

# 一种新的近重复监控视频检测算法\*

郭丁云<sup>1</sup>, 杨艳芳<sup>2</sup>, 朱俊俊<sup>1</sup>, 齐美彬<sup>1</sup>

(1.合肥工业大学 计算机与信息学院, 安徽 合肥 230009;

2.合肥工业大学 电子科学与应用物理学院, 安徽 合肥 230009)

**摘要:** 设计了一种使用视频镜头时序特征来实现级联式检测近重复视频的算法。首先在进行关键帧特征提取之前, 直接在镜头层次上提取时序特征, 初步滤除完全不同的视频, 然后对剩下的视频帧提取全局颜色特征和 SURF 特征进行逐步检测, 最终获得与查询视频近重复的视频。对实验室的监控视频进行小范围的验证实验, 实验结果表明, 该算法与不用时序特征的方法相比有一定的有效性和准确性。

**关键词:** 监控视频; 近重复检测; 时序特征; 全局颜色特征; SURF

中图分类号: TP391

文献标识码: A

文章编号: 1674-7720(2013)11-0053-04

## A new near-duplicate detection algorithm for surveillance videos

Guo Dingyun<sup>1</sup>, Yang Yanfang<sup>2</sup>, Zhu Junjun<sup>1</sup>, Qi Meibin<sup>1</sup>

(1.School of Computer and Information, Hefei University of Technology, Hefei 230009, China;

2.School of Electronic Science & Applied Physics, Hefei University of Technology, Hefei 230009, China)

**Abstract:** This paper introduces an approach using the temporal feature in the shot to detect near-duplicate videos in a cascaded way. Firstly, before extracting features of the key frames, extract the temporal feature directly in the shot level to preliminarily filter out some dissimilar videos and then extract global color features and SURF feature of the rest key frames step by step to obtain the ultimate videos that are duplicated with the query video. By performing the verification experiment on the surveillance videos in our laboratory, the experimental results show the usefulness and accuracy of our proposed algorithm compared with the one without the temporal feature.

**Key words:** surveillance video; near-duplicate detection; temporal feature; global color feature; SURF

近年来, 数字视频应用飞速发展, 在视频监控领域, 需要从成千上万个摄像头产生的海量视频中找到含有用户关心的线索的视频, 比如穿红色衣服女子唱歌的视频片段, 检索需要耗费大量的人力物力。因此, 视频检索、视频摘要以及视频编解码等应用应运而生, 如何在海量监控视频数据中快速、准确地检测到相同的视频片段已成为多媒体内容分析和视频检索中的一个重要课题。

传统的近重复视频检测一般采用参考文献[1]中描述的算法, 即首先将视频通过时间采样或镜头边界检测算法检测出多个子镜头, 提取出各子镜头中的一个或多个关键帧图像; 然后使用某些高维特征向量(如颜色直

方图、局部二值模式(LBP)等)表示这些关键帧, 作为视频的全局签名; 最后通过某种相似性度量函数来计算两个视频的关键帧序列, 以此来检测两个视频是不是重复或近重复视频。参考文献[2]提出了一种基于全局签名的视频重复检测算法, 提出了一种称为视频直方图的视频签名, 用来表示视频特征向量在特征空间的分布情况。参考文献[3]介绍了一种结合时间-空间分布信息的时序特征和色彩范围的视频重复检测算法。参考文献[4]提出将视频关键帧的全局特征与局部关键点结合起来检测相似视频, 即采用分层的方法, 通过比较视频签名过滤掉一些完全不同的视频, 从而减少基于局部关键点的相似关键帧检测的计算量, 然后再用局部关键点检测剩下的变化较大的视频。这种以视频内容为视频序

\* 基金项目: 国家自然科学基金项目(61172164)

列匹配依据的方法一般都是在所有关键帧图像上直接提取全局颜色特征,再使用图像匹配算法比较相似性,在大量关键帧图像的特征匹配中不仅占用大量计算时间,而且会丢失视频序列的时间一致性信息。

受参考文献[4]启发,本文在提取关键帧图像的特征提取之前先在镜头层次提取出一种新的特征作为镜头的时序特征,再提取全局颜色特征和更精确的 SURF 局部特征进行检测,最终得到与查询视频近重复的视频,算法具体框架如图 1 所示。

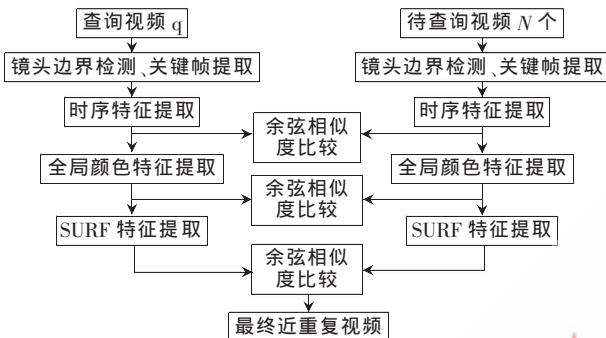


图1 本文算法设计框架

## 1 近重复监控视频检测算法

### 1.1 镜头边界检测

镜头边界检测是视频索引和检索的第一步,即找出视频序列中发生镜头变换的位置,以进一步将视频分成独立的镜头片段。本文采用的是基于边缘轮廓差值法来检测镜头边界的方法<sup>[5]</sup>,使用 Canny 边缘检测算子依次逐帧检测出各图像帧的边缘,将消失的旧边缘和增加的新边缘的差异不连续值与相邻帧的不连续值作比较,通过设置合适的阈值,可以达到一定的自适应性,以满足不同视频的镜头边界检测。具体实现步骤如下:

(1) 初始化读入视频的第 1 帧图像,将图像转化为灰度图像,以便使用 Canny 边缘检测算子检测出其边缘,统计其边缘元素个数,并通过形态学操作得到该图像的膨胀图像,将得到的边缘二进制图像反色。

(2) 按照读入图像帧的顺序从第 2 帧开始直到视频的最后一帧,使用步骤(1)中的方法将各图像帧依次进行边缘检测,得到边缘元素个数、膨胀图像及反色边缘二进制图像。

(3) 通过图像帧的两两依次进行与操作并统计前一图像帧消失的边缘点数目和后一图像帧新增的边缘点数目,定义一个边缘轮廓差值函数 framedif 来进行镜头边界点的评价度量:

$$\text{framedif}(i-1) = \max\left(\frac{\text{out}}{p_{i-1}}, \frac{\text{in}}{p_i}\right) \quad (1)$$

其中,  $i$  的值从 2 到最后一帧,  $p_{i-1}$  是前一图像帧的边缘数,  $p_i$  是后一图像帧的边缘数,  $\text{out}$  是前一图像帧消失的边缘点数目,  $\text{in}$  是后一图像帧新增的边缘点的数目。

(4) 为达到一定的自适应性以满足不同视频的镜头

边界检测,本文选取的比较度量是后一帧的边缘轮廓差值与前一帧的边缘轮廓差值的比值  $\text{ratioidif}$ ,即:

$$\text{ratioidif} = \frac{\text{framedif}(i)}{\text{framedif}(i-1)} \quad (2)$$

当  $\text{ratioidif}$  的值大于预先设置的阈值  $T$  时,就说明两个图像帧之间发生了镜头变化。

另外,考虑到有些监控视频中会有一些运动幅度变化很小的渐变视频,其镜头边界点非常不明显,其最后一帧的边缘轮廓差值与前一帧的边缘轮廓差值相差不大,这种情况下使用边缘轮廓差值法将检测不到符合大于阈值条件的镜头边界点。为了增加实验的广泛适用性和准确性,对于使用边缘轮廓差值法检测不到镜头边界点的情况,则只取其中一帧作为镜头边界点,将整个视频分成两个子镜头,然后再按照同样的关键帧提取算法进行检测。

### 1.2 关键帧图像提取

视频的匹配检测归根到底就是图像的匹配检测,本文采用的是基于帧差的欧式距离来提取各个子镜头中的关键帧,欧式距离越大,说明帧间的差异越大,越可能出现关键帧。帧差欧氏距离的表达式为:

$$\text{eulerdisdiff}(i) = \sqrt{\sum_{i=1}^{N-2} [(x_{i+2}-x_{i+1})-(x_{i+1}-x_i)]^2} \quad (3)$$

首先逐帧计算出各图像的帧差欧氏距离,即对第  $i+2$  帧图像的灰度值与第  $i+1$  帧图像的灰度值的差减去第  $i+1$  帧图像的灰度值与第  $i$  帧图像的灰度值的差的平方进行求和,对和值求算术平方根,然后计算这些帧差欧氏距离的极值以及极值点对应的函数值,再计算各函数值的均值,极值点中函数值大于均值的点即为所要选取的关键帧图像。

### 1.3 时序特征提取

视频是由一系列连续记录的帧图像组成的,是一个二维图像流序列,检测出的子镜头也是按照时间顺序排列的,因此每个子镜头中提取出的关键帧图像也必然存在着很大的时间一致性。前面提到,对于两个视频序列的匹配,大多数学者都是直接对各子镜头中每个关键帧图像提取全局或局部特征,在关键帧层次上进行两两比较,该方法在大规模视频数据集查找所需视频时,若对很多与查询视频明显不相同的干扰视频进行检测,则会花费大量时间。

本文提出一种新的特征来表示镜头的时序特征,直接在关键帧所在的子镜头层次上进行比较,可在提取关键帧图像的全局或局部特征之前,快速滤除与查询视频很不相同的无关视频。具体做法如下:

(1) 在基于视频序列连续一致性的前提下,先计算查询视频与待查询视频中所有视频每个关键帧文件夹中的帧图像数目,按照在关键帧中的时间顺序,将各个关键帧图像分成 9 块,计算其中间第 5 块的颜色特征。

(2) 将各个镜头中所有关键帧图像第 5 块的颜色特征组成一个行数为关键帧数目、列数为 36 的矩阵,即可用来表示每一个关键帧集合中的所有帧图像在连续时间上的时序特征。

(3) 将得到的查询视频的各关键帧子镜头的时序特征分别与待查询视频的所有关键帧子镜头的时序特征进行余弦相似度比较。

(4) 将计算得到的比较值与事先设定的阈值进行比较,若存在大于这一阈值的时序特征,就取出待查询视频中满足这一条件的关键帧子镜头,以进行进一步检测;若没有满足条件的关键帧子镜头,则说明这个待查询视频与查询视频很不相同,可以将其滤除,从而避免了再进行视频特征提取和两两特征匹配,大大节省检测时间。

#### 1.4 全局颜色特征提取

全局特征中颜色特征一般是颜色直方图,能简单描述一幅图像中颜色的全局分布。在这里也简单地提取出视频关键帧图像的 36 维的颜色直方图来表示这些关键帧,作为视频的全局颜色特征进行特征相似性比较。

#### 1.5 SURF 特征提取

局部特征描述图像局部区域信息,计算容易,能抵抗局部遮挡,对视觉变换不敏感。常用的局部关键点检测算子有 Difference-of-Gaussian (DoG) 算子、Harris-Affine 算子等。最常用的局部关键点描述子有 SIFT 算子、PCA-SIFT 算子、GLOH 算子以及依赖于积分图像的图像卷积的 SURF 算子<sup>[6]</sup>。鉴于 SURF 算法对于图像旋转、平移、缩放和噪声影响具有较好的鲁棒性,而且计算速度比 SIFT 算法快很多,本文把 SURF 的这种优势应用到下一步的局部特征提取中。

## 2 实验结果与分析

本文的实验是一个小范围的验证实验,所使用的视频来源于合肥工业大学计算机与信息学院 DSP 联合实验室各个小实验室的监控视频。由于每次拍摄的监控视频一般时长为 1 h~2 h,在验证实验中,重新构造了实验数据集,对于不同时间不同地点拍摄的各类视频进行截取,分割成时长均约为 1 分 30 秒的视频。通过混合不同时间、不同地点的视频,构造出每类包含 20 个这样视频的待查询视频集。在每类的 20 个视频中,选取一个视频为查询视频之后,其他的视频中,一类是包含查询视频的近重复视频,另一类是与查询视频不相同的视频,下面对这些视频进行验证实验。

本实验所用的计算机配置为 Intel (R) Core (TM) i3 2.93 GHz CPU, 2 GB 内存, 实验平台为 Matlab 7.11.0 (R2010b)。实验所用的实验评价标准为:(1) 定性分析使用这种算法的有效性和准确性;(2) 大规模视频集中进行拷贝检测和近重复检测常用的准确率和召回率。

实验中对很多不同同源视频的两类都进行了验证,

限于文章篇幅,本文只对其中一类的实验结果进行说明。选取的视频如图 2 所示,其中 1.avi 为预先假设的查询视频,2.avi 视频为选取的与查询视频在不同场景、不同主体人物下拍摄的监控视频,即看作无关的干扰视频,这里该类视频共有 7 个,剩下的如 3.avi 视频为与查询视频重复或近重复的视频,包含与查询视频有相同场景但主题人物不同和有不同场景但有相同主体人物动作的情况。这些视频的帧率均为 25 f/s,码率均为 46.1 Mb/s,画面尺寸为 320 pix×240 pix,包括 1.avi 在内的这 20 个视频中共有 13 个与查询视频重复或近重复的视频,有 7 个不相同的视频。



图 2 本文中选取的监控视频

在实验中进行时序一致性特征和分块颜色特征比较时,本文采取一般算法使用的直接将余弦值与事先设定的一个阈值相比较的度量方法。满足条件的关键帧集合对应的被查询视频有可能是要检测出来的重复或近重复视频,将再进行下一步更精确的检测;而不满足条件的即为与查询视频不相同的视频,直接滤除掉,以减少后面的检测量。但是在进行 SURF 特征的比较时,本文采用的是一种同时满足两个阈值的度量方法。每一对 SURF 特征相比较的余弦度量值的维数为  $M_i \times N_j$ ,其中  $M_i$  表示查询视频中一个镜头中第  $i$  个关键帧的局部关键点数目, $N_j$  表示一个被查询视频对应的一个镜头中第  $j$  个关键帧的局部关键点数目。当比较这个余弦值是否满足近重复检测的条件时,首先计算这个矩阵中的元素个数,即  $M_i \times N_j$  个;再计算矩阵中大于第一个阈值  $T_1$  的元素个数,记为  $M$ ;然后比较  $M$  和  $M_i \times N_j$  的比值是否大于第二个阈值  $T_2$ 。同时满足这两个阈值条件的关键帧集对应的被查询视频就是最终得到的重复或近重复视频。

本实验中,当查询视频 1.avi 与 20 个待查询视频相比较时,通过每一步都可以滤除掉一部分视频,并最终检测出重复或近重复视频。在进行时序一致性特征比较时,当阈值设置为 0.985 时,两个完全不同场景、不同人物的视频首先被滤除;在进行全局颜色特征比较时,当阈值设置为 0.99 时,滤除 3 个不同的视频;最后,在进行 SURF 特征比较时,当设置阈值  $T_1=0.6$ 、阈值  $T_2=0.7$  时,滤除 1 个视频。实验最终剩下 14 个视频,其中 13 个重复或近重复视频被全部检测出来,7 个不相同的视频只滤除了 6 个,还有一个不相同的视频没有被滤除而和近重复视频一起被误检出来。假设 CAS\_TF 表示本文提出的使用了视频镜头时序特征的方法,M\_GLOF 和 M\_HIER 分别表示直接使用全局颜色特征和联合全局特征与局部特征的分层方法,则使用这 3 种方法进行小范

围的视频检测的实验结果比较如表 1 所示,其中  $p/q$  表示在  $q$  个待查询视频中返回  $p$  个重复或近重复视频。

表 1 实验结果比较

$p/q$	初始 视频	时序特征 比较后	分块颜色特 征比较后	SURF 特 征比较后	召回率	准确率
M_GLOF	13/20	.....	13/17	.....	13/13=1.0	13/17=0.765
M_HIER	13/20	.....	13/17	13/16	13/13=1.0	13/16=0.813
CAS_TF	13/20	13/18	13/15	13/14	13/13=1.0	13/14=0.929

由表 1 可见,3 种方法达到了相同的召回率,即 3 种方法都检测出了所要检测的重复或近重复视频,但是本文方法的准确率明显高于前两种方法。由于本文方法先在镜头层次使用了镜头的时序一致性特征,在进行每个关键帧图像的特征提取之前进行比较,滤除了一部分不相同的视频,减少了关键帧图像层次上特征的直接比较。本文提出的代表时序特征的过程简单,计算也比较容易,为后面进行全局特征和局部特征的比较缩小了范围,减少了计算量,若应用到大规模的视频数据集中,会有力地减少耗时,提高检测速度。

本文提出了在关键帧层次上进行特征提取和匹配之前,首先使用一种新的特征代表整个镜头随时间变化的时序特征,预先滤除不同视频,以减少与不相同视频的比较,节省检测时间,再联合全局特征与局部特征进行一步步的拷贝检测。通过对实验室的监控视频进行验证实验,验证了对于给定的查询视频,可以得到与该查询视频重复或近重复的关键帧镜头及相应的关键帧,即准确检测出与查询视频重复或近重复的视频。本文方法有一定的有用性和准确性,尤其对于运动场景和景物变化较大的监控视频检测效果更好。未来的研究工作主要有:将本方法应用到大规模网络视频集的近重复检测中,使用合适的索引结构和检索方案,与其他近重复检测方法进行比较,以获得更好的检测效率和准确率;优化阈值设置方案,以能够自适应地对不同内容的查询视

频进行特征比较,而不必每次设置不同的阈值。

## 参考文献

- [1] Shang L F, Yang L J, Wang F, et al. Real-time large scale near-duplicate web video retrieval [C]. Proceedings of ACM International Conference on Multimedia (MM), 2010: 531-540.
- [2] Lu L, Wei L, Xian S H, et al. Video histogram: a novel signature for efficient Web video duplicate detection [J]. Lecture Notes in Computer Science, 2007: 94-103.
- [3] Yuan J, Duan L Y, RANGANATH S, et al. Fast and robust short video clip search for copy detection [C]. Proceedings of Pacific-Rim Conference on Multimedia (PCM), 2004: 479-488.
- [4] Wu X, NGO C W, ALEXANDER G H, et al. Real-time near-duplicate elimination for Web video search with content and context[C]. IEEE Transactions on Multimedia, 2009, 11 (2): 196-207.
- [5] LIENHART R. Comparison of automatic shot boundary detection algorithms[C]. SPIE, 1999, 3656:290-301.
- [6] BAY H, TUYTELAARS T, VAN G L. SURF: speeded up robust features [J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.

(收稿日期:2013-03-15)

## 作者简介:

郭丁云,女,1989年生,硕士研究生,主要研究方向:视频图像处理、近重复视频检测。

杨艳芳,女,1970年生,副教授,主要研究方向:数学物理方法、数字图像处理。

朱俊俊,女,1988年生,硕士研究生,主要研究方向:数字图像分析与处理、代表性图像选取。