一种分级检索 MPEG 视频的方法*

刘 阳, 许松涛, 吴志美+

(中国科学院 软件研究所,北京 100080)

A Hierarchical Retrieval Method for MPEG Video

LIU Yang, XU Song-Tao, WU Zhi-Mei+

(Institute of Software, The Chinese Academy of Sciences, Beijing 100080, China)

+ Corresponding author: Phn: 86-10-62645407, Fax: 86-10-62645410, E-mail: wzm@isdn.iscas.ac.cn

http://www.ios.ac.cn

Received 2002-02-06; Accepted 2002-05-21

Liu Y, Xu ST, Wu ZM. A hierarchical retrieval method for MPEG video. *Journal of Software*, 2003,14(3): 675~681.

Abstract: Video retrieval is a current hot research area. Most of the past algorithms are done in pixel domain, which need many decode calculations. What is more, the same matching algorithm is used to all the video clips in the same way, which wasts many unnecessary calculations. A new hierarchical retrieval method based on example video for MPEG video is proposed: firstly the dct_dc_size field in I frames is used to locate the suspected videos quickly, then the retrieval result scope can be further reduced by analyzing the spatiotemporal distribution of motion vector in B frames using tomography method, at last DC images precise matching analysis is used to validate the retrieval result. The experimental results show that this method needs a few calculations and has higher precision ratio.

Key words: video retrieval; MPEG; hierarchical; dct_dc_size; spatiotemporal analysis; DC image

摘 要: 视频检索是当前的一个研究热点.以前的检索方法大多在像素域中进行,需要较大的解码运算量;且不加区分地对所有视频片断采用统一的匹配算法,浪费了许多不必要的计算.提出了一种基于样本的分级检索MPEG 视频的新方法:首先用 I 帧的 dct_dc_size 字段快速粗检,然后用断层摄影(tomography)法分析 B 帧运动矢量的时空分布特性以进一步缩小结果集,最后用 DC 图像的精确匹配方法验证检索结果.试验结果表明,本方法所需计算量较小,且可保证较高的检索精度.

关键词: 视频检索;MPEG;分级;dct_dc_size;时空分析;DC 图像

中图法分类号: TP391 **文献标识码**: A

随着多媒体技术的应用和计算技术的发展,有越来越多的学者投入到视频检索的研究中,并提出了有效的检索方法[1~3].视频检索不同于字符检索,在很大程度上也不同于图像检索,因为图像检索只需分析单幅图像的

* Supported by the National Grand Fundamental Research 973 Program of China under Grant No.G1998030407 (国家重点基础研究发展规划(973)); the Foundation of Beijing Science Committee of China under Grant No.H011710010123 (北京科委基金)

第一作者简介: 刘阳(1975－),男,山东莱芜人,博士生,主要研究领域为视频分析,多媒体通信.

颜色或纹理等特征,而视频检索是在视频序列库中查找特定的连续帧,还需要分析连续帧间的相关特征.

基于样本的视频检索是根据用户提交的视频样本,在数据库中查找与之相似的视频片断(clip),其基本思想是先提取视频样本的某些特征,然后根据与各视频片断比较的相似度得到检索结果.与其他视频分析,例如镜头分割和关键帧提取一样,可将检索方法分为两类:一类是在像素域中分析连续帧的颜色或对象边界等特征,以前大多数方法属于此类,进展相对较快;另一类是在压缩域中直接分析 DC 系数或光流场等特征,对这类方法的研究较少,因为无须对压缩视频完全解码,需要较少的运算量,这对大型视频数据库检索和实时检索是非常有益的.另外,当前方法不加区分地对明显不同和基本相似的片断采用统一的匹配算法,事实上大部分片断只需简单分析就可判定不相似,而对相似度较高的片断则需精确分析,但很少有方法考虑到这一点.文献[4,5]对当前的主要检索方法进行了综述.

本文提出一种基于样本的分级检索 MPEG Ⅰ或 MPEG Ⅱ(下文统称为 MPEG)视频的新方法,同时对原有的 I 帧 dct_dc_size 字段的分析算法进行了改进,用断层摄影法分析运动矢量的时空分布特性.

1 分级检索的原理和方法

1.1 分级检索方法概述

与文本数据不同,视频数据有不同的相似性评价标准,所选取的比较特征的优劣决定了算法的性能,它的主要评价标准是其计算存储代价和对不同视频的区分性.本文选取的比较特征是以 MPEG 原始码流中存在的字段为基础的,无须复杂变换,且具有较好的区分性.

在视频数据库中,与样本相似的视频仅是一小部分,不必对所有片断进行精确比较,只需分析 I 帧中表示 DC 差值编码长度的 dct_dc_size 字段就可以过滤与样本有较大差异的片断,这样的快速粗检是分级检索的第 1 级,然后用断层摄影(tomography)法分析 B 帧运动矢量的时空分布特性以进行精确分析,最后用 DC 图像精确匹配法对结果进行验证.

1.2 分析I帧的dct_dc_size字段快速粗检

MPEG 标准规定[6],每个图像组 GOP 中存在一个 I 帧,I 帧主要包含 DCT 变换后的直流系数 DC 和交流系数 AC,可独立解码.DC 值与原 8*8 矩阵元素的平均值成正比,它是视频分析的一个重要特征.MPEG 码流中存在的并不是 DC 的实际值,而是相邻块间的 DC 差值,当前块的 DC 值等于参考块的 DC 值加上当前块的 DC 差值,MPEG 标准规定了参考值为 0 时的情况,例如在每个组块(slice)开始处,其主要目的是增大压缩比例.图 1 给出宏块(*i*,*j*)内部 4 个亮度块间的参考示意图,其中数字和箭头表示参考顺序,其他字符意义将在下文给出.

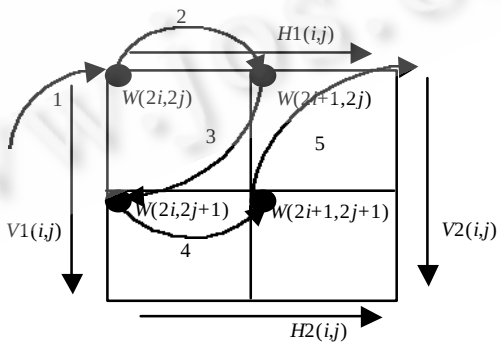

Fig.1 DC value reference mode of the four blocks in macroblock (*i*,*j*)

图 1 宏块(*i*,*j*)内部 4 个亮度块间的参考示意图

每个块的 DC 差值是与预测块间的亮度差或色度差的平均值成正比的,相似视频对应的 DC 差值也必相似.同 DC 值一样,DC 差值也是视频分析的重要特征.在 MPEG 码流中,用 dct_dc_size 字段表示 DC 差值的编码长度,差值越大,编码长度越大,相应的 dct_dc_size 字段也就表示差值的取值范围.以前对此特征研究较少,仅检索

到文献[7]用该特征进行镜头分割并得到较好效果.根据 MPEG 标准,设 diff(i)表示当 dct_dc_size=i 时 DC 差值的取值范围,当 $i\geq 1$ 时,diff(i)的取值范围为$\pm(2^{i-1},2^i-1)$.文献[7]直接将 dct_dc_size 字段作为特征值,这样会存在两方面问题:一是未考虑量化系数和差值的正负;二是未考虑 dct_dc_size 表示的是编码长度,而不是实际差值.直接作为特征值有较大误差,例如:很明显 diff(8)−diff(7)>diff(3)−diff(2),为克服以上不足,定义如下的加权特征值:

$$W(x,y)=\begin{cases}0 & \text{当}i=0;\\ Q(x,y)\cdot\mathrm{sign}(\mathrm{diff}(x,y))\,\mathrm{int}\left(2^{i-1}+2^i-1/2\right) & \text{当}i\geq 1.\end{cases} \tag{1}$$

式(1)中 $W(x,y)$表示块(x,y)的加权特征值,i 为块(x,y)的 dct_dc_size 字段值,$Q(x,y)$表示其量化系数,sign(diff(x,y))根据实际差值数据的最高有效位 MSB 取正号或负号,它位于 dct_dc_size 字段 VLC 编码的后面,int()表示取整运算.

对图 1 所示的宏块(i,j),由式(1)进一步定义以下比文献[7]更合理的特征值:

块$(2i+1,2j)$与块$(2i,2j)$间的水平加权特征值:

$$H1(i,j)=W(2i+1,2j); \tag{2}$$

块$(2i+1,2j+1)$与块$(2i,2j+1)$间的水平加权特征值:

$$H2(i,j)=W(2i+1,2j+1); \tag{3}$$

块$(2i,2j+1)$与块$(2i,2j)$间的垂直加权特征值:

$$V1(i,j)=W(2i+1,2j)+W(2i,2j+1); \tag{4}$$

块$(2i+1,2j+1)$与块$(2i+1,2j)$间的垂直加权特征值:

$$V2(i,j)=W(2i,2j+1)+W(2i+1,2j+1); \tag{5}$$

宏块(i,j)第 1 个块和宏块$(i-1,j)$最后块间的加权特征值:

$$G(i,j)=W(2i,2j); \tag{6}$$

宏块(i,j)的总加权特征值:

$$T(i,j)=H1(i,j)+H2(i,j)+V1(i,j)+V2(i,j)+G(i,j). \tag{7}$$

在本级快速粗检中,我们利用以上的加权特征值,计算样本和被比较片断的 I 帧的相似度.首先提取样本中每个 I 帧的加权特征值,然后与数据库的各片断的相应 I 帧相比较.对每对 I 帧的帧间差定义如下:

$$D(k)=\sum_{j=0}^{N-1}\sum_{i=0}^{M-1}\left|Tr(i,j)-Tc(i,j)\right| \Big/ \sum_{j=0}^{N-1}\sum_{i=0}^{M-1}\left|Tr(i,j)\right|. \tag{8}$$

式(8)中 $D(k)$表示第 k 对 I 帧的帧间差,$Tr(i,j)$和 $Tc(i,j)$分别表示样本和被比较片断的第 k 个 I 帧的宏块(i,j)的加权特征值,假设视频分辨率为 $16M\times 16N$.

接下来,需计算样本和被比较片断中所有对应 I 帧间的平均帧间差,并定义如下不等式判断其相似性:

$$\frac{\sum_{k=0}^{S-1}D(k)}{S}\leq\varsigma_1, \tag{9}$$

式(9)中 S 表示样本包含的 I 帧数量. h_1 是判断是否相似的阈值,调整其大小可控制检索精度.因本级检测目的是过滤明显不匹配的片断,h_1 设置原则是尽可能保留基本相似的片断,一般地,当分辨率为 352×288 时,h_1 应小于 2,具体实现时可灵活设置,例如可预先统计样本与所有片断的帧间差,不具体设置h_1 值,然后按平均帧间差由小到大选取一定数量的片断.

不直接用 DC 值作为检索特征值,是因为本级检索是快速粗检,目的是为初步排除明显不相似的视频片断,使用 DC 值会使比较过于精细.另一方面,用 dct_dc_size 字段可降低解码运算量.

1.3 分析B帧的运动矢量时空分布特性缩小检索范围

上面的粗检只能保证结果与样本粗略相似,因为在 MPEG 码流中 I 帧只占较小比例,用小部分数据分析整体数据的相似性必然会降低准确性.事实上,上文所述改进的 I 帧检索的基本原理也可用于 JPEG 格式的图像检索[8],连续帧随时间变化的相关性还未使用.另外,检索算法需根据样本精确定位视频结果的起止点,仅分析 I 帧显然是不足的,还需进一步分析其他类型帧的相关特性.

在 MPEG 码流中,B 帧和 P 帧主要根据相邻预测帧间的相关性采用预测编码,包含的主要数据有:宏块编码类型、运动矢量和预测误差.因预测误差极不稳定,参考价值较小,很少有算法将其作为检索特征.运动矢量表示向前或向后预测的参考宏块的相对位置,它反映了连续帧间对应宏块的相关性,分析其特性可以得到一段视频内部的运动结构,运动矢量分析在视频分析中得到广泛应用.参考其他检索方法,在保证一定精度的前提下,为减少计算量,在本级检测中仅分析 B 帧的运动矢量.

具体的情况我们需要分析连续 B 帧运动矢量的空间和时间分布特性来定.空间分布特性是指单帧内各宏块的运动矢量在本帧内的分布特性.时间分布特性是指一段时间内连续帧各宏块的运动矢量随时间变化的分布特性.文献[2]仅利用了运动矢量的空间分布特性,忽略了时间分布特性.为分析方便,我们将连续帧的时空图像序列组成图 2 所示的立方体,x-y 平面表示单帧的运动矢量的空间分布图,坐标大小由视频分辨率决定,x-t 和 y-t 平面表示连续帧的运动矢量随时间变化的分布图,坐标大小由视频分辨率和帧数量决定.

我们用断层摄影法分析该三维空间,计算机断层摄影 CT(computed tomography)技术已成功应用在医学疾病检测上,文献[9]用该方法进行镜头分割.断层摄影法首先将三维空间中某点的 x 或 y 或 z 坐标固定,然后再分析其他两维坐标平面的分布特性.图 2 给出一个断层摄影法的示意图.例如 x 轴的投影可表示如下:

$$f_{tomo}(y,t\,|\,x^*) = f(x^*,y,t)\,, \tag{10}$$

式(10)中, $f_{tomo}(y,t\,|\,x^*)$ 表示 x 为常数时 $f(x,y,t)$的 y-t 投影函数,代表 x 坐标固定时 y 轴上的运动矢量随时间变化的过程,我们也可将空间特性形成的 x-y 平面看做是 t 为常数时的断层摄影.这样,运动矢量的空间分布特性可直接在 x-y 平面上分析,运动矢量的时间特性用断层摄影法在 x-t 和 y-t 平面上分析.运动矢量的时空分布特性可采用统一的二维平面分析方法,以下论述的二维平面上的运动矢量特性的分析方法以 x-y 平面为例,同样也适用于 x-t 和 y-t 平面.

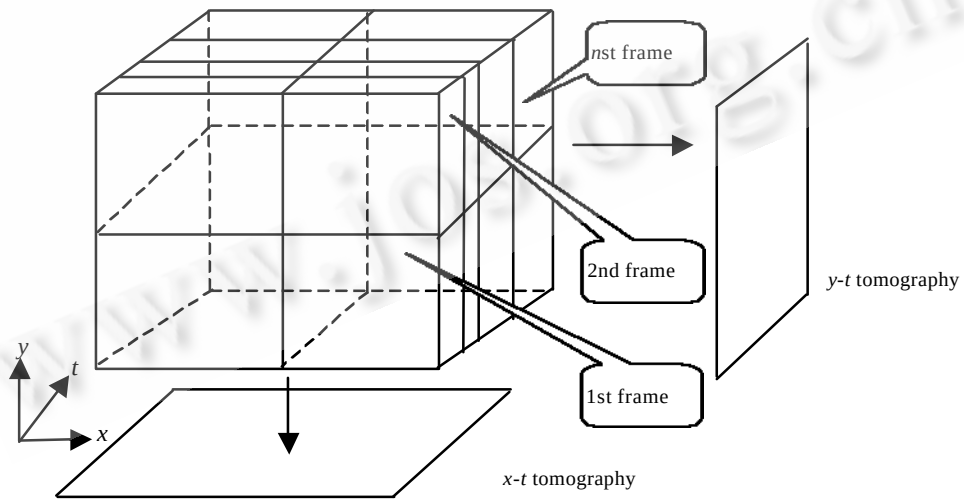

Fig.2 Tomography method for spatial and temporal distribution feature of motion vector

图 2 运动矢量的时空分布特性的断层摄影法示意图

视频分析中可用直方图、方差分析、边界提取等方法进行特征比较,文献[10]通过试验比较指出,直方图分析是最为有效的方法.直方图通过比较元素分布到预先设定的各离散小区间的分布差异性来判定相似性,其缺

点是丢失了元素的空间信息,两个不同的视频可能有相同直方图.为了在一定程度上克服此缺陷,我们将坐标平面平分为 $U\times V$ 个子区域,在每个子区域中分别进行直方图分析,这样可保留部分的空间信息.具体的 UV 取值可综合视频分辨率和计算复杂度等因素综合考虑,例如,一般地,当视频分辨率为 352×288 时,可分解为 2×2 个子区域.

运动矢量是二维向量,在 MPEG 码流中分别以其横坐标分量和纵坐标分量表示,可直接分别对它们进行直方图分析,避免用大小和方向分析而引入的额外的转换计算量.结合预测编码的搜索范围,在具体实现中,可将分析区间限制在[−15,15]之间,对极少数此范围之外的元素采取饱和化(saturate)处理.另外,对内部编码宏块单独统计.

在对样本和被比较片断的某对 B 帧子区域(u,v)内的运动矢量进行分析时,首先对其横坐标分量的直方图距离定义如下:

$$C_x(u,v)=\frac{1}{L}\sum_{z=1}^{L}\left|H_{rx}(z)-H_{cx}(z)\right|, \tag{11}$$

其中 $C_x(u,v)$表示横坐标分量的直方图距离,L 表示对横坐标分量划分的区间数,$H_{rx}(z)$和 $H_{cx}(z)$分别表示样本和被比较片断在该子区域内横坐标分量位于第 z 个区间的数量.采用同样的方法,我们可以得到纵坐标分量的直方图距离.

对该子区域内运动矢量的横坐标分量和纵坐标分量的直方图距离之和定义如下:

$$C(u,v)=\acute{a}C_x(u,v)+(1-\acute{a})C_y(u,v), \tag{12}$$

其中 $C(u,v)$表示在子区域(u,v)内运动矢量的直方图距离,$C_x(u,v)$和 $C_y(u,v)$分别表示式(11)定义的横坐标分量和纵坐标分量的直方图距离,a表示横坐标分量的加权值,$1-a$ 表示纵坐标分量的加权值,一般地,取a=0.5.

对分解为 $U\times V$ 个子区域的每对 B 帧的运动矢量的帧间差定义如下:

$$G_t(t)=\frac{1}{U\times V}\sum_{v=0}^{V-1}\sum_{u=0}^{U-1}C(u,v), \tag{13}$$

其中 $G_t(t)$表示如图 2 所示的三维空间中时间坐标为常数 t 时 x-y 平面上的帧间差.

用上文所述的断层摄影法,同样可得到 x-t 和 y-t 平面上的所有 B 帧运动矢量的时间分布的帧间差.对连续 B 帧运动矢量的空间和时间分布的差值定义如下:

$$G=\sum_{t=0}^{R-1}G_t(t)+\sum_{x=0}^{M-1}G_x(x)+\sum_{y=0}^{N-1}G_y(y), \tag{14}$$

其中 G 表示连续 R 个 B 帧的时空分布总差值,$G_t(t)$已在式(13)中给出其定义,$G_x(x)$表示如图 2 所示的三维空间中横坐标为常数 x 时 y-t 平面上的帧间差,$G_y(y)$表示纵坐标为常数 y 时 x-t 平面上的帧间差,视频的分辨率为 $16M\times16N$.

定义如下不等式以判断样本与被比较片断是否相似:

$$\frac{G}{3R}\le\varsigma_2, \tag{15}$$

其中 h_2 是判断是否相似的阈值,调整其大小可控制检索精度,一般地,当视频分辨率为 352×288 时,h_2 应小于 10,它也可采用与h_1一样的灵活设置方法.

1.4 利用DC图像的精确匹配法验证检索结果

满足以上两级检测条件的视频片断基本保证与样本相似,为进一步验证检索结果,可采用比前两级检测更严格的检测方法,对 DC 图像进行精确匹配分析.DC 图像是由压缩域的 DC 系数组成的图像,每个块用其 DC 系数近似代替.可将 DC 图像看做是原始图像的近似缩影,验证其相似性可得到原始视频片断的相似性.I 帧的 DC 系数可直接在视频流中提取,B 帧和 P 帧的 DC 系数可用文献[11]的算法近似求得,这样形成的 DC 图像与帧类型无关,验证过程不必区分帧类型.

DC 图像的相似性可通过计算由样本和被比较片断中 DC 图像的系数组成的矩阵间的距离来实现,我们用欧式距离(euclidean distance)计算其差异性,该法对噪声有较好的鲁棒性.对 DC 图像的帧间差定义如下:

$$B(k)=\sqrt{\sum_{q=0}^{N-1}\sum_{p=0}^{M-1}(E_r(p,q)-E_c(p,q))^2}, \tag{16}$$

其中 $B(k)$表示第 k 对 DC 图像的帧间差, $E_r(p,q)$ 和 $E_c(p,q)$ 分别表示样本和被比较片断的第 k 对 DC 图像点(p,q)的 DC 值,DC 图像的分辨率为 $M\times N$.

定义如下不等式以判断样本与被比较片断是否相似:

$$\frac{1}{T}\sum_{k=0}^{T-1}B(k)\le\varsigma_3, \tag{17}$$

其中 T 表示样本包含的帧总数量,h_3是判断是否相似的阈值.因为这是检测的最后一级,应定义h_3比较小以保证检测的精确性.同时规定,当不满足以上不等式时,可否定前面的检测结论.另外,当第 1 级检测中式(9)的比较值等于 0 时,可直接跳到这一级.

2 试 验

为验证本方法的性能,我们参考其他视频检索方法试验的原则和方式,从 ftp://202.38.126.48 下载了 62 个足球射门的视频片断,它们均采用 MPEG-1 标准压缩,包含 12 106 帧,分辨率为 352×288.用如图 3 所示的视频片断作为样本,该样本包括 30 帧,其中 I 帧有 3 个、B 帧有 18 个.

Fig.3 Example clips of the experimentation

图 3 试验的样本视频

按上文所述的方法对各个视频片断进行分析,因仅需很少的 MPEG 解码和简单的比较运算,比 MPEG 实时解码的计算量要少,具有 MPEG 软解压能力的机器完全可实时检测.为了更直观地给出检索结果,试验中没有设置各级检测中使用的阈值 $h_1h_2h_3$ 的具体取值,而按各级检索中计算的帧间差得到样本与各个片断的相似度.图 4(a)~图 4(e)从左到右给出了按相似度由大到小排列的前 4 个检索结果.

(a) (b) (c) (d) (e)

Fig.4 Retrieval results of the experimentation from left to right

图 4 由左到右排列的检索结果

3 结 论

本文提出了一种分级检索 MPEG 视频的新方法.整个检索过程在压缩域中进行,因不必完全解码,因此所需计算量较小,同时与其他检索算法相比,不用预先提取关键帧,减少了预处理过程,可实时检索.另外,经过多级比较,本方法可保证有较高的检索精度.在第 1 级检索中改进了原来分析 I 帧 dct_dc_size 字段的算法,在第 2 级检索中使用断层摄影法分析运动矢量的时空分布特性.阈值的自适应问题是我们下一步要进行研究的工作.

References:

[1] Zhang HJ, Low CY, Smoliar SW, Wu JH. Video parsing retrieval and browsing: an integrated and content-based solution. In: Adelson EH, Bergen JR, eds. Proceedings of the 3rd ACM International Conference on Multimedia' 95. San Francisco: ACM Press, 1995. 15~24.

[2] Jain AK, Vailaya A, Xiong W. Query by video clip. Multimedia Systems: Special Issue on Video Libraries, 1999,7(5):369~384.

[3] Zhuang YT, Liu XM, Wu Y, Pan YH. A new approach to retrieve video by example video clip. Chinese Journal of Computers, 2000,23(3):300~305 (in Chinese with English Abstract).

[4] Atsuo Y, Tadao I. Survey on content-based retrieval for multimedia databases. IEEE Transactions on Knowledge and Data Engineering, 1999,11(1):81~93.

[5] Ngo CW, Pong TC, Chin RT. A survey of video parsing and image indexing techniques in compressed domain. In: Vass J, Zhuang S, eds. Proceedings of the Symposium on Image, Speech, Signal Processing, and Robotics' 98. Hong Kong, 1998,1:231~236.

[6] Zhong YZ, Qiao BX, Qi W. General International Coding Standard for Motive Pictures and Audio——MPEG-2. Beijing: Tsinghua University Press, 1997. 142~319 (in Chinese).

[7] Won CS, Park DK, Yoo SJ. Extracting image features from MPEG-2 compressed stream. In: Ishwar KS, Ramesh J, eds. SPIE Proceedings of the Storage and Retrieval for Image and Video Databases VI. San Jose, CA: SPIE Press, 1998,3312:426~435.

[8] Liu Y, Wu ZM. A new JPEG image retrieval method in compressed domain. In: Proceedings of the 14th IEEE International Conference on Application-Specific System, Architectures and Processors. Hague, 2003.

[9] Akutsu A, Tonomura Y. Video tomography: an efficient method for camerawork extraction and motion analysis. In: Wayne N, Ramesh J, eds. Proceedings of the 4th ACM International Conference on Multimedia' 94. San Francisco: ACM Press, 1994. 349~356.

[10] Boreczky JS, Rowe LA. Comparison of video shot boundary detection techniques. In: Sethi K, Jain RC, Ishwar KS, eds. Proceedings of the SPIE Conference on Storage and Retrieval for Still Images and Video Databases IV. San Jose, CA: SPIE Press, 1996,2664:170~179.

[11] Yeo BL, Liu B. On the extraction of DC sequence from MPEG compressed video. In: Depommier R, Hsu A, eds. Proceedings of the International Conference on Image Processing' 95. Washington, DC: IEEE Press, 1995. 260~263.

附中文参考文献:

[3] 庄越挺,刘小明,吴翌,潘云鹤.通过例子视频进行视频检索的新方法.计算机学报,2000,23(3):300～305.

[6] 钟玉琢,乔秉新,祁卫.运动图象及其伴音通用编码国际标准——MPEG-2.北京:清华大学出版社,1997.142～319.