

结合 SIFT 算法的视频场景突变检测

李 枫, 赵 岩*, 王世刚, 陈贺新
(吉林大学 通信工程学院, 吉林 长春 130012)

摘要: 视频场景变化检测对于视频的标注以及语义检索具有非常重要的作用。本文提出了一种结合 SIFT(Scale Invariant Feature Transformation) 特征点提取的场景变化检测算法。首先利用 SIFT 算法分别提取出视频前后帧的特征点并分别统计其数量, 然后对视频前后帧进行图像匹配, 统计匹配上的特征点数量, 最后将该帧的匹配特征点数量与该帧前一帧的特征点数量做比值, 从而通过该比值判断场景变化情况。实验结果表明, 视频场景突变检测率平均可以达到 95.79%。本算法可以在视频帧进行图像匹配的过程中对场景的变化情况进行判断, 因此该算法不仅应用范围较广, 还可以保证场景变化检测的精度, 仿真结果证明了算法的有效性。

关键词: SIFT; 特征点匹配; 场景变化检测

中图分类号: TN919.8 文献标识码: A doi:10.3788/CO.20160901.0074

Video scene mutation change detection combined with SIFT algorithm

LI Feng, ZHAO Yan*, WANG Shi-gang, CHEN He-xin
(College of Communication Engineering, Jilin University, Changchun 130012, China)

* Corresponding author, E-mail: zhao_y@jlu.edu.cn

Abstract: Video scene change detection has a very important role for video annotation and semantic search. This paper proposes a scene mutation change detection algorithm combined with SIFT(Scale Invariant Feature Transformation) feature point extraction. Firstly, the feature points of two adjacent video frames are extracted respectively using SIFT algorithm and the number of them is counted respectively. Then image matching of the two adjacent frames of the video is performed and the number of matching feature points is counted. Finally, the ratio between the number of matching feature points of the current frame and the number of matching feature points of its previous frame is calculated, so as to judge the scene change by this ratio. The average scene mutation change detection rate in the experimental results can reach 95.79%. The proposed algorithm can judge scene change during image matching. Therefore, the algorithm can not only be applied widely, but also guarantee the accuracy of scene change detection. Experimental results show the effectiveness of the proposed algorithm.

Key words: SIFT; feature point matching; scene change detection

1 引言

图像和视频作为信息的载体之一,已经融入到了人们的生活之中。场景变化检测算法在关键帧检索、视频检索、视频编码以及码率控制等方面具有广泛的应用,因此研究场景变化检测算法对视频技术的进一步发展有很大帮助。

目前常见的场景突变检测算法有像素比较法、直方图比较法、基于边缘的方法和基于运动矢量的方法等^[1-5]。像素比较法比较简单,但是对物体运动和噪声比较敏感;直方图比较法忽略了图像中的位置信息;基于边缘的方法和基于运动矢量的方法能够获得较好的检测效果,但是计算量较大。随着研究的深入,一些算法有了更加精准的检测结果。例如, Vlachos^[6]曾经提出利用相位相关性进行场景突变检测的方法。 Yumi Eom^[7]等人提出了在 HEVC 的编码码流中利用帧特征对超高清(UHD)视频进行场景变化检测,解决了传统场景变化检测方法无法面对 UHD 视频数据量大的问题。 Soongi Hong^[8]等人为解决传统算法在面对不同视频时阈值选择单一的问题上提出了先提取视频直方图,而后通过对数正态分布和参数模型估计的方法,实现了自适应阈值的视频场景变化检测。不过一些算法因为计算量过大,同时没有明显提高检测性能而没有得到普及和应用。例如,文献[9]利用压缩域结合神经网络进行边界检测,当场景中出现光强变化以及镜头移动等情况时,这种基于神经网络的场景变化检测算法顽健性仍然不足。文献[10]提出一种基于 ORB(面向快速和短暂旋转)的场景变化检测算法,不过,该算法中的阈值与视频类型或者镜头转换方式有关,需要在实际应用中根据具体数据进行调整。

针对以上问题,如何在视频中高效准确的判断场景变化成为本文研究的重点内容。故本文提出了一种结合 SIFT 特征点匹配算法的场景突变检测算法。

2 本文算法

SIFT 算法自 David Lowe 于 1999 年提出后,由于其在物体识别、图像拼接、机器人地图感知与导航、影像追踪等领域的广泛应用而备受人们的关注^[11-16]。SIFT 算法具有尺度不变性、旋转不变性,同时对于仿射变化,噪音以及光照变化具有很好的稳定性,因此经过了十几年的发展依然是一种经典的特征提取算法。

SIFT 算法主要有以下几个特点:(1)独特性,信息量丰富,可以在特征数据库中进行快速、准确的匹配;(2)多量性,即使少数的几个对象也可以提取数量较多的 SIFT 特征向量;(3)高速性,经典的 SIFT 匹配算法经过降低维度、改变灰度级等方法处理之后甚至可以达到实时的要求;(4)可扩展性,可以很方便地与其他形式的特征向量进行联合。因此本文将作为图像匹配和识别的基础,根据 SIFT 算法的这些特点进行视频场景突变检测。

本文算法流程图见图 1。基于 SIFT 特征点提取的场景突变检测算法主要分为 3 个部分:

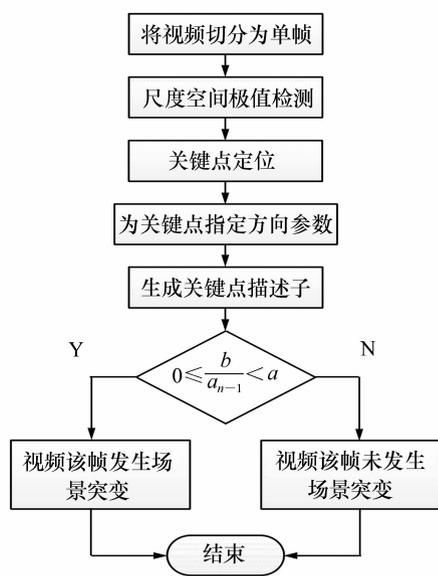


图1 本文算法流程图

Fig. 1 Flow chart of the algorithm proposed in this paper

(1) SIFT 算法提取特征点; (2) 基于 SIFT 算法的图像匹配; (3) 视频场景变化检测。

2.1 SIFT 算法提取特征点

SIFT 算法的本质是在不同的尺度空间上寻找特征点并计算出特征点的方向。Lowe 将 SIFT 算法主要分为 4 个步骤:

(1) 尺度空间极值检测: 首先对图像实现高斯模糊并通过降采样建立高斯金字塔, 利用高斯金字塔表示尺度空间, 然后每一个像素点要和它所有的相邻点比较, 看其是否比它的图像域和尺度域的相邻点大或者小。以确保在尺度空间和二维图像空间都检测到极值点。

(2) 关键点的位置确定: 尺度空间极值检测得到的是离散空间的关键点, 为了得到稳健的特征点, 必须对尺度空间高斯差分 (DOG) 函数进行拟合, 同时去除由于 DOG 算子的边缘响应而产生的边缘点。

(3) 关键点的方向确定: 计算关键点的图像梯度, 并用直方图统计邻域内像素的梯度和方向, 梯度直方图将 $0 \sim 360^\circ$ 的方向范围分为 36 个柱, 其中每柱 10° 。直方图的峰值方向代表了关键点的主方向。

(4) 关键点描述: 对于每一个关键点, 拥有 3 个信息: 位置、尺度以及方向。特征点描述的目的是用一个具有独特性的特征向量把特征点描述出来。Lowe 建议描述子使用在关键点尺度空间内 4×4 的窗口中计算的 8 个方向的梯度信息, 共 $4 \times 4 \times 8 = 128$ 维向量表征。

2.2 基于 SIFT 算法的图像匹配

在 SIFT 算法的实际应用过程中, SIFT 特征点的匹配是最重要的步骤之一。SIFT 特征点使用在关键点尺度空间内 4×4 的窗口中计算的 8 个方向的梯度信息, 共 $4 \times 4 \times 8 = 128$ 维向量表征, 因此 SIFT 特征点的匹配问题就可以转化为在 128 维空间中特征向量的最近邻搜索问题。由于特征描述子是一个 128 维向量, 因此任意向量的距离定义都可以作为特征点的距离, 比较常用的有 Euclid 距离和协方差和反余弦函数计算出的距离。由于 Euclid 距离计算量较大, 会影响场景变化检测的计算速度, 因此本文采用协方差和反余弦函数计算, 如式(1):

$$d(u, v) = \arccos\left(\sum_{i=1}^{128} u_i \times v_i\right). \quad (1)$$

计算输入图像的每个特征点与匹配图像中所有特征点的距离后将其按照从小到大的顺序排列得到距离向量 d 。计算最近距离 $d(1)$ 与次近距离 $d(2)$ 的比值, 即 $d(1)/d(2)$, 如果该比值小于阈值, 则认为匹配, 否则认为不匹配。

2.3 视频场景变化检测

SIFT 特征匹配具有显著性, 也就是说特征点在其匹配的图像中进行 SIFT 特征点搜索时, 正确匹配的概率非常高, 故本文将其作为图像匹配和识别的基础, 利用 SIFT 匹配算法的这个特点进行视频的场景突变检测。

将图像中相邻两帧即当前帧与前一帧利用 SIFT 算法进行特征点匹配, 两帧匹配后得到的特征点数量记为 b 。为了通过相邻两帧的匹配率来反映视频场景突变的情况, 将特征点匹配数量 b 与前一帧的特征点数量 a_{n-1} 作比值, 当其满足式(2)时, 认定当前帧发生了场景突变, 经实验验证 α 取 0.3。

$$0 \leq \frac{b}{a_{n-1}} \leq \alpha, \quad (2)$$

式中, n 表示当前帧, $n-1$ 表示前一帧, a_{n-1} 表示前一帧特征点数量, b 表示当前帧与前一帧利用 SIFT 算法进行特征点匹配后得到的匹配特征点数量。

3 实验结果

3.1 算法验证

本程序是在 VS2013 + OpenCV2.4.10 开发



图2 无场景突变举例(115、116帧)

Fig. 2 Examples of no scene mutation change (the 115th and 116th frames)



图3 场景突变举例(114、115 帧)

Fig. 3 Examples of scene mutation change (the 114th and 115th frames)

环境下实现,主要包括 SIFT 特征点提取、特征点匹配与场景变化检测的判断三个部分,实验过程结果见图 2 与图 3。

本文截取了实验视频中 114、115、116 三帧,其中 114 与 115 帧发生了场景突变,115 与 116 帧无场景突变,具体实验数据见表 1。通过实验可以看出当相邻两帧没有发生场景突变的时候,其相邻两帧的相似性高,匹配率大,如图 2 所示;当相邻两帧发生了场景突变的时候,由于相邻两帧的场景完全变化,基本没有相似性,匹配率也随之大大降低,如图 3 所示。因此可通过 SIFT 特征点匹配算法实现场景突变检测。

表 1 实验数据举例

Tab. 1 Examples of experimental data

视频帧	特征点数量	与前一帧匹配的特征点数量	与前一帧特征点匹配率/%	是否发生场景突变
114	276	238	85.9	否
115	299	26	9.42	是
116	298	199	66.56	否

3.2 SIFT 特征点提取实时性验证

由于视频的场景突变检测算法在实际的应用当中实时性指标较为重要,因此本文通过视频片

段验证 SIFT 算法在特征点提取方面是否可以达到实时性的要求。

表 2 SIFT 特征点提取耗时统计

Tab. 2 Time consumption of SIFT feature point extraction

视频片段	总帧数	SIFT 特征点提取所用时间/s	平均 1 s 完成特征点提取的视频帧数/s
动画(哆啦 A 梦)	1 922	72.717 6	26.431 0

从表 2 可以看出,对于场景较复杂的动画,其 1 922 帧的特征点提取时间为 72.717 6 s,平均 1 s 可以提取 26.431 0 frames。人眼能识别的最低帧数是 24 frames/s,就是说画面每秒钟变换 24 次以上人眼就会觉得连贯,低于 24 frames/s 就会觉得有停顿感。常见的电视等播放器可以达到 25 frames/s,本实验中可以达到 26 frames/s 左右,因此基本可以达到 SIFT 特征点实时提取的要求。

3.3 视频场景突变检测算法实验数据

为了验证本文算法的有效性,本文挑选使用了不同视频片段作为实验数据,包括新闻、电影与动画等,均为 AVI 格式视频。实验数据中场景突

变点信息如表 3 所示。

表 3 视频中场景突变点

Tab. 3 Scene change points of videos in our experiments

视频片段	总帧数	场景突变点数目
电影	3 479	117
动画	1 922	100
新闻	2 180	39
MV	3 022	147

评价检测算法最常用的是查全率和查准率,标准检测算法的有效性指标定义如下:

$$P_{re} = N_c / (N_c + N_m), \quad (3)$$

$$P_{pre} = N_c / (N_c + N_f), \quad (4)$$

式中, N_m, N_f, N_c 分别为漏检数、误检数、正确检测数。在场景突变检测中, 查全率用来证明这个场景变化检测算法在实际应用中的有效性, 虚检率用来说明当达到一定检全率时所付出的代价。

表4 本算法实验结果

Tab.4 Results obtained using the algorithm proposed in this paper

视频片段	N_c	N_m	N_f	$P_{re}/\%$	$P_{pre}/\%$
电影	112	5	10	95.73	91.80
动画	98	2	7	98.00	93.33
新闻	37	2	0	94.87	100.00
MV	138	8	11	94.56	92.67

3.4 本算法与其他算法比较

文献[17]中提出针对二维视频编码的运动帧差图像的多测度联合场景突变检测算法, 对视频 ishop, 本文与文献[17]比较结果见表5。视频 ishop 是一段场景变化较多的 MV 视频, 总共有 2 752 帧, 其中场景突变总数为 53 个。从表 5 可以看出, 本文算法的查全率比文献[17]高 3.78%, 查准率比文献[17]高 4.88%。

表5 视频 ishop 的比较结果

Tab.5 Comparison results of ishop video

方法	N_c	N_m	N_f	$P_{re}/\%$	$P_{pre}/\%$
本文	51	2	7	96.23	87.93
文献[17]	49	4	10	92.45	83.05

文献[18]在信息论的基础上, 提出了一种结合信息论和运动信息的镜头检测算法, 通过视频探索《北极熊》与文献[18]比较结果见表6所示, 视频探索《北极熊》是一段纪录片, 总共有 39 687 frames, 其中场景突变共有 127 个。本文算法的查全率比其高 1.58%, 查准率比其高 0.83%。由于此视频段中北极场景较多, 白雪皑

视频场景突变检测的实验结果如表4所示。实验结果中平均查全率为 95.79%, 平均查准率为 94.45%。查准率略低于查全率, 这是由于视频中的人工编辑多而复杂, 场景内出现大范围的突然的运动变化, 从而容易造成的误判。

皑的背景下特征点提取受到一定的限制, 故本文算法与文献[18]比较后提高较少。

表6 视频探索《北极熊》的比较结果

Tab.6 Comparison results of 'Exploring Arctic Bear'

方法	N_c	N_m	N_f	$P_{re}/\%$	$P_{pre}/\%$
本文	121	6	5	95.28	96.03
文献[18]	119	8	6	93.70	95.20

4 结 论

本文提出了一种基于 SIFT 特征点匹配的场景突变检测算法。通过 SIFT 特征点提取与特征点匹配算法相结合, 实现了视频的场景突变检测。本算法可以在视频帧进行图像匹配的过程中对场景的变化情况进行判断, 同时 SIFT 特征点提取可以满足实时性的要求, 因此该算法不仅应用范围较广, 还可以保证场景突变检测的精度, 视频场景突变检测平均查全率为 95.79%, 平均查准率为 94.45%, 表明了该算法在视频场景突变检测中的有效性。

参考文献:

- [1] 邹晓燕. 基于 H.264 压缩域的视频检索研究[D]. 山东: 山东大学, 2011.
ZOU X Y. Video retrieval based H.264 compressed domain[D]. Shandong: Shandong University, 2011. (in Chinese)
- [2] 薛立勤, 张秀娟. 基于运动分析的视频检索方法[J]. 计算机工程与应用, 2008, 44(25): 152-154.

- XUE L Q,ZHANG X J. Video retrieval method based on motion analysis[J]. *Computer Engineering and Applications*, 2008,44(25):152-154. (in Chinese)
- [3] 朱耀麟,李倩. 视频检索常用的镜头分割方法的研究[J]. 电视技术,2014,38(3):178-181.
ZHU Y L,LI Q. Survey of used methods for partitioning video into shots in video indexing[J]. *Ideo Engineering*,2014,38(3):178-181. (in Chinese)
- [4] 魏玮,刘静,王丹丹. 视频镜头分割方法综述[J]. 计算机系统应用,2013,22(1):5-8.
WEI W,LIU J,WANG D D. Survey of methods for partitioning video into shots in video[J]. *Computer System & Applications*,2013,22(1):5-8. (in Chinese)
- [5] 刘艳红. 视频镜头分割算法综述[J]. 科技创新与应用,2014(16):49-50.
LIU Y H. Survey of methods for partitioning video into shots in video[J]. *Scientific and Technological Innovation and Application*,2014(16):49-50. (in Chinese)
- [6] VLACHOS T. Cut detection in video sequences using phase correlation[J]. *IEEE Signal Processing Letters*,2000,7(7):173-175.
- [7] EOM Y,PARK S,YOO S,*et al.*. *An analysis of Scene Change Detection*[M]. Anaheim,California,USA: International Conference on Semantic Computing,2015:470-474.
- [8] HONG S,CHO B,CHOE Y. *Adaptive Thresholding for Scene Change Detection*[M]. Berlin:International Conference on Consumer Electronics,2013:75-78.
- [9] 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.
- [10] 唐剑琪,谢林江,袁庆生,等. 基于 ORB 的镜头边界检测算法[J]. 通信学报,2013,34(11):184-190.
TANG J Q,XIE L J,YUAN Q SH,*et al.*. ORB-based shot boundary detection algorithm[J]. *Communications*,2013,34(11):184-190. (in Chinese)
- [11] 吴伟交. 基于 SIFT 特征点的图像匹配算法[D]. 武汉:华中科技大学,2013.
WU W J. Image matching algorithm based on SIFT feature points[D]. Wuhan:Huazhong University of Science and Technology,2013. (in Chinese)
- [12] 许佳佳. 结合 Harris 与 SIFT 算子的图像快速配准算法[J]. 中国光学,2015,8(4):574-581.
XU J J. Fast image registration method based on Harris and SIFT algorithm[J]. *Chinese Optics*,2015,8(4):574-581. (in Chinese)
- [13] 高文,朱明,贺柏根,等. 目标跟踪技术综述[J]. 中国光学,2014,7(3):365-375.
GAO W,ZHU M,HE B G,*et al.*. Overview of target tracking technology[J]. *Chinese Optics*,2014,7(3):365-375. (in Chinese)
- [14] 聂海涛,龙科慧,马军,等. 采用改进尺度不变特征变换在多变背景下实现快速目标识别[J]. 光学精密工程,2015,23(8):2349-2356.
NIE H T,LONG K H,MA J,*et al.*. Fast object recognition under multiple varying background using improved SIFT method[J]. *Opt. Precision Eng.*,2015,23(8):2349-2356. (in Chinese)
- [15] 王睿,朱正丹. 融合全局-颜色信息的尺度不变特征变换[J]. 光学精密工程,2015,23(1):295-301.
WANG R,ZHU ZH D. SIFT matching with color invariant characteristics and global context[J]. *Opt. Precision Eng.*, 2015,23(1):295-301. (in Chinese)
- [16] 刘志文,刘定生,刘鹏. 应用尺度不变特征变换的多源遥感影像特征点匹配[J]. 光学精密工程,2013,21(8):2146-2153.
LIU ZH W,LIU D SH,LIU P. SIFT feature matching algorithm of multi-source remote image[J]. *Opt. Precision Eng.*, 2013,21(8):2146-2153. (in Chinese)
- [17] 马彦卓,常义林,杨海涛. 应用于视频编码的实时多测度联合突变场景切换检测算法[J]. 光子学报,2010,39(6):1058-1063.
MA Y ZH,CHANG Y L,YANG H T. Used in video coding mutations in real-time multi measure joint scene change de-

tection algorithm[J]. *Acta Photonica Sinica*, 2010, 39(6):1058-1063. (in Chinese)

[18] 易璨. 基于信息熵和运动信息的视频镜头检测[D]. 湘潭:湘潭大学, 2006.

YI C. Shot detection based on information entropy and motion information[D]. Xiangtan: Xiangtan University, 2006. (in Chinese)

作者简介:



李 枫(1990—),女,黑龙江佳木斯人,硕士研究生,主要从事场景变化检测方面的研究。E-mail: 1006407131@qq.com



赵 岩(1971—),女,吉林辽源人,教授,博士生导师,2003年于吉林大学获得博士学位,主要从事图像与视频编码、立体视频处理方面的研究。E-mail: zhao_y@jlu.edu.cn

《发光学报》

—EI 核心期刊(物理学类;无线电电子学、电信技术类)

《发光学报》是中国物理学会发光分会与中国科学院长春光学精密机械与物理研究所共同主办的中国物理学会发光分会的学术会刊。该刊是以发光学、凝聚态物质中的激发过程为专业方向的综合性学术刊物。

《发光学报》于1980年创刊,曾于1992年,1996年,2000年和2004年连续四次被《中文核心期刊要目总览》评为“物理学类核心期刊”,并于2000年同时被评为“无线电电子学、电信技术类核心期刊”。2000年获中国科学院优秀期刊二等奖。现已被《中国学术期刊(光盘版)》、《中国期刊网》和“万方数据资源系统”等列为源期刊。英国《科学文摘》(SA)自1999年;美国《化学文摘》(CA)和俄罗斯《文摘杂志》(AJ)自2000年;美国《剑桥科学文摘社网站》自2002年;日本《科技文献速报》(CBST, JICST)自2003年已定期收录检索该刊论文;2008年被荷兰“Elsevier Bibliographic Databases”确定为源期刊;2010年被美国“EI”确定为源期刊。2001年在国家科技部组织的“中国期刊方阵”的评定中,《发光学报》被评为“双效期刊”。2002年获中国科学院2001~2002年度科学出版基金“择重”资助。2004年被选入《中国知识资源总库·中国科技精品库》。本刊内容丰富、信息量大,主要反映本学科专业领域的科研和技术成就,及时报道国内外的学术动态,开展学术讨论和交流,为提高我国该学科的学术水平服务。

《发光学报》自2011年改为月刊,A4开本,144页,国内外公开发行。国内定价:40元,全年480元,全国各地邮局均可订阅。《发光学报》欢迎广大作者、读者广为利用,踊跃投稿。

地 址:长春市东南湖大路3888号

《发光学报》编辑部

邮 编:130033

电 话:(0431)86176862, 84613407

E-mail: fgxbt@126.com

国内统一刊号:CN 22-1116/04

国际标准刊号:ISSN 1000-7032

国内邮发代号:12-312

国外发行代号:4863BM

http://www.fgxb.org