

Journal of Image
and Graphics

中国图象图形学报



ISSN1006-8961
CN11-3758/TB

2013 5
Vol.18 No.

中国科学院遥感应用研究所
中国图象图形学学会主办
北京应用物理与计算数学研究所

中国图象图形学报

Zhongguo Tuxiang Tuxing Xuebao

2013 年 5 月 第 18 卷 第 5 期(总第 205 期)

目 次

综述

- 中国图像工程:2012 章毓晋(483)
动态场景下的交通标识检测与识别研究进展 刘华平, 李建民, 胡晓林, 孙富春(493)

图像处理与编码

- 基于 PCA 硬阈值收缩的平滑投影 Landweber 图像压缩感知重构 李然, 干宗良, 朱秀昌(504)

图像分析和识别

- 利用视觉显著性和粒子滤波的运动目标跟踪 张巧荣, 冯新扬(515)
融合边缘测度的自然场景彩色图像区域分割 代沁伶, 王雷光, 洪亮(523)
场景语义树图像标注方法 刘咏梅, 杨帆, 于林森(529)
基于视锥细胞色适应生理机制的颜色恒常性计算 袁兴生, 王正志, 项凤涛(537)

医学图像处理

- 改进 FCM 和 LFP 相结合的白细胞图像分类 庞春颖, 刘记奎, 韩立喜(545)
局部高斯分布拟合的脑 MR 图像分割及有偏场校正 崔文超, 王毅, 樊养余, 冯燕, 郝重阳(552)

遥感图像处理

- 局部投影可分离的高光谱图像异常检测 刘明, 杜小平, 夏鲁瑞(558)
基于结构特征的遥感影像匹配 寇媛, 徐景中(565)

“第 9 届中国计算机图形学大会”专栏

- 结合浅景深与构图的图像质量评价 顾婷婷, 郭延文, 殷昆燕(574)
结合精确大气散射图计算的图像快速去雾 甘佳佳, 肖春霞(583)
多特征综合的视频拷贝检测 林莹, 杨扬, 凌康, 肖金伟, 武港山(591)
真实感雨线绘制 江隆盛, 吴恩华(600)
误差云映射的深度融合 3 维重建 陈亚当, 蔡仲谋, 吴恩华(607)

- “第八届中国系统建模与仿真技术高层论坛”征文通知 封 3

Journal of Image and Graphics

(Monthly, Started in 1996)

Vol. 18 No. 5 May 2013

Contents

Review

- Image engineering in China: 2012 Zhang Yujin(483)
- Recent progress in detection and recognition of the traffic signs in dynamic scenes
..... Liu Huaping, Li Jianmin, Hu Xiaolin, Sun Fuchun(493)

Image Processing and Coding

- Smoothed projected Landweber image compressed sensing reconstruction using hard thresholding based on principal components analysis
..... Li Ran, Gan Zongliang, Zhu Xiuchang(504)

Image Analysis and Recognition

- Object tracking based on visual saliency and particle filter Zhang Qiaorong, Feng Xinyang(515)
- Regional segmentation of natural scene color images by integrating edge cues Dai Qinling, Wang Leiguang, Hong Liang(523)
- Automatic image annotation based on scene semantic trees Liu Yongmei, Yang Fan, Yu Linsen(529)
- Color constancy algorithm based on biological mechanism; color adaptation of cone cells
..... Yuan Xingsheng, Wang Zhengzhi, Xiang Fengtao(537)

Medical Image Processing

- White blood cells image classification based on improving the connection of FCM and LFP
..... Pang Chunying, Liu Jikui, Han Lixi(545)
- Multiphase level set method for segmentation and bias correction of brain MR images
..... Cui Wenchao, Wang Yi, Fan Yangyu, Feng Yan, Hao Chongyang(552)

Remote Sensing Image Processing

- Anomaly detection method based on separable local projection in hyperspectral imagery ... Liu Ming, Du Xiaoping, Xia Lurui(558)
- Algorithm for remote sensing images matching based on the structure characteristics Kou Yuan, Xu Jingzhong(565)

Issum of the Chinagraph '2012

- Image quality assessment combining low DoF and composition Gu Tingting, Guo Yanwen, Yin Kunyan(574)
- Fast image dehazing based on accurate scattering map Gan Jiajia, Xiao Chunxia(583)
- Video copy detection based on multiple visual features synthesizing
..... Lin Ying, Yang Yang, Ling Kang, Xiao Jinwei, Wu Gangshan(591)
- Photorealistic rendering of rain streaks Jiang Longsheng, Wu Enhua(600)
- Three dimensional reconstructions based on error clouds mapping Chen Yadang, Cai Zhongmou, Wu Enhua(607)

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2013)05-0591-09

论文引用格式: 林莹, 杨扬, 凌康, 肖金伟, 武港山. 多特征综合的视频拷贝检测[J]. 中国图象图形学报, 2013, 18(5): 591-599.

多特征综合的视频拷贝检测

林莹, 杨扬, 凌康, 肖金伟, 武港山

南京大学计算机软件新技术国家重点实验室, 南京 210046;

南京大学计算机科学与技术系, 南京 210046

摘要: 基于内容的视频拷贝检测是多媒体领域的一个研究热点。由于拷贝变换的多样性和综合性, 单一特征难以获得很好的检测效果。提出一种多特征综合的方法来提高视频拷贝检测的效果。除了使用传统的局部和全局视觉特征外, 还使用非正交二值子空间(NBS)方法来表示视频内容, 并在其基础上使用归一化互相关(NCC)来提高拷贝视频内容相似度计算的效果。在此基础上, 还采用多种措施对拷贝视频的判定结果进行精化。实验结果表明, 该套方案对多种拷贝变换具有很强的鲁棒性, 并且能够得到很好的检测精度。

关键词: 多特征综合; 视频拷贝检测; 非正交二值子空间(NBS); 归一化互相关(NCC)

Video copy detection based on multiple visual features synthesizing

Lin Ying, Yang Yang, Ling Kang, Xiao Jinwei, Wu Gangshan

State Key Laboratory for Novel Software Technology at Nanjing University, Nanjing 210046, China;

Department of Computer Science and Technology, Nanjing University, Nanjing 210046, China

Abstract: Nowadays, content based video copy detection has become a widely studied issue. Due to the uncertainty and diversity of video copy transformation, it is difficult to achieve great performance based on single visual features in video copy detection. In this paper, we propose a new method, which uses a multiple visual feature synthesizing method to solve the problem. Besides traditional local and global visual features, we additionally employ nonorthogonal binary subspace (NBS) as special visual feature to represent the video content. On this basis, the normalized cross correlation (NCC) is used to improve the performance of similarity matching of the video content. We also used other measures to improve the detection accuracy. The experiment results show that our system is robust to various video transformations, and achieves better detection accuracy.

Key words: multiple visual feature synthesizing; video copy detection; nonorthogonal binary subspace (NBS); normalized cross correlation (NCC)

0 引言

伴随着信息化时代的来临, 计算机、网络和多媒体技术得到了前所未有的普及与发展, 数字视频数据成倍增长。这些视频常常会在没有任何监督的情

况下被任意编辑、复制和转载, 成为很多不同的拷贝版本。如何在海量视频数据中检测拷贝视频已经成为一个亟待解决的问题。视频拷贝检测还是很多应用的核心问题, 如版权执法、改进网页检索、广告监控、基于实例的视频检索、减少冗余和概念跟踪等^[1]。近来, 研究者们越来越重视基于内容的视频

收稿日期: 2012-08-28; 修回日期: 2013-01-09

基金项目: 国家自然科学基金项目(61021062); 国家高技术研究发展计划(863)基金项目(2011AA01A202)

第一作者简介: 林莹(1986—), 女, 2013年于南京大学获工学硕士学位, 主要研究领域为多媒体检索。E-mail: keller0618@163.com

通讯作者: 武港山, E-mail: gswu@nju.edu.cn

拷贝检测的研究,但是自动检测技术效果一直不佳,如何提高自动检测的精度成为一个挑战。

根据 TRECVID 官方网站的描述,视频拷贝可以归结为一个视频通过不同的变换方法,如添加、裁剪、修改(外观、色彩、对比度、编码)、嵌入等,转换成另一个视频^[1]。由于拷贝变换的方法种类繁多,甚至是多种变换方法的叠加,这给检测处理带来了很大的难度。由此可知,视频拷贝检测技术的一个核心问题就是怎样提高拷贝检测技术对多种多样的拷贝变换的鲁棒性,其研究重点在于寻找有效的表示方法来描述视频的不变特性。在解决这个问题之后,拷贝检测又将面临着另一个核心问题,就是怎样进一步来提高视频拷贝检测的精度,其研究重点在于寻找适当的视频相似度匹配算法。

针对以上问题,提出一种基于多特征综合的视频拷贝检测方法提高系统对于多种多样的拷贝变换的鲁棒性,解决海量视频数据的高效检索问题;并进

一步使用基于 NBS^[2] 的图像相似度比对技术方法,来提高视频帧内容相似度比对的精度;最后提出一种改进的时序一致性的帧序列匹配方法解决拷贝帧的精确定位问题。

1 视频拷贝处理及检测方法分析

视频拷贝处理的方式多种多样,TRECVID 总结了常见的一些拷贝变换,并根据他们不同的性质将拷贝变换分为以下几种类型:摄像机重采样、画中画、模式添加、重新编码、伽马值调整、降低质量(模糊、伽马、丢帧、对比度、比例变换、压缩、白噪声等)、后期制作(剪切、偏移、对比度、插入字幕、垂直镜面、模式插入、画中画等)以及多种变换方法的任意组合。目前,研究人员都是针对这些变换类型来开展拷贝检测技术研究的。

图 1 是一些拷贝变换方法的图例。



图 1 拷贝变换

Fig. 1 Video copy transformation

多种多样的拷贝变换方法给拷贝检测带来了极大的挑战。拷贝检测的第 1 个核心问题在于寻找有效的表示方法来描述视频内容,以提高拷贝检测对多种拷贝变换的鲁棒性。针对这个问题,学者们提出了很多方法来解决这个问题^[3-8]。这些方法一般是基于两种类型的特征:局部特征和全局特征。

由于局部特征^[9-11]在图像检索方面取得的成功,最近研究人员也开始把局部特征应用于拷贝检测。文献[12]将局部特征 Harris 算子用于关键帧匹配。文献[13]对每帧图像检测 Harris 特征点,提取特征点附近区域的边缘直方图作为局部特征。文献[14]提出一种用 Harris 兴趣点的图像拷贝检测方案。通过 Harris 检测器提取图像的局部特征点,

并以 Harris 兴趣点为中心的多重圆环区域的信息熵作为特征向量,计算特征相似性判断检测图像是否为原始作品的拷贝。总的来说,基于局部特征的拷贝检测系统存在两方面的问题:一方面,基于特征点的算法可以避免局部变化问题,但是由于特征点的数目巨大,造成对计算和存储的需求较大^[3]的问题;另一方面,这类算法的前提是拷贝变化不会改变关键点空间分布顺序^[15],如果受到一些静态区域(如背景)的影响,会产生一些错报情况。

全局特征(如颜色直方图、颜色布局描述子、序数亮度特征)对于模糊、压缩和伽马等变换比较稳定,所以也被用于视频拷贝检测。在文献[16-17]中使用图像分块颜色直方图特征进行图像相似度比

较,这种方法简单快速,但是对于颜色、对比度改变的拷贝变换不具鲁棒性。为了解决这样的问题,研究者又提出基于分块亮度顺序(OM)的视频拷贝检测方法^[5,18-20]。文献[20]首先将OM方法应用于立体图像匹配。基于OM的方法对视频亮度、色调、尺寸变化都有较好的鲁棒性,并具有计算复杂度较低的特点和较好的检测性能^[21]。文献[19]提出一种改进的结合时序信息的OM方法,能检测添加边框的拷贝视频。但由于OM分块数少,对视频内容的区分能力较弱,检测精度有待提高。总的来说,全局特征对图像全局统一的变换具有鲁棒性。然而对于会改变图像全局特征分布的局部变换会降低这类算法的匹配效果^[22]。

综上所述,单独使用全局特征或局部特征都仅能应对少数几种特定的拷贝变换,而实际应用中的拷贝变换是复杂多样的,单一的特征不能很好地解决问题。所以基于多特征综合的方法必将成为未来视频拷贝检测的一个发展趋势。文献[15]提出一种多特征匹配的拷贝检测算法,采用级联式检测过滤框架,在提取视频帧图像的全局特征用于检测画面轻微变化的重复视频片段后,再使用局部特征等来检测各种复杂变化后的拷贝内容。但是与大多数现有方法相同,文献[15]没有对查询视频进行预处理,这必将影响检测的精度。

拷贝检测的第2个核心问题在于寻找有效的视频相似度匹配算法来提高检测精度,由于生成拷贝的变换方法的不确定性,拷贝检测的内容相似性比对方法很难对每种变换都有很好的鲁棒性,使得检测难度大大增加。

根据以上两个拷贝检测的核心问题和现有方法存在的不足,提出了一种完整的拷贝检测框架来应对基于内容的视频拷贝检测。该方法综合局部和全局特征来解决视频拷贝检测问题,以避免单特征拷贝检测系统难以应对多种多样的拷贝变换的难题,由于其有很好的召回率,并且效率较高,可以很好地解决海量视频数据的粗检索。然后进一步提出使用基于NBS的图像相似度比对方法来提高视频内容相似性比对的计算精度。

2 多特征的拷贝检测系统框架设计

为了提高系统对于多种拷贝变换的鲁棒性,本文综合了传统的局部和全局视觉特征,并使用基于

非正交二值子空间(NBS)的归一化互相关(NCC)精细过滤过程来提高拷贝视频帧内容相似度计算的效果。除此之外,还采用其他一些措施,如预处理和改进的基于时序一致性的帧序列匹配,来对拷贝视频的判定结果进行精化。

图2展示了本文方法的系统框架。在这个系统框架中,拷贝检测工作分为五个关键步骤:查询视频预处理、特征提取、粗检索、NCC精细过滤和改进的基于时序一致性的帧序列匹配。此方法跟以往方法的最大不同在于:1)综合多种特征来提高检测对于各种拷贝变换的鲁棒性;2)采用基于NBS的NCC^[2]方法对视频帧内容进行更精细的比对;3)使用改进的基于时序一致性的帧序列匹配方法以得到更加准确的帧序列。

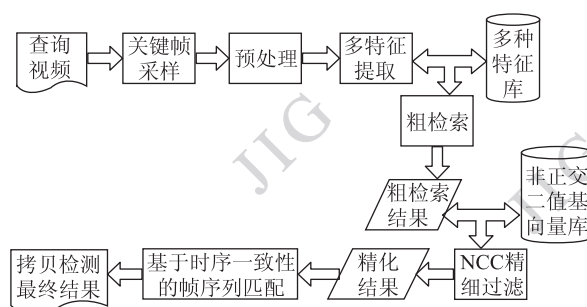


图2 拷贝视频检索系统构架图

Fig. 2 Video copy detection system architecture diagram

首先每隔1s对视频进行一次关键帧采样,仅把采样所得的关键帧用于接下来的处理。接着对查询视频进行预处理,对添加字幕、边框、偏移等拷贝变换进行反变换操作,以减小这些变换对检测结果的影响;然后提取关键帧和关键帧中心的颜色自相关图^[23]和SURF(speeded up robust features)^[24]特征,使用快速近似的最近邻索引库(FLANN)^[25]建立被查询视频特征向量的索引,以便进行快速高效的第1阶段粗检索;接着,为了提高检索精度,首次提出使用基于NBS图像表示方法的NCC过滤掉粗检索阶段返回的无效帧;最后应用一种改进的基于时序一致性的帧序列匹配方法由帧级别的结果得到最后视频级别的检测结果。

2.1 预处理

根据TRECVID对拷贝变换的分类,一些变换方法对检测的影响是可以预先消除的。为了消除加边框、添加字幕、改变画面比例、偏移和画面裁切等一些变换方法对检测效果的影响,在预处理阶段对查

询视频进行了去除边框以及屏蔽字幕等操作。因为采用的 SURF 特征对于一些图像质量缺失和人工编码比较稳定^[24],而颜色自相关图特征对于噪声等比较稳定,所以预处理步骤中对于改变亮度、模糊和加白噪声等变换没有进行特殊处理。

对于剪切、偏移等拷贝变换,应用边框检测算法检测到黑色边框并将它们去除,然后生成一个无黑色边框的新的视频。对于加入上下边框而改变画面原始比例的视频,不但去掉黑框并且把视频放回到原始比例。图 3(a)(b)清楚地展示了变换结果。

去掉黑框后,应用模式插入检测算法检测到插入到视频的字幕和静态图片等区域,然后为每个查询视频生成一个二值的掩码,如图 3(c)所示。这个掩码文件把人工插入内容的区域掩盖住,在特征提取步骤中这些区域的特征将不被提取。

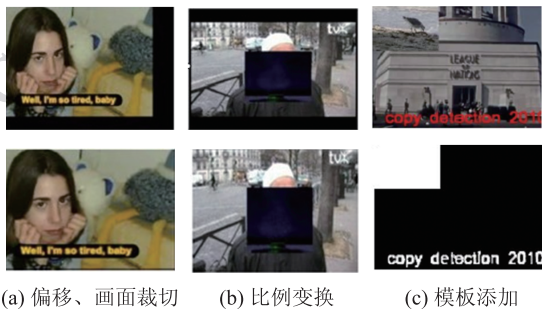


图 3 预处理

Fig. 3 Preprocess

2.2 特征提取

视频拷贝检测技术的第 1 个核心问题就是怎样提高拷贝检测对多种多样的拷贝变换的鲁棒性,其研究重点在于寻找有效的表示方法来描述视频的不变特性。由于每种特征的性质不同,单一的一种鲁棒的特征只能应对少量几种的拷贝变换,而实际情况中拷贝变换是复杂多样的。因此采取多种特征综合的方法来应对复杂的拷贝检测问题已势在必行。

2.2.1 常规特征提取及综合

由于全局和局部特征都具有其独特的性质,对于不同的拷贝变换有着不同的敏感度和鲁棒性,所以采取综合局部和全局特征的方法来弥补他们各自的不足。

由于图像在经过模糊、对比度、亮度变化等全局拷贝变换后,其颜色自相关图特征仍基本保持稳定^[23],因此选择颜色自相关图特征来应对改变关键帧全局内容拷贝变换。

但是关键帧的局部内容变换(如画面剪裁、模板添加等)将改变全局信息的分布,所以用全局特征进行匹配的效果不佳。相比于全局特征,局部特征对复杂的编辑和后期处理等变换有较好的效果。

在基于局部特征的视频拷贝检测系统中,选择合适的描述子对提高系统的性能具有决定性的作用^[26]。当前应用比较广泛的几种典型的局部特征有尺度不变特征变换(SIFT)、梯度位置方向直方图特征(GLOH)、PCA-SIFT 特征、SURF^[24]特征等。文献[27]对主要的局部描述子进行了比较研究,评价结果表明 SIFT 描述子在对象识别方面具有最好的性能,它具有旋转、尺度和亮度不变性,同时对仿射形变、视角改变和噪声等也有一定的鲁棒性。SURF 是 SIFT 特征的加速版本,计算复杂度比 SIFT 大大地降低,但是能够获得与 SIFT 相近的性能。文献[26]也证实了在拷贝检测领域中,SURF 与其他局部描述子相比,不但计算速度快,而且对不同的拷贝变换也很好的鲁棒性,所以选取 SURF 特征来应对改变关键帧局部内容的拷贝变换。而为了减少模板添加、偏移等变换对视频内容全局信息分布的影响,还提取了关键帧中心 50% 面积的 SURF 和颜色自相关图特征(以下称之为中心 SURF 和中心颜色特征)。

对于几种特征的综合策略,采取最简单的综合手段,而这种综合对于提高召回率有所帮助,在本系统中,由于有后续的精化检索处理流程,这里未对特征综合的权重作深入研究。

2.2.2 视频精确比对特征提取

仅仅使用全局和局部特征的综合,可以提高检测系统对于不同拷贝的鲁棒性,这种综合的策略可以得到比较好的召回率,而且效率相对较高,主要解决了海量视频数据的粗检索。但是粗检索阶段有可能会返回一些相似度不高的帧。为了要进一步提高视频帧的相似度比对精度,受到文献[2]的启发,又使用了一种 NBS 图像表示方法,并在此基础上以 NBS 模板匹配的方法来提高帧的相似度匹配的精度。

NBS 模版表示的基本思想是用一组较少数量的类似哈尔(Haar)的二值盒子函数(二值的基,如图 4 所示)对模板进行重构,这种表示方法非常便捷。在模板匹配时可以借助预先计算好的积分图像来提高匹配的效率和也能保证匹配的准确性。

图 5 所示为本文实验中用 40 个非正交二值基

向量重构一幅眼睛模板图像的示例。由于 NBS 是用二值的基来表示一幅图像或者一组图像,所以将这些二值基称为 NBS 特征。这些基更贴近图像的语义层面,并且它们的内积计算非常便利^[2]。

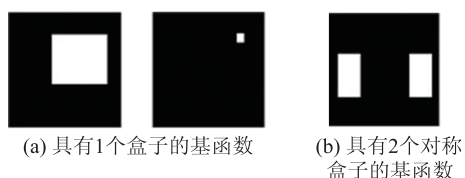


图4 基函数

Fig. 4 Box functions

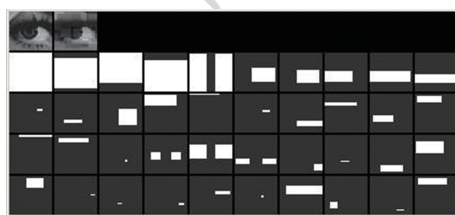


图5 由前40个非正交二值基向量重构的图像

Fig. 5 Reconstruction of the picture using 40 NBS base vectors

然而在一组预定义的非正交基向量字典中寻找最优的子空间表示是一个 NP 问题。所以这类问题经常用一种贪婪算法来寻找次优的解决方法。在本文的系统中,使用 OOMP^[2] 算法来寻找最佳的子空间表示。首先使用 MSER^[28] 算法对被查询视频的关键帧进行分割,并提取每个小块的 NBS 特征,并在后续处理中利用 NBS 对被查询视频的分割块与查询视频进行模板匹配,以此提高视频内容相似性比对的精度。

2.3 视频拷贝检测策略

单一特征不能应对所有拷贝变换,本文为此综合全局和局部特征的拷贝检测方法解决这个问题。但仅仅这样并不能提高视频内容相似度比对的精度。故提出一种粗检索和精细过滤的两步检索策略。该策略可以有效地提高检索的准确率。

在粗检索阶段,充分利用全局和局部特征对拷贝视频的关键帧进行初步检测。其中基于局部特征的拷贝检测系统所面临的一个普遍难题就是高效的特征索引^[25],当今学者们已经提出很多索引方法,如 ANN^[29]、LSH^[30] 和 kd-树算法^[31-32] 等。这里使用 FLANN^[33] 来有效解决局部特征的高效索引问题。首先分别建立被查询视频关键帧的 SURF、颜色自相关图、中心 SURF 和中心颜色自相关图 4 种

特征向量的索引,然后分别返回与查询视频关键帧特征向量最邻近的 30 个特征向量,最后把 4 种结果进行综合。返回的最邻近的特征向量的个数对于最终的实验结果有一定的影响,通过实验我们发现 30 是一个比较好的选择,在后续实验中,都采用 30 作为返回的最邻近特征向量的个数。

在精细过滤阶段,为了对查询视频和被查询视频的关键帧进行精细的相似性比对,首先提取使用 MSER 分割后的被查询视频的关键帧小块的 NBS 特征,并在此基础上利用 NCC 对这些小块与查询视频关键帧进行模板匹配。对于如何判定被查询视频关键帧是否与查询视频关键帧的内容相似,采取以下策略:使用 NCC 检测被查询视频关键帧的分割小块中是否有能与查询视频关键帧中的某部分匹配的,如果匹配的小块数超过某一阈值则认定这个被查询视频关键帧是一个准确的结果,并按照匹配结果给予这一帧打分;否则将认定这一帧为无效帧并丢弃。这样可以排除上一步粗检索返回结果中的无效被查询视频关键帧,并且可以给予有效的关键帧更高的打分。

3 拷贝视频的精确定位技术

视频拷贝检测的第 2 个核心问题就是怎样进一步提高视频拷贝检测的精度,其研究重点在于寻找适当的视频相似度匹配算法。为了提高拷贝视频检测的准确率,在传统的检索框架之上又增加了一个拷贝视频精确定位的步骤,并首次提出使用基于 NBS 的 NCC 方法对视频帧内容进行精细的相似性比对,并提出一种改进的基于时序一致性的帧序列匹配方法对帧进行精确定位。

3.1 基于 NBS 的图像相似度比对技术

基于 NBS 的图像相似度比对是精细过滤阶段的主要精髓,而它的技术关键就在于如何对图像内容进行精确的相似度比对。NCC 是一种图像的模板匹配方法,由于效果良好,在实际应用中被广泛使用。所以提出使用 NCC 来解决图像内容的相似度比对问题。一个大小为 $(2h+1) \times (2w+1)$ 的模板 t 在图像 x 中坐标为 (u, v) , NCC 计算

$$c(u, v) = \frac{\sum_{i=-h}^h \sum_{j=-w}^w X(i, j) T(i, j)}{\sqrt{\sum_{i=-h}^h \sum_{j=-w}^w X(i, j)^2} \sqrt{\sum_{i=-h}^h \sum_{j=-w}^w T(i, j)^2}} \quad (1)$$

$$\begin{aligned} X(i, j) &= x(u+i, v+j) - \bar{x} \\ T(i, j) &= t(h+i, w+j) - \bar{t} \end{aligned} \quad (2)$$

图 6 是实验中 NCC 的一个应用实例,其清晰地展示了 NCC 的过程和原理:对于一个图像小块,首先应用 NBS 对其进行重构,再使用 NCC 模板匹配过程对这个小块在原图像中进行快速匹配。



图 6 NCC 应用实例

Fig. 6 NCC application example

3.2 改进的时序一致性的帧序列匹配方法

经过拷贝视频关键帧内容的相似度精确比对之后,使用一种基于时序一致性的帧序列匹配方法由帧级别的结果得到视频级别的最终结果。

受文献[12]帧序列匹配方法算法的启发,采用相关分数矩阵(图 7)进行帧序列匹配,并对其进行了一些改进。把文献[12]中二值的相关分数改变成精细过滤步骤中 NCC 的相似帧打分,这种评分标准与文献[12]相比,可以使得正确查询结果和错误查询结果的区分度更大,以便更好地降低检测的错报率,提高检测的准确率。

算法的前提是被查询视频至少有一个关键帧与查询视频的一个关键帧相匹配。假设查询视频和被查询视频分别有 Q 和 R 个关键帧。定义 $p(i, j)$ 为查询关键帧 i 与被查询关键帧 j 的一个匹配对, $N(i, j)$ 是 $p(i, j)$ NCC 的打分, $T_q(i)$ 为查询关键帧 i 的时间戳, $T_r(j)$ 为被查询视频的时间戳。图 7 展示了相关分数矩阵,所有的匹配对用一个黑色方块表示。

对其中某一个匹配对 $p(i, j)$, 首先计算时间戳差: $\Delta = T_r(j) - T_q(i)$ 。然后找到所有时间戳差在 $[\Delta - \delta, \Delta + \delta]$ 范围内的匹配对集合 F (这里设置 δ 为 5 s), 如图 7 两条虚线表示时间戳差的区域内所有的黑色方块。匹配对 $p(i, j)$ 的扩展相关分数就是虚线范围内所有匹配对 (黑色方块) 的 NCC 打分之总和。为减少查询视频关键帧数量 N 的影响, 把扩展相关分数用 $1/\log(N)$ 标准化 (如图 7 中 $N = i_5 - i_1 + 1$)。

在计算所有匹配对的扩展相关分数之后, 选取分数最大者作为查询视频与被查询视频匹配的打

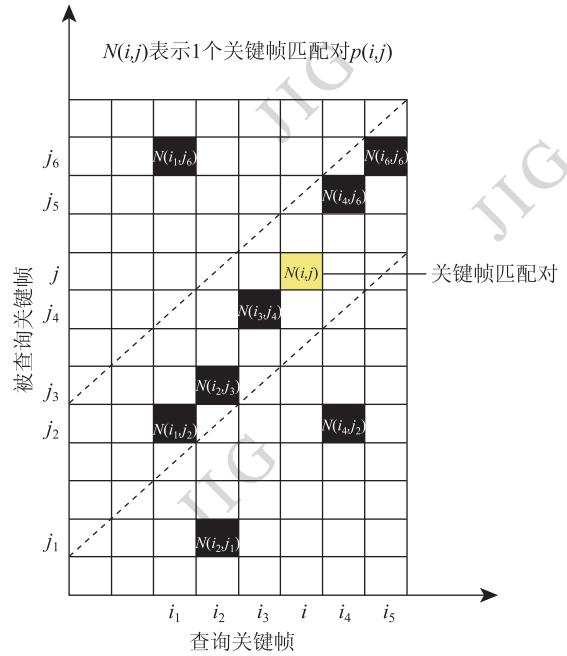


图 7 相关分数矩阵图

Fig. 7 Relevance scores matrix

分。最后对单个被查询视频所有返回的查询视频的打分排序, 选取分数最高的前几名最为视频拷贝检测的最终结果。

4 实验结果与分析

针对视频拷贝检测的核心问题, 通过实验验证本文提出的视频拷贝检测方法对于不同拷贝变换的鲁棒性、拷贝检测的准确性以及系统的时间效率。

实验在硬件配置为 Intel Xeon E5450 CPU 3.00 GHz quad x2, 32 G 内存的 PC 机上进行。实验数据从 TRECVID2010 拷贝检测比赛的数据库中选取, 通过选择一部分视频, 然后对其使用较为常见的视频拷贝变换手段, 得到仿真视频数据集, 最后在原始视频数据集上考察本文算法的性能。

TRECVID2010 的数据库中共有 11 525 个视频文件, 选择 3 500 个作为实验数据集, 然后从中选取 300 个视频 (时长 1 ~ 5 min), 分别进行常见的拷贝变换: 改变亮度、加高斯噪声、加边框、改变长宽比例、偏移、添加字幕、模糊、改变分辨率以及多种变换的组合, 作为查询视频。

应用视频拷贝检测方法常用的评价标准: 召回率 R 和准确率 P 来对实验结果进行评定, 分别定义为

$$R = \frac{T}{A} \quad (3)$$

$$P = \frac{T}{T+F} \quad (4)$$

式中, T 表示系统检测到的正确拷贝视频数, A 表示拷贝视频总数, F 表示系统检测到的错误拷贝视频数。

4.1 视频拷贝检测准确性分析

在视频的拷贝检测中,检测者不但需要将所有可疑的对象检测出来,即拥有好的检测召回率,同时在检测结果中正确的拷贝对象还应该具有较高的准确率。在召回率相同的情况下,系统的检测准确率越高,检测性能越好^[34]。

图8所示为本文方法与传统的几种拷贝检测方法 OM^[5]、TOM^[19] 和文献[34]方法的 PR 曲线比较曲线图。

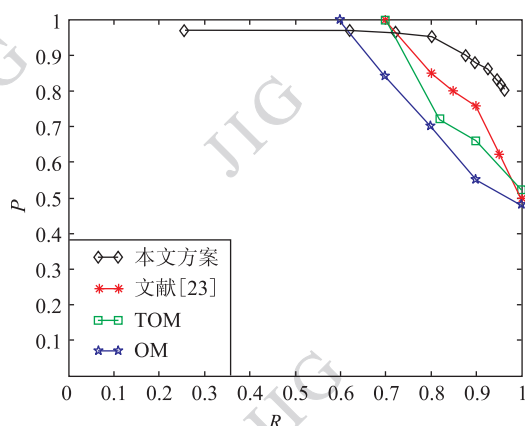


图8 实验效果对比图

Fig. 8 Experimental performance comparison figure

由图8可见,本文系统在召回率很高的前提下,保证了检索的准确率。本文的实验数据比文献[34]的数据量更大,视频内容更丰富,变换种类更复杂,而且还增加了多种变换的组合变换。当召回率为0.87时本文方法的准确率达到0.9,而OM、TOM和文献[34]方法的准确率仅为0.53、0.65和0.75。虽然TOM和文献[34]的方案在准确率为0.7时召回率达到了1,但是本文方法却具有更好的平均检测效果。这表明本文提出的基于NBS的NCC结果精化过程有效地提高了系统的检测精度。

4.2 多特征综合鲁棒性分析

图9展示了系统对每种变换不同的效果,显而易见的,系统处理多种不同的拷贝变换时都具有很稳定的效果,具有较好的鲁棒性,这归功于多种特征

综合的设计,每种特征都弥补了其他特征的不足点。其中系统对于亮度变换和增加字幕变换的检测效果较其他变换稍差一些。原因可能是亮度变换后对颜色自相关图特征的影响比较大,而添加字幕变换对于SURF特征的影响比较大,从而影响了最终的检测效果。

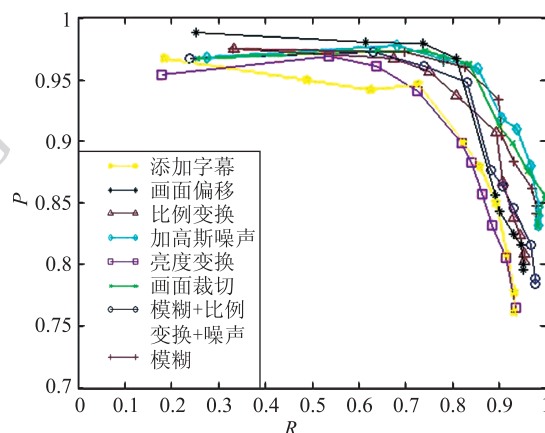


图9 不同变换的实验效果图

Fig. 9 Different transformation experimental performance figure

由于系统采取了时间间隔的帧采样方式,所以减少了大量的冗余数据和计算时间。对于播放总时长约为8.08 h的300个查询视频,系统的处理耗时约3.12 h。本文系统的处理效率与文献[3]的相当。

由此可知,本文方法对于拷贝视频内容相似度比对的敏感性和对于多种拷贝变换的鲁棒性都较强。这说明,基于多特征的视频拷贝检测方法是有效的。

5 结 论

视频拷贝检测技术的两个核心问题就是如何提高拷贝检测对多种多样的拷贝变换的鲁棒性,以及如何进一步提高视频拷贝检测的精度。针对这两个问题提出了一种基于多特征综合的方法来进行基于内容的视频拷贝检测。该方法的贡献在于首次提出使用基于NBS的图像相似度比对方法来对视频内容相似度进行精确比对,提高了检索精度;系统的整体框架和多特征综合的检测处理设计来应对多种多样的拷贝变换,以提高检测的鲁棒性;使用了一种改进的基于时序一致性的帧序列匹配方法以得到更加准确的帧序列。实验效果表明,本文方法在保持召

回率很高的前提下,得到了良好的检测准确率。

未来将对多种变换相结合的拷贝视频以及多种特征综合的权重进行进一步研究,以达到更好的检测效果,并进一步提升检测效率。

参考文献 (References)

- [1] Paul O. Guideline for TRCVID 2010[EB/OL]. (2011-04-26) [2011-12-01]. <http://www-nlpir.nist.gov/projects/tv2010/tv2010.html#ccd>
- [2] Tang F, Crabb R, Tao H. Representing images using nonorthogonal haar-like bases[J]. *Pattern Anal. Mach. Intell.*, 2007, 29(12): 2120-2134. [DOI:10.1109/TPAMI.2007.1123]
- [3] Guo J B, Li J T, Zhang Y D, et al. Video copy detection based on trajectory behavior pattern[J]. *Journal of Computer-Aided Design and Computer Graphics*, 2010, 22(6): 943-958. [郭俊波, 李锦涛, 张勇东, 等. 基于轨迹行为模式特征的视频拷贝检测算法[J]. *计算机辅助设计与图形学学报*, 2010, 22(6): 943-958.]
- [4] Zhao W L, Ngo C W. Scale-rotation invariant pattern entropy for keypoint-based near-duplicate detection[J]. *IEEE Trans. on Image Processing*, 2009, 18(2): 412-423. [DOI:10.1109/TIP.2008.2008900]
- [5] Hua X, Chen X, Zhang H. Robust video signature based on ordinal measure[C]//*Proceedings of International Conference on Image Processing*. Beijing, China: Microsoft Res., 2004: 685-688. [DOI:10.1109/ICIP.2004.1418847]
- [6] Usman M, Kim C. Real time video copy detection under the environments of video degradation and editing[C]//*Proceedings of the 10th International Conference on Advanced Communication Technology*. Daejeon: Inf. & Commun. Univ., 2008: 1583-1588. [DOI:10.1109/ICACT.2008.4494083]
- [7] Paisitkriangkrai S, Mei T, Zhang J. Scalable clip-based near-duplicate video detection with ordinal measure[C]//*Proceedings of the ACM International Conference on Image and Video Retrieval*. New York, USA: ACM, 2010: 121-128. [DOI:10.1145/1816041.1816062]
- [8] Law-To J, Chen L, Laptev I. Video copy detection: a comparative study[C]//*Proceedings of the 6th ACM International Conference on Image and Video Retrieval*. New York, USA: ACM, 2007: 371-378. [DOI:10.1145/1282280.1282336]
- [9] Jégou H, Douze M, Schmid C. Hamming embedding and weak geometric consistency for large scale image search[J]. *Lecture Notes in Computer Science*, 2008, 5302: 304-317. [DOI:10.1007/978-3-540-88682-2_24]
- [10] Nistér D, Stewénus H. Scalable recognition with a vocabulary tree[C]//*Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Lexington, USA: University of Kentucky, 2006(2): 2161-2168. [DOI:10.1109/CVPR.2006.264]
- [11] Philbin J, Chum O, Isard M. Object retrieval with large vocabularies and fast spatial matching[C]//*Computer Vision and Pattern Recognition*. Oxford: University of Oxford, 2007: 1-8. [DOI:10.1109/CVPR.2007.383172]
- [12] Joly A, Buisson C, Frelieot. Contentbased copy retrieval using distortion based Probabilistic similarity search[J]. *IEEE Transactions on Multimedia*, 2007, 9(2): 293-306. [DOI:10.1109/TMM.2006.886278]
- [13] Law-to J, Gouet-brunet V, Buisson O, et al. Local behaviour labelling for content based video copy detection[C]//*Proceedings of the 18th International Conference on Pattern Recognition*. Hong Kong, China: INA \ INRIA, 2006: 232-235. [DOI:10.1109/ICPR.2006.767]
- [14] Ling H F, Chen Y, Zhou F H, et al. Copy detection scheme based on Harris interest points area[J]. *Computer Application*, 2009, 46(1): 224-249. [凌贺飞, 陈勇, 邹复好, 等. 基于Harris兴趣点区域的图像拷贝检测算法[J]. *计算机应用*, 2009, 46(1): 244-249]
- [15] Wu X, Gao K, Guo J B, et al. Video Copy Detection Based on Multiple Visual Feature Matching[J]. *Journal of Computer Aided Design and Computer Graphics*, 2010, 22(11): 1856-1865. [吴潇, 高科, 郭俊波, 等. 基于多特征匹配的视频拷贝检测算法[J]. *计算机辅助设计与图形学学报*, 2010, 22(11): 1856-1865]
- [16] Hsu W, Chua T S, Pung H K. An integrated color-spatial approach to content-based image retrieval[C]//*Proceedings of ACM Multimedia*. San Francisco, CA: ACM, 1995: 305-313. [DOI:10.1145/217279.215284]
- [17] Wu W, Zhang Y D, Tang S. A hierarchical scheme for rapid video copy detection[C]//*Proceedings of IEEE International Workshop on Computer Vision Applications*. Beijing, China: Key Lab. of Intell. Inf. Process., Chinese Acad. of Sci., 2008: 1-6. [DOI:10.1109/WACV.2008.4543992]
- [18] Mohan R. Video sequence matching[C]//*Proceedings of the International Conference on Audio, Speech and Signal Processing*. Yorktown Heights, NY: IEEE, 1998, 6: 3697-3700. [DOI:10.1109/ICASSP.1998.679686]
- [19] Kim C, Asudev B. Spatiotemporal sequence matching for efficient video copy detection[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2005, 15(1): 127-132. [DOI:10.1109/TCSVT.2004.836751(410)1]
- [20] Bhat D N, Nayar K. Ordinal measures for image correspondence[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(4): 415-423. [DOI:10.1109/34.677275]
- [21] Hampapur Ar, Hyun K H, Bolle R. Comparison of sequence matching techniques for video copy detection[C]//*Proceedings of the Storage and Retrieval for Media Databases*. San Jose, California, USA: SPIE, 2002: 194-201.
- [22] Gao K, Lin S, Zhang Y D, et al. Object-based image retrieval using spatial context[J]. *Journal of Computer-Aided Design and Computer Graphics*, 2008, 20(11): 1452-1458 [高科, 林守勋,

- 张勇东,等. 基于空间上下文的目标图像检索[J]. 计算机辅助设计与图形学学报, 2008, 20(11):1452-1458]
- [23] Jing H. Color spatial image indexing and applications[D]. NY, USA: Cornell University Department of Computer Science Dissertation, 1998.
- [24] Bay H, Tuytelaars T, Van Gool L. Surf: speeded up robust features[J]. Lecture Notes in Computer Science, 2006, 3951: 404-417. [DOI: 10.1007/11744023_32]
- [25] Yusuke U, Motilal A, Shigeyuki S. Accurate content-based video copy detection with efficient feature indexing[C]//Proceedings of the 1st ACM International Conference on Multimedia Retrieval. New York: ACM, 2011: 19. [DOI: 10.1145/1991996.1992015]
- [26] Wen C H. Content-based video copy detection[D]. Shanghai: FuDan University, 2011. [文朝辉. 基于内容的视频拷贝检测研究[D]. 上海: 复旦大学, 2011.]
- [27] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors[J]. IEEE Trans. on Pat. Analysis and Machine Intelligence, 2005, 27(10): 1615-1630. [DOI: 10.1109/TPAMI.2005.188]
- [28] Michael D, Horst B. Efficient maximally stable extremal region (MSER) tracking[C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Styria, Austria: Graz University of Technology, 2006: 553-560. [DOI: 10.1109/CVPR.2006.107]
- [29] Andoni A. Near-optimal hashing algorithms for approximate nearest neighbour in high dimensions[C]//Annual IEEE Symposium on Foundations of Computer Science. Cambridge, MA: MIT, 2006: 459-468. [DOI: 10.1109/FOCS.2006.49]
- [30] Arya S, Mount D M, Netanyahu N S, et al. An optimal algorithm for approximate nearest neighbour searching in fixed dimensions[J]. Journal of ACM, 1998, 45(6): 891-923. [DOI: 10.1145/293347.293348]
- [31] Philbin J, Chum O, Isard M, et al. Object retrieval with large vocabularies and fast spatial matching[C]//IEEE Computer Vision and Pattern Recognition. Oxford: University of Oxford, 2007: 1-8. [DOI: 10.1109/CVPR.2007.383172]
- [32] Silpa-Anan C, Hartley R. Optimised kd-trees for fast image descriptor matching[C]//IEEE Computer Vision and Pattern Recognition. Canberra, Australia: ACT, 2008: 1-8. [DOI: 10.1109/CVPR.2008.4587638]
- [33] Muja M, Lowe D G. Fast approximate nearest neighbors with automatic algorithmic configuration[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Oxford: University of Oxford, 2007: 1-8. [DOI: 10.1109/CVPR.2007.383172]
- [34] Zhao Y X, Liu G J, Dai Y W. Video copy detection based on local ordinal[J]. Journal of Computer-Aided Design and Computer Graphic, 2009, 21(9): 1340-1343. [赵玉鑫, 刘光杰, 戴跃伟, 等. 基于局部排序的视频拷贝检测[J]. 计算机辅助设计与图形学学报, 2009, 21(9): 1340-1343.]