

基于行列块编码的视频查询方法

张洪德

刘雨 唐波

(重庆通信学院通信理论教研室, 重庆 400035) (国防科技大学电子科学与工程学院, 长沙 410073)

摘要 基于内容的视频分析方法是多媒体检索领域内最近兴起的研究课题,多数研究集中在视频序列结构化、索引、识别和浏览等技术上,而对用户常用的视频查询却研究较少。为此讨论了基于“窗口对”查询方法存在的不足,并根据视频编码制式的特点、视频本身特性和常用的视频分析方法,提出了一种可处理多种视频格式的方法。该法基于视频帧行列块编码,采用了一种逐帧匹配的流式查询方法。该方法可用少量的字节来表征视频帧的颜色和位置信息,其占用内存少,简单易得,计算量小,而且对于移位和亮度的变化均不敏感,使得查全率和准确率大为提高,取得了较为理想的实验结果。

关键词 基于内容的视频分析 DC 图象 视频相似性 样本查询

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 1006-8961(2002)12-1274-05

A New Video Inquiry Technology Based on Row-column Coding

ZHANG Hong-de

(Chongqing Communication College, Communication Theory Laboratory, Chongqing 400035)

LIU Yu, TNAG Bo

(National University of Defense Technology, School of Electronic Science and Engineering, Changsha 410073)

Abstract The technology of content-based video analysis is a new research task in the multimedia retrieval field, but the most research is centralized in structuring video sequence, indexing recognizing and browsing, and video inquiry is seldom focused. This paper analyses the shortcoming of the “window pairs” technology proposed by Nevenka. The way of selecting window pairs is arbitrary in that method, so it is inappropriate sometimes. This paper suggests a new method about row and column coding based on the video coding characteristics, video peculiarities and normal technology of video analysis. The character this method used describes the color and position information of video frame. Most video about different formats could be processed using the frame-by-frame matching and streaming inquiry technology by this method. This method needs a few bytes to figure a picture, occupying fewer memory, and using less computing. The experiment improved that this technology is robust to frame shift and luminance variable, achieved better experiment results that the Recall and Precision are highly improved.

Keywords Content-based video analysis, DC image, Video similarity, QBE

0 前言

对于视频文档而言,因其数据量大,内容丰富多样,仅用传统的基于关键字的查询和检索已远远不能适应实际的应用要求。近几年来,基于视频内容的分析技术已经引起了国内外学者的研究兴趣。这种

技术主要是利用视频内部(本身)的信息进行有效的存储、浏览、查询及检索等操作,其中的视频分析技术有:视频结构构造^[1,2],内容索引和识别,视频注释和浏览等技术。从用户的观点来看,最重要的是基于视频段内容的有效检索,常用的是根据用户提交的视频样本进行视频数据库的匹配查询,即基于样本的视频查询(QBE)。

基于内容的视频查询,其基本思想是首先将视频样本进行结构化,并构造结构化后的视频特征(含镜头和关键帧特征),然后利用得到的特征进行视频特征库查询. Nevenka 提出了一种基于“窗口对”的方法^[3],其原理是首先选取合适的窗口对(window pairs)如图 1 所示,对于每个窗口对利用其窗口内块的 DC 系数和运动矢量得到一个 128bit 的特征(signatures),然后对整个序列求其特征,得到一个特征序列(signature sequence),再根据相似度从特征库中进行匹配查询. 对于这种方法虽取得了较好的实验效果,但窗口对选取的随机性,使得“查全率”和“准确率”都大为降低,而且该算法是基于视频结构化的,查询中未充分利用视频时间上的跨度性,所以在镜头相似、各镜头内部长度被抽取的情况下会误检.

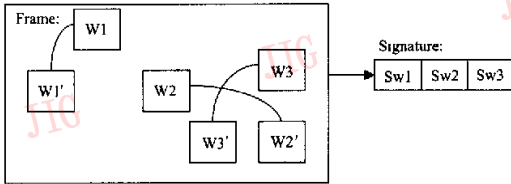


图 1 视频帧中的窗口对示例

本文的方法将用户提交的视频段看作是一个帧序列,首先对视频帧进行行列编码,得到视频序列编码,然后根据相似度量进行逐次匹配的流式查询,最后返回查询结果. 其中所用特征充分利用了视频本身的编码信息和常用的视频分析方法,可处理常用制式的视频,行列全编码特征包含了颜色和相对位置信息,此外,匹配度量中还利用了视频段固有的特征,即时间上的顺序性和跨度性,符合人的视频理解,实验中,该方法比现有方法取得了更好的查全率和准确率.

1 常用视频格式

目前比较常用的视频格式主要有 AVI、H261、MJPEG 和 MPEG 等. 尽管各种格式均不相同,压缩比也各有千秋,但其共同的特点就是可以直接提取部分图象信息. 对于 AVI 而言,虽然有多种压缩格式,但用 Windows 本身提供的 SDK,即可直接得到图象本身的像素值;对于 MJPEG 压缩格式而言,虽然不能直接得到各像素值,但可利用 DCT 的直流系数得到缩微图象(DC 图象);对于 H261 和 MPEG

而言,不但可利用 DCT 的直流系数得到缩微图象,而且还可直接提取其运动信息^[4]. 其中 I 帧中的 DC 值很简单,是块中系数的平均值,B、P 帧则用到运动矢量,主要公式如下

$$DC(P_{ret})_{00} = \sum_{j=1}^4 \left(\sum_{m=0}^7 \sum_{l=0}^7 W_{ml} ((DCT(P_i))_{ml}) \right) \quad (1)$$

其中, $W_{ml} = (DCT(S_{11}))_{0m} \times (DCT(S_{12}))_{l0}$, S_{11} 为形如 $\begin{bmatrix} 0 & 0 \\ I_{h_1} & 0 \end{bmatrix}$ 或 $\begin{bmatrix} 0 & I_{h_1} \\ 0 & 0 \end{bmatrix}$ 的位置矩阵, S_{12} 为形如 $\begin{bmatrix} 0 & 0 \\ I_{w_1} & 0 \end{bmatrix}$ 或 $\begin{bmatrix} 0 & I_{w_1} \\ 0 & 0 \end{bmatrix}$ 的位置矩阵.

直接对提取的空间缩微图象(spatially reduced images)进行分析的方法,对于有效镜头的检测、快速镜头的浏览和编辑以及对视频进行时间上的处理以进行视频表达和内容注释都是有效的,而且计算量小,对视频处理而言不丢失有效信息. 运动信息可从压缩视频流中直接提取,用以表示视频的运动信息,可辅助缩微图象进行视频的分析.

2 视频查询系统

视频查询系统(图 2)是多媒体检索的重要组成部分,按照流程可分为以下几个模块:

切变检测模块是视频结构化的第 1 步,也是整个视频分析系统的关键部分,其主要任务是将视频分为内容相同或相似的镜头或镜头组合;接下来是关键帧及其特征提取,由于一个镜头内,往往有数十个甚至上百个视频帧,需处理的数据量太大,因此可以选取最能代表镜头的关键帧,将对镜头的处理转化为对关键帧的处理,这样既缩小了需要处理的数据量,又不丢失有效信息,而提取关键帧的特征,为用户提供了从不同的角度进行视频浏览或检索的依据. 提取特征后,即可将其放入视频数据库中供用户使用.

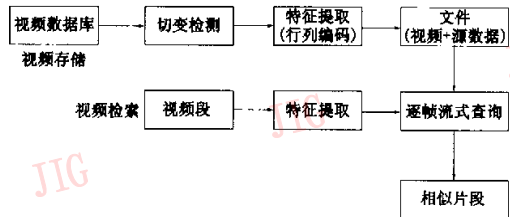


图 2 视频查询系统示意图

当用户提交视频样本时,进行以上类似处理,最后得到样本特征,从而可以在视频库中,利用相似度量进行视频查询。

3 基于 DC 和 M 的行列块编码

在视频段检测中,用得最多的是像素信息和直方图特征^[5],其中,前者有效地利用了位置信息,但对运动和噪声过于敏感;而后者则相反,对运动和噪声不是很敏感,但没有充分利用位置信息。因此许多人提出了基于块的检测方法,用得最多的是首先对视频帧进行分块处理,然后对后续帧进行块的匹配,最后计算各块变化是否超过某设定门限,通过统计超过门限的块个数,以确定是否发生切变。这样既可利用颜色信息也可利用其位置信息。Nevenka 提出的基于“窗口对”的方法也结合利用了这两者信息,窗口对的选取是任意的,也就是说,位置信息是随意的,其优点是可以针对视频帧内特定的领域进行查询,但视频库中的视频内容有其通用性,不适合此法,而且有时还可能由于窗口对选取不当而造成漏检。

基于以上的思想提出了一种基于 DC 和 M 信息的行列块编码方法,其中 DC 是直流系数, M 是运动矢量信息。对于 AVI 视频, DC 可取平均得到, M 为 0;对于 JPEG 视频, DC 可从流中提取得到, M 也为 0;对于 MPEG, DC 和 M 均可从流中得到。编码依据当前块和相邻块的距离,这样可以利用位置信息。下面仅以基于 DC 的编码为例,予以说明。

首先将图象分为 $M \times N$ 个块,然后对每个块利用块内的各 DC,求得代表该块的 DC 值,即

$$DC_{i,j} = f(DC_{i,j}^1, DC_{i,j}^2, \dots, DC_{i,j}^m) \quad (2)$$

$$i = 1, 2, \dots, M; j = 1, 2, \dots, N$$

$$f = \frac{1}{n_{ij}} \sum_{m=1}^{n_{ij}} \lambda_m DC_{i,j}^m$$

λ_m 为各块的加权值; n_{ij} 为第 (i, j) 个块内的 DC 个数。

然后对帧进行编码,编码后行编码为 LB, 共 $(M-1) \times N$ bits, 列编码为 CB, 共 $(N-1) \times M$ bits。这样只需保存 $(M-1) \times N + (N-1) \times M$ bits 的编码帧即可。

$$LB_{i,j} = f(DC_{i,j}, DC_{i+1,j}) \quad i = 1, 2, \dots, M-1; j = 1, 2, \dots, N$$

$$CB_{i,j} = f(DC_{i,j}, DC_{i,j+1}) \quad i = 1, 2, \dots, M; j = 1, 2, \dots, N-1 \quad (3)$$

其中, f 可取经过阈值检测的二值运算, 其值取 0 或 1。

如果 M 和 N 均取为 8, 则各行和各列编码仅需用 1 Byte 来保存, 共需 $(7+7)$ Bytes; 若取为 16, 则各行和各列编码需 2 Bytes 来保存, 共需 $(7 \times 2 + 7 \times 2)$ Bytes。该算法中, 编码利用的信息可以从原始数据中直接获得(参见第 2 节), 不需要复杂的提取算法, 因此计算速度快。

在视频查询中, 常用查全率和准确率、漏检率和误检率来作为查询的性能评价标准, 这两组标准都是相互矛盾的, 必须兼顾考虑, 其中对查全率(漏检率)要求较高, 因此查询的目标应该是在保证查全率(漏检率)的基础之上, 提高准确率(降低误检率)。

上面提到的编码方式包含了视频帧的直流信息和运动矢量信息, 其反映了视频帧的绝大部分信息量, 其中直流信息综合利用了视频帧的颜色和位置信息。从编码过程中可知, 编码后的数据和各块的绝对亮度无关, 而且较小的移位对编码数据影响不大(如果移位过大, 可认为其不是检测结果, 以降低误检率), 从编码方式上确保了对亮度变化和有一定移位情况的检测, 提高了查全率, 降低了漏检率。

4 基于编码的视频查询

在对视频帧进行编码后, 即可得到已编码过后的帧序列, 假设视频样本和待匹配的视频库文件为

$$X = (x_1, x_2, \dots, x_m)$$

$$Y = (y_1, y_2, \dots, y_n)$$

则视频查询就是从 Y 中找出是否含有 Y' , 使得下式成立:

$$\text{sim}(Y', X) \approx 1 \quad Y' = (y_i, y_{i+1}, \dots, y_{i+m-1}) \quad (4)$$

视频样本有不同于其他媒体的特点, 即时间上的顺序性和跨度性。这里假定相似视频的视频帧顺序一致, 而且长度一致, 因此对编码后视频帧序列采用逐次匹配(顺序性)的流式(跨度性)查询。

查询算法如下:

GetCodedVideoFromSample();

for($i=0; i < n-m+1; i++$)

{

 GetCorrespondCodedVideoFromDB();

 For($j=0; j < m; j++$)

 {

$$Dis = \text{RowDistance} \left(\sum_{i=0}^{(M-1) \times N} LB_{i,j}, \Delta LB_{i,j} \right) \parallel$$

```

ColDistance(  $\sum_{p=0}^{N-1} \sum_{q=0}^{M-1} C B_{i,p} A C B_{j+1,q}$  )
If( (dis > minDis) AND similar(ShotsLength))
    Sim += 1/m;
}
}

```

5 实验结果及结论

实验利用 MS VC 作为开发平台,因为该开发平台具有和底层结合紧密,运算速度快等特点.素材

选用了具有一定代表性的视频文件,其素材本身变化有缓有急,色度变化范围较大.一般的视频分析实验都是采用选取多种类别各异的视频作为实验素材,但因选取的主观性,不能保证视频内容的变化覆盖其变化范围,所以在下面的实验中固定一个素材,在此基础上自己构造变异的素材,比如采取步进 10 做移位、亮度增强等处理(部分结果见图 3).这样可以人为地构造覆盖某个特征(如亮度等)整个变化范围的视频素材,比以前的方法具有更好的说服力,可以更好地检验算法的优劣.

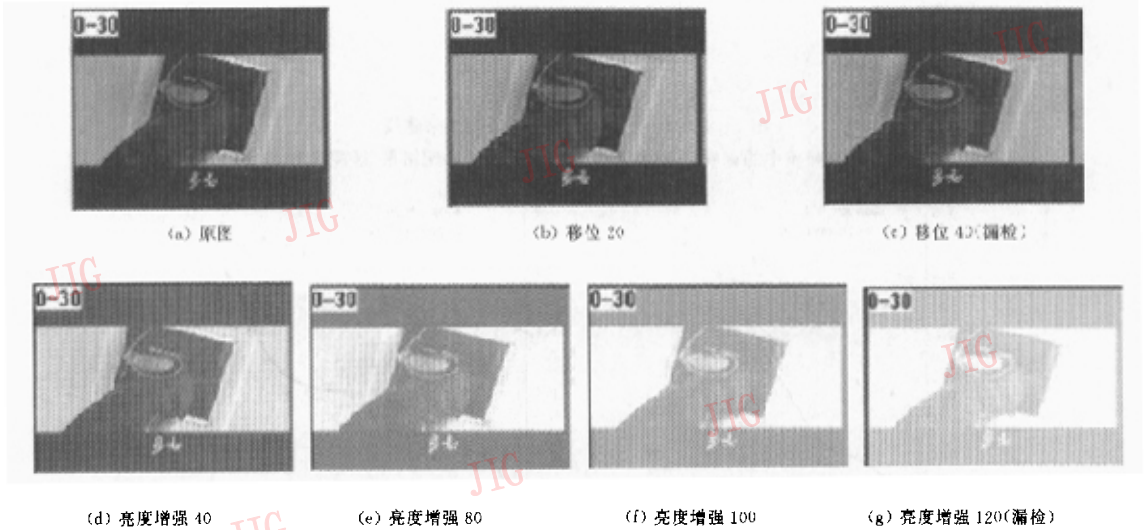


图 3 样本首帧处理后的结果

试验中,取 $M=N=8$,仅统计处理了 DC 信息.视频样本含 3 个镜头,仅选择了 3 帧关键帧进行处理,视频库文件共 845 帧.样本编码后共 $(7+7) \times 3=42\text{Bytes}$,可用 42Bytes 保存.对色彩进行了各种处理,进行流式匹配后得到结果如图 4 所示.从结果中可以明显地看到,当移位 20 时尚有显著波谷,可以检测到,但移位过大到 40 时就漏检,此时从素材上看失真已经很大;随着亮度逐渐增强,匹配结果曲线中的波谷处变得越来越不明显,当亮度增强到 120 时已极不明显,造成漏检,同样此时素材失真已非常大,可认为不是查询结果.

为得到该算法对移位和亮度变化的敏感程度,处理 5 种不同类型的素材,并分别进行移位和亮度变化处理,然后和原始素材进行匹配计算,得到如图 5 所示的匹配结果.对于移位处理,当移位不是 3 的

整数倍时,余数表示视频帧色调同时发生变化,此时的匹配结果有所偏差;只有当移位值是 3 的整数倍时,才仅是对帧的移位处理.所以该曲线是周期性的、单调性变化(图中仅取值 3 的倍数,故呈单调变化).而对于亮度的变化则是亮度偏差越大,越不匹配,因此从图中可以看到,对于亮度的变化基本上呈单调趋势,当亮度变化小于 80,移位 在 0~40,且是 3 倍数时匹配效果最好,匹配值均在 0~0.4 之间.图 5 中还画出了 Nevenka 算法的性能,可以看出这种方法对这位移、亮度极为敏感,漏检率高.

试验结果表明,本文提出的视频查询方法计算量少,对视频帧的移位和亮度变化均不敏感,通用性较好,可以满足视频查询的“在保证查全率的基础上,提高准确率”的要求.但此算法在特征的选取上,尚未利用视频帧间的相关信息,这是下一步工作的研究方向.

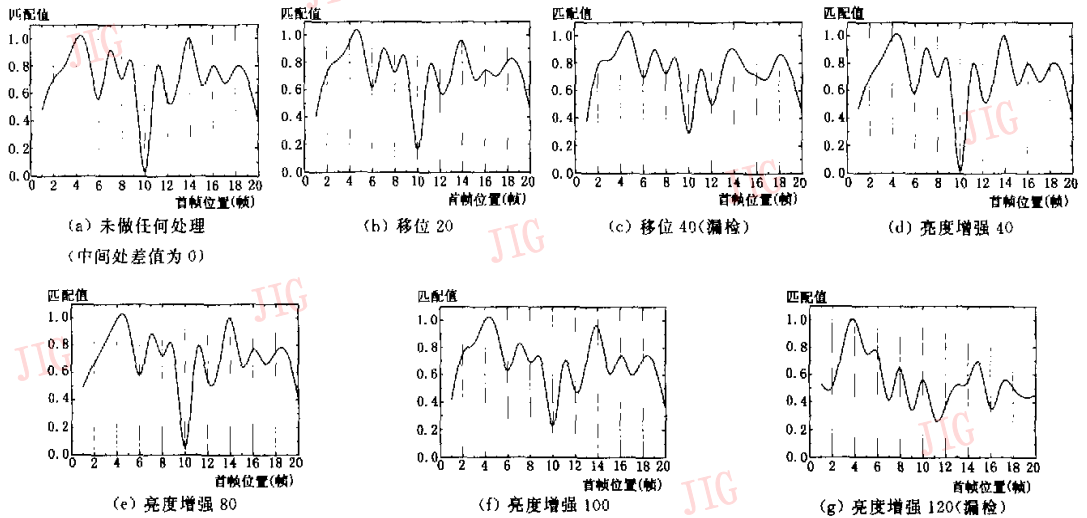


图4 各种样本处理后的流式查询结果曲线

(横坐标为匹配时视频流中的首帧位置,纵坐标为该位置处的匹配结果,匹配度等于0为最相似)

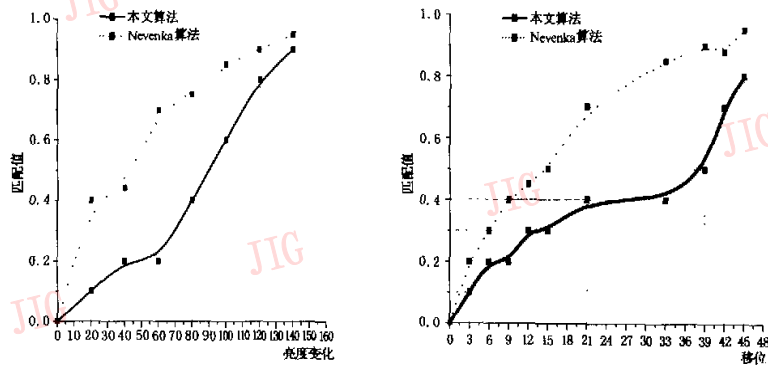


图5 亮度变化和移位对两种算法匹配结果的影响

参考文献

- 1 Zhang H J, Wu Jianhua, Zhang D. An integrated system for content-based video retrieval and browsing [J]. Pattern Recognition, 1997, 30(4): 643~657.
- 2 Zhang H J, Low C Y, Smolier. Video parsing using compressed data [A]. In: Proc. IS and T/SPIE Conf. Image and Video Processing I [C], San Jose, California, 1994: 142~149.
- 3 Nevenka Dimitrova, Mohamed Abdel-Mottaleb. Content-based video retrieval by example video clip [J]. SPIE, 1997, 3022: 59~70.
- 4 Yeo B L, Liu B. On the extraction of DC sequence from mpeg compressed video [J]. IEEE, 1995, 2: 260~263.
- 5 John S B, Lawrence A R. Comparison of video shot boundary detection techniques [J]. SPIE, 1996, 2664: 170~179.



张洪德 1977年生, 1999年获国防科技大学电子工程学院硕士学位. 主攻方向为视频图象检索、多媒体通信等.



刘雨 1963年生, 硕士生导师, 现任国防科大信息工程教研室主任、高工. 获部委级科技进步一等奖2项, 二等奖1项, 三等奖3项, 出版著作2部. 研究方向为图形图象处理、数字视频技术.



唐波 1972年生, 讲师, 1996年获国防科技大学硕士学位, 曾获部委级科技成果二等奖两项, 三等奖一项. 研究方向为数字图象处理、数字视频处理.