性进行研究。

参考文献:

- [1] GARGI U, KASTURI R, STRAYER S H. Performance characterization of video-shot-change detection methods[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2000, 10(1):1-13.
- [2] AMEL A M, ABDESSALEM B A, ABDELLATIF M. Video shot boundary detection using motion activity descriptor[J]. Journal of Telecommunications, 2010, 2(1):54-59.
- [3] SHEN K, DELP E J. A fast algorithm for video parsing using MPEG compressed sequences[A]. Proc International Conference on Image Processing(ICIP)[C]. Washington, DC, USA, 1995, 2:252-255.
- [4] LEE M H, YOO H W, JANG D S. Video scene change detection using neural network: improved ART2[J]. Expert Systems with Applications, 2006, 31(1):13-25.
- [5] LIU Z, GIBBON D, ZAVESKY E, et al. AT&T research at trecvid 2006[EB/OL]. <http://www-nlpir.nist.gov/projects/tvpubs/tv6.papers/att.pdf>, 2006.
- [6] LIU Z, ZAVESKY E, GIBBON D, et al. AT&T research at trecvid 2007[EB/OL]. <http://www-nlpir.nist.gov/projects/tvpubs/tv7.papers/att.pdf>, 2007.
- [7] PARK M H, PARK R H, LEE S W. Shot boundary detection using scale invariant feature matching[A]. Proc SPIE Visual Communications and Image Processing(VICIP)[C], 2006, 6077:569-577.
- [8] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.
- [9] HUANG C R, LEE H P, CHEN C S. Shot change detection via local keypoint matching[J]. IEEE Transactions on Multimedia, 2008, 10(6):1097-1107.
- [10] HUANG C R, CHEN C S, CHUNG P C. Contrast context histogram—an efficient discriminating local descriptor for object recognition and image matching[J]. Pattern Recognition, 2008, 41:3071-3077.
- [11] BIRINCI M, KIRANYAZ S, GABBOUJ M. Video shot boundary detection by structural analysis of local image features[A]. 12th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)[C]. Delft, Netherlands, 2011.

[12] BAY H, ESS A, TUYTELAARS T, *et al.* SURF: speeded up robust features[J]. *Computer Vision and Image Understanding (CVIU)*, 2008, 110(3):346-359.

[13] RUBLEE E, RABAUDE V, KONOLIGE K, *et al.* ORB: an efficient alternative to SIFT or SURF[A]. *IEEE International Conference on Computer Vision (ICCV)[C]*. Barcelona, Spain, 2011. 2564-2571.

[14] CALONDER M, LEPETIT V, STRECHA C, *et al.* Brief: binary robust independent elementary features[A]. *Computer Vision-ECCV[C]*. 2010. 778-792.

[15] CCH[EB/OL]. <http://imp.iis.sinica.edu.tw>.2013.

[16] The open video project[EB/OL]. <http://www.open-video.org>.2013.

[17] JOYCE R A, LIU B. Temporal segmentation of video using frame and histogram space[J]. *IEEE Transactions on Multimedia*, 2006, 8(1): 130-140.



袁庆生 (1980-), 男, 山东济南人, 国家计算机网络应急技术处理协调中心工程师, 主要研究方向为网络与信息安全。



张冬明[通信作者] (1977-), 男, 江苏盐城人, 中国科学院计算技术研究所副研究员, 主要研究方向为多媒体内容检索、模式识别、视频编码等。E-mail:dmzhang@ict.ac.cn。

作者简介:



唐剑琪 (1974-), 女, 黑龙江依兰人, 国家计算机网络应急技术处理协调中心高级工程师, 主要研究方向为网络与信息安全与图像处理。



包秀国 (1963-), 男, 江苏如皋人, 国家计算机网络应急技术处理协调中心教授级高级工程师、博士生导师, 主要研究方向为网络与信息安全。



谢林江 (1987-), 男, 重庆人, 北京邮电大学硕士生, 主要研究方向为多媒体内容分析。



郭薇 (1981-), 女, 山东泰安人, 工业和信息化部电信研究院工程师, 主要研究方向为网络与信息安全管理。