

基于二叉树和随机邻域嵌入的视频指纹算法

聂秀山¹, 刘 琚^{2,3}, 孙建德^{2,3}

(1. 山东财经大学计算机科学与技术学院, 山东济南 250014; 2. 山东大学信息科学与工程学院, 山东济南 250100;
3. 海信数字多媒体技术国家重点实验室, 山东青岛 266067)

摘 要: 针对互联网视频认证问题, 提出了一种基于二叉树和随机邻域嵌入的视频指纹算法, 该算法首先利用二叉树和归一分割理论选取视频代表帧, 然后以代表帧亮度的离散余弦变换系数作为视频在高维空间的特征, 利用随机邻域嵌入方法把视频高维特征投影到三维空间, 以视频在三维空间中相邻投影点距离序列的均值和方差构造匹配密钥, 以距离序列向量二值量化构造视频指纹; 对于视频指纹匹配, 首先利用匹配密钥进行初级匹配, 缩小匹配范围, 然后在候选指纹组内进一步匹配, 从而达到视频认证的目的.

关键词: 视频指纹; 视频拷贝检测; 随机邻域嵌入; 二叉树

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112 (2013) 12-2409-07

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2013.12.014

A Video Fingerprinting Algorithm Based on Binary Tree and Stochastic Neighbor Embedding

NIE Xiu-shan¹, LIU Ju^{2,3}, SUN Jian-de^{2,3}

(1. School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, Shandong 250014, China;
2. School of Information Science and Engineering, Shandong University, Jinan, Shandong 250100, China;
3. Hisense State Key Laboratory of Digital-Media Technology, Qingdao, Shandong 266061, China)

Abstract: In order to identify video copies on the Internet, a video fingerprinting algorithm based on binary tree and stochastic neighbor embedding is proposed in this paper. In this scheme, representative frames are selected based on binary tree and normal cut, and the Discrete Cosine Transformation (DCT) coefficients of luminance of representative frames are taken as the high dimensional features of the video. The features are mapped into three-dimensional space using stochastic neighbor embedding. A matching key is generated based on the mean and variance of distance vector between adjacent points in this three-dimensional space, and the video fingerprint is generated by binarizing. Moreover, during video fingerprint matching, the matching key is used in the first-stage matching to reduce the search range, and then a further matching is carried out in the candidate fingerprint to identify video copies.

Key words: video fingerprinting; video identification; copy detection; stochastic neighbor embedding; binary tree

1 引言

随着计算机与信息技术的发展, 互联网已经成为日常生活中不可或缺的一部分, 特别是随着多媒体技术的迅速发展, 互联网视频网站越来越多, 视频内容变的越来越丰富, 但随之而来的网络信息安全问题也变的日渐突出. 网络用户可以随意的下载、编辑视频并再次上传, 由此网络中出现大量“拷贝视频”, 这大大降低了视频检索的效率. 同时, 由于互联网上对于海量视频的管理缺乏规划和统一性, 常常引起一些诸如知识产权等纠纷,

盗版和非法下载的出现也危害了版权所有者的利益, 挫伤了创新的积极性. 去除网络中的“拷贝视频”以及打击盗版进行有效的版权管理从本质上来说都是视频内容认证和识别问题. 基于内容的视频指纹技术正是解决上述问题的主要方法. 基于内容的视频指纹, 又被称为视频哈希, 与传统的数字视频水印^[1,2]不同, 它不需要在视频发布或传播之前, 人工地在视频对象中嵌入附加信息, 因此, 视频指纹对于尚未发布和已经发布的视频都有效. 利用视频指纹技术可以在视频检索中搜索到多余的视频拷贝, 并将其屏蔽, 有助于规范数字视频版权、

传播等的管理;因此,在个人视频制作与网络传播日益流行的情况下,视频指纹具有重要的理论价值和应用价值,已经成为了多媒体信息处理领域的研究热点.

对视频指纹的研究,近年来国内外涌现出一批优秀的研究成果,如 Joly^[3]等利用帧内角点局部特征构造签名,对视频中全局运动强度变化较大的关键帧提取特征作为指纹向量来实现视频认证. Lee 等^[4]将视频帧分割成固定数量不重叠的若干块,计算各块的 CGO (Centroids of Gradient Orientation)生成检测指纹.同时,一些研究者提出利用鲁棒哈希作为视频指纹,以鲁棒视频哈希的相似性衡量作为视频认证的主要方式. Zhou^[5]等利用视频帧分块的偏微分特征作为视频哈希,通过视频哈希的匹配来实现视频认证; B. Coskun 等^[6]通过三维 DCT 变换来获取视频在时空域上的频率特征构成鲁棒哈希. Esmacili 等^[7]提出了一种基于时域信息帧的三维 DCT 的视频指纹算法. 国内的研究者诸如李锦涛教授^[8,9]、黄庆明教授^[10]、戴跃伟教授^[11,12]、刘璐教授^[13,14]等带领的研究组也取得很多重要的研究成果.

视频片段由一个或多个场景组成,每个场景由内容连续或相同的若干帧组成. 如何从众多的视频帧中选取关键帧来提取视频指纹是视频指纹算法首要考虑的问题. 目前的关键帧提取方法大多依靠镜头分割和摄像机参数,计算复杂且对设定的分割参数敏感. 本文提出了利用代表帧表示视频内容的方法弥补了这一缺陷. 另外,在现代社会中,视频的应用越来越广泛,随着视频的内容越来越丰富,容量越来越大,对于视频认证技术,如何简洁有效的浏览和表示视频数据也是需要考虑的问题. 本文从图建模的角度入手,利用二叉树选取代表帧. 利用随机邻域嵌入 (Stochastic Neighbor Embedding, SNE) 把视频数据降维到一个低维空间上进行表示,以达到数据约简的目的,从而在低维空间上构造视频指纹,实现视频认证. 本文的主要贡献总结如下:

(1) 提出基于二叉树的代表帧提取方法. 该方法利用成熟的图论和归一分割优化理论来提取代表帧. 首先把视频建模为无向图,然后利用二叉树把视频帧分成若干簇,在每个簇中选取部分帧作为代表帧.

(2) 把维数约简应用到视频分析中. 从直观的数据表示形式上来看,视频存在于由若干像素点组成的高维空间上,这对检索的快速执行是非常不利的. 因此有必要对视频维数约简. 本文利用随机邻域嵌入把视频帧投影到低维空间上,在低维空间构造视频指纹,大大降低了算法的复杂性.

2 视频指纹方案

本文算法的流程图如图 1 所示,共分为视频预处理、代表帧提取、随机邻域嵌入投影和视频指纹生成与

匹配四个部分,具体内容在以下各节详述.

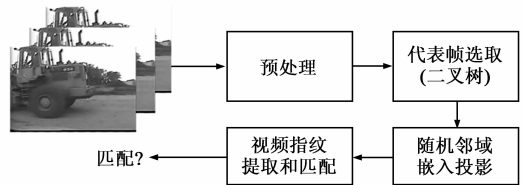


图1 视频指纹系统

2.1 预处理

视频预处理主要包括视频帧率和分辨率的修改,而帧率和分辨率的改变不会影响视频内容. 因此,视频预处理是可行的. 预处理的目的是为了增加算法的鲁棒性和视频分辨率以及帧率的统一性. 在本文算法中,帧率通过重采样的方法设为 5fps,而每帧的分辨率统一设为 100×100 .

2.2 代表帧选取

视频帧之间存在大量的冗余信息,且视频帧的数量巨大,对每帧提取特征作为视频指纹的方法计算复杂度太高,因此,本文在视频不同的部分选取一些帧作为代表帧,利用代表帧内容表示视频内容,这样做既能表达视频内容,又可以减小算法的复杂性. 代表帧除了和传统关键帧一样能表示视频内容之外,还和关键帧有很多不同之处. 传统关键帧提取算法大多都是基于视频镜头分割的,镜头界限分为突变和渐变,渐变又包括淡入淡出、溶解、扫换等编辑效果^[15]. 镜头的分割不仅计算复杂度高,而且易受噪声、光线变化等参数影响. 本文采取的利用成熟的二叉树理论提取代表帧方法,不仅能够代表视频内容,而且具有较好的稳定性,这恰恰是视频指纹技术所要求的.

本文利用二叉树思想来选取代表帧. 二叉树示意图如图 2 所示. 对于视频 V 的所有帧,利用归一理论把所有帧分为两部分,对于每一部分再进行二分,以此类推,至达到预先设定的层数为止,则视频所有帧递归二分的过程就构成一个二叉树. 二叉树最低层的叶节点对应于一组帧,本文称之为簇,从每个簇中选取部分帧作为代表帧. 算法的主要思路包括以下三个部分.

(1) 视频帧的二叉树建模

从宏观的角度来看,可以把视频的每一帧当作高维空间中的一个点,点的坐标由帧的各种属性决定,点与点之间根据帧之间的关系连线就可以构成一个图. 因此,视频可以看作高维空间里的一个无向权重图 $G = (V, E)$,其中 V 和 E 分别代表图顶点集和边集. 因此,划分视频帧的过程就转化为划分图 G 顶点的过程,

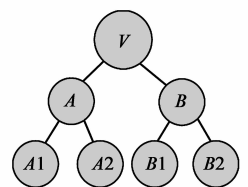


图2 视频的二叉树示意图

在图论中,图 $G = (V, E)$ 通过去除两部分相连的边,可以被分解为不相交的集合 A 和 B , 其中 $A \cup B = V, A \cap B = \emptyset$. 对于集合 A 和 B 同样分别可以分解为两部分 A_1 和 A_2, B_1 和 B_2 . 以此下去,视频帧划分的过程就等价于构造图 G 的一个二叉树的过程.

(2) 二叉树的构造

如何有效合理构造图 G 的二叉树是视频代表帧提取的关键. 本文采用归一化分割理论来构造二叉树. 本文以根节点 V 为例, 详细介绍二叉树下一级节点生成过程, 在图 2 中, 根节点 V 被分成两个子节点 A 和 B , 两个子节点相异程度等于在图 G 去除两部分相连边的权重和, 此相异程度称为分割 (cut), 即为 $\sum_{i \in A, j \in B} w(i, j)$. 其中 $w(i, j)$ 为两点之间边的权重, 研究如何优化分割的理论, 即图分割理论被广泛应用于图像分割等领域, 但传统分割理论容易出现较小节点集合的不自然情况. 为避免此类问题, 文献 [16, 17] 提出了归一化分割理论, 本文利用归一化分割理论把根节点 V 进行二分.

归一化分割理论通过最小化式 (1) 把图 G 的顶点集最优的分为两部分 A 和 B :

$$M_{\text{cut}}(A, B) = \frac{\text{cut}(A, B)}{\text{cut}(A, V)} + \frac{\text{cut}(A, B)}{\text{cut}(B, V)} \quad (1)$$

其中, $\text{cut}(A, B)$ 是两集合 A 和 B 的图分割值. $\text{cut}(A, B) = \sum_{i \in A, j \in B} w(i, j)$, $\text{cut}(A, V) = \sum_{i \in A, j \in V} w(i, j)$, $\text{cut}(B, V) = \sum_{i \in B, j \in V} w(i, j)$. 对图 G 的最优二分法即最小化目标函数式 (1). 权重 $w(i, j)$ 代表了图中两点的距离相近程度, 本文定义权重如下:

$$w(i, j) = \exp\left\{\frac{-k \cdot \text{sim}(i, j)}{|f_j - f_i|}\right\} \quad (2)$$

$$\text{sim}(i, j) = \max_{u \in P} \{H_i(u), H_j(u)\} \quad (3)$$

其中, $\text{sim}(i, j)$ 和 $|f_j - f_i|$ 分别代表第 i 帧和第 j 帧亮度相似值和时域距离, k 是一个常数. P 是两帧相同亮度等级的集合, $H_i(u), H_j(u)$ 分别是第 i 帧和第 j 帧在亮度 u 等级上的一化直方图的值. 从式 (2) 可以看出, 权重的计算充分考虑了视频的时空因素.

对于式 (1) 的优化问题, 具体计算过程可参见文献 [17]. 若顶点 i 属于集合 A , 则设 $x_i = 1$, 否则 $x_i = -1$. 设 $d(i) = \sum_j w(i, j)$, 则式 (1) 可转化为:

$$M_{\text{cut}}(A, B) = \frac{\text{cut}(A, B)}{\text{cut}(A, V)} + \frac{\text{cut}(A, B)}{\text{cut}(B, V)} = \frac{\sum_{(x_i > 0, x_j < 0)} -w(i, j) x_i x_j}{\sum_{x_i > 0} d(i)} + \frac{\sum_{(x_i < 0, x_j > 0)} -w(i, j) x_i x_j}{\sum_{x_i < 0} d(i)} \quad (4)$$

$$\text{令 } k = \frac{\sum_{x_i > 0} d(i)}{\sum_i d(i)}, \text{ 式 (4) 可继续转化为}$$

$$M_{\text{cut}}(A, B) = \frac{\mathbf{x}^T(\mathbf{D} - \mathbf{W})\mathbf{x} + \mathbf{1}^T(\mathbf{D} - \mathbf{W})\mathbf{1}}{k(1-k)\mathbf{1}^T\mathbf{D}\mathbf{1}} + \frac{(2(1-2k)\mathbf{1}^T(\mathbf{D} - \mathbf{W})\mathbf{1})}{k(1-k)\mathbf{1}^T\mathbf{D}\mathbf{1}} = \frac{\mathbf{y}^T(\mathbf{D} - \mathbf{W})\mathbf{y}}{\mathbf{y}^T\mathbf{D}\mathbf{y}} \quad (5)$$

其中利用 $\frac{1+\mathbf{x}}{2}$ 代表 $x_i = 1$, $\frac{1-\mathbf{x}}{2}$ 代表 $x_i = -1$, $\mathbf{y} = (1+\mathbf{x}) - b(1-\mathbf{x})$, $b = k/(1-k)$.

式 (5) 的最优化问题是一个广义瑞利商的问题, 可转化为如下的广义特征值问题求解:

$$(\mathbf{D} - \mathbf{W})\mathbf{y} = \lambda\mathbf{D}\mathbf{y} \quad (6)$$

其中 \mathbf{D} 是对角矩阵, 对角线元素为 $\sum_j w(i, j)$, \mathbf{W} 是由 $w(i, j)$ 组成的对称矩阵. 其中第二小的特征值对应的特征向量用于把顶点集分割成 A 和 B . 对于由 A 和 B 组成的子图, 递归的利用式 (1) 进行二分, 直到每个二叉树的叶节点对应的视频簇内任意两点之间的权重小于阈值 t . 阈值 t 定义为图 G 权重均值和方差的和值.

(3) 代表帧选取

二叉树的每个叶节点对应与视频的一个簇, 本文在每个簇中选取固定数量的帧作为代表帧, 选取方法如下:

对第 m 个簇中, 计算簇内每帧的平均四阶累积量, 记为 c_i , 则该簇内所有帧的平均四阶累积量构成序列 $\{c_i\}_{1 \times T}$, T 为簇的数目. 计算该序列的均值和方差, 分别记为 μ 和 σ , 选取该簇中平均累积量约等于 $\mu, \mu + \sigma$ 和 $\mu - \sigma$ 的三类帧作为代表帧.

代表帧的数目和二叉树叶节点的数目以及每簇中选择的帧的数目有关, 假设二叉树共有 j 层, 每簇中选择的帧数为 n , 则共有 $n \cdot 2^j$ 个代表帧. 本文仿真实验中, 取 $j = 4, n = 20$.

对于视频指纹算法来说, 视频在遭受攻击前后代表帧提取的稳定性是非常重要的. 本文算法的代表帧提取的稳定性主要是指视频利用二叉树分组的稳定性, 而传统关键帧提取算法稳定性主要是指视频镜头分割的稳定性, 为验证本文代表帧选取方法在遭受攻击后的稳健性, 本文选取五个不同的视频, 分别施加最典型的攻击手段——随机噪声攻击, 并把本文提取代表帧算法的稳定性和传统利用阈值和帧间差异提取关键帧的方法 [18] 作比较, 实验结果如图 3 所示. 代表帧/关键帧选取的相关性定义为攻击前后相同的代表帧/关键帧数目与原代表帧/关键帧数目的比值. 由图 3 可以看出, 本文代表帧选取方法明显优于传统关键帧选取方法.

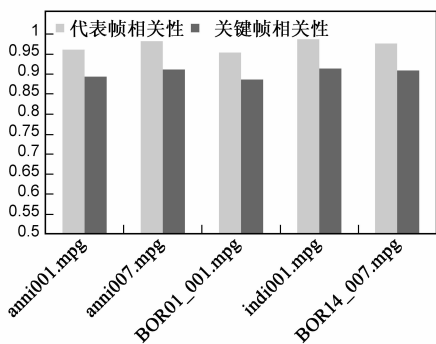


图3 不同视频在噪声攻击下代表帧/关键帧选取的相关性

2.3 随机邻域嵌入投影

数据的有效表达是模式识别和特征提取中的重要问题.一幅图像所包含的大量信息若使用传统的欧氏向量表示,会导致计算机难以计算图像间的差异,进而难以对图像进行准确的识别及有效的检索.而使用降维算法可以有效地发现高维数据分布的内在几何结构,挖掘隐藏在低维数据中的本征信息与内在规律,得到图像数据的内在表达.这种低维内在空间上对图像的表达比在高维欧氏空间中更能体现不同图像之间的内在差异,也就是说,能够更有效地反映出图像内容的异同.而视频的每个帧就是一幅静态图像,对于视频来说,丰富的场景和镜头信息等局部信息对视频内容表达非常重要,因此,为很好的保持视频的局部信息,本文采用基于统计特性的局部流形降维算法——随机邻域嵌入对视频进行降维.

2.3.1 随机邻域嵌入

随机邻域嵌入的基本思想是通过最小化高维空间和低维空间数据点之间条件概率的相对熵来实现维数约简.设 y_i 和 y_j 是高维空间数据点 x_i 和 x_j 在低维空间中的映射点, $p_{j|i}$ 和 $q_{j|i}$ 分别为点 x_j 和 y_j 相对于 x_i 和 y_i 的条件概率.在理想情况下有式 $p_{j|i} = q_{j|i}$ 成立,即低维空间映射点完全保持了高维数据结构和关系.随机邻域嵌入通过最小化相对熵(或交叉熵)来寻找低维映射点的.相对熵目标函数定义如下:

$$C = \sum_i D(P_i || Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}} \quad (7)$$

其中, P_i 代表给定点 x_i 后,其他点在高维空间的条件概率分布函数; Q_i 代表给定映射点 y_i 后,其他映射点在低维空间的条件概率分布函数. $D(\cdot)$ 为 Kullback-Leibler 差异 (Kullback-Leibler Divergence), 又称为相对熵.此最小化问题,可以通过梯度下降算法求解.

随机邻域嵌入利用统计方法和数据的局部特征把高维空间中的数据投影到低维空间上,因此利用随机邻域嵌入对视频进行维数约简是非常恰当的.随机邻域算法的具体过程可参考文献[19].

2.3.2 视频投影

视频的每一帧可以看作以像素为坐标值的高维空间中的点,但是对于视频指纹算法来说,像素的鲁棒性太弱,因此,本文采用帧亮度分量的 DCT 变换的低频系数作为帧在高维空间中的坐标,然后利用随机邻域嵌入实现视频降维.具体过程如算法 1 所示:

算法 1: 视频投影算法

输入: 视频代表帧.

输出: 低维空间映射点范数序列

开始: 对视频进行预处理,并提取代表帧,方法如本文 2.1, 2.2 所述;

For $k = 1; N$ Do // N 为代表帧的数目

 计算亮度系数的 DCT 变换,取前 100 个较大的 DCT 系数,定义为 $\{D_k^0: 1 \leq k \leq N\}$,此向量即为该帧在高维空间中的点坐标向量.

End

所有代表帧的高维坐标向量组成矩阵 $D: D = \{D_k^0: 1 \leq k \leq N\}$.

利用随机邻域嵌入的方法,把视频各代表帧投影到三维空间中,得到点列 $v = \{v_i\}_{1 \times N}$,计算相邻点的 F-范数 $d_m = \|v_{i+1} - v_i\|_F, 1 \leq m \leq N-1$,则得到一个范数序列 $d^* = \{d_m\}, 1 \leq m \leq N-1$,用来生成视频指纹序列.

结束

视频“BOR04_01.mpg”帧截图及其在三维空间中的投影如图 4 所示.

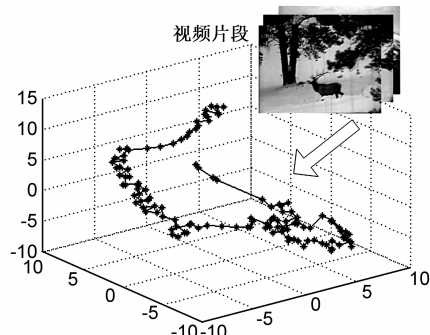


图4 视频片段“BOR04_01.mpg”在三维空间中的投影轨迹

2.4 视频指纹的生成和匹配

2.4.1 视频指纹的生成

本文对低维空间中相邻点的距离序列进行二值化,以此比特序列作为视频的指纹,方法如下:

首先生成一个范围在 $[0, 1]$, 零均值服从均匀分布的随机序列 $p = \{p_m\}$ 作为密钥,根据式(8),生成哈希序列 $h = \{h_k\}$.

$$h_m = \begin{cases} 1 & \text{if } |d_m * p_m| \geq \theta \\ 0 & \text{if } |d_m * p_m| < \theta \end{cases}, 1 \leq m \leq N-1 \quad (8)$$

其中 θ 为阈值,计算方法如下:

$$\theta = \text{median}(|d_m * p_m|), 1 \leq m \leq N-1 \quad (9)$$

对于低维空间相邻点距离序列,计算其均值 μ 和方差 σ ,以 $u = \mu + b * \sigma$ 作为匹配密钥在视频匹配中用

来缩小视频匹配范围,在本文实验中 b 取 0.6. 因此,视频指纹 F 由两部分组成,即 $F = \{h, u\}$.

2.4.2 视频指纹的匹配

视频指纹匹配是视频认证过程重要的环节之一,由本文算法可知,视频指纹的长度和代表帧的数目基本相同(代表帧数目减 1),因此本文按照以下两个步骤实现视频指纹的匹配.

步骤 1 对于待测视频,首先根据匹配密钥 u ,得到一个候选匹配指纹集合,大大缩小了匹配的范围.

步骤 2 在候选集合里,利用二值化的指纹序列进行视频指纹进行细致匹配.

3 仿真实验结果和实验分析

3.1 实验设计

本文的实验视频来自于视频数据库 MUSCLE-VCD-2007,为验证方法的性能,共进行两个方面的实验.

(1)鲁棒性和区分性测试.把该方法应用于视频数据库中随机抽取的二百个视频,每个视频帧的数目在 800 到 3000 之间,计算在各种攻击下的误码率平均值.本文设定阈值为 0.15,若待测视频与原始视频的指纹序列相比误码率小于阈值,则说明待测视频是原始视频的一个拷贝.

(2)视频检测的查全率.查全率 R_c 定义如下:

$$R_c = N_{tp} / N_{ep} \quad (10)$$

其中 N_{tp} 是检索出正确的视频拷贝数, N_{ep} 是数据库中视频拷贝数.

3.2 实验结果及分析

在算法鲁棒性和区分性实验中.本文把实验结果和文献[3]、[13]和[14]比较.实验结果如表 1.由表 1 可以看出,本文的算法在所有攻击条件下,误码率都小于文献[3]的方法.在和文献[13]的比较中,除了噪声攻击和帧丢弃攻击下性能比文献[13]稍差外,其他均好于文献[13]的方法,这是因为文献[13]中使用的多维标度是一个全局流形学习算法而本文算法采用的随机邻域嵌入算法更关注视频的局部结构,噪声攻击和帧丢弃会破坏视频帧之间相邻关系这一局部结构,导致算法在这些攻击下性能稍差,但本文算法在这两种攻击下的误码率仍小于既定阈值.在和文献[14]中方法的比较中,可以看出,本文算法性能在模糊、帧丢弃、放缩和 MPEG 压缩等多数攻击下明显好于文献[14].与文献[13]和[14]相比,本文算法虽然在个别攻击下性能稍差,但仍然达到较高的性能指标.另外,本文算法在算法时间效率上,即同样的视频数据库中视频指纹检测时间明显少于文献[13,14]算法.视频指纹检测时间包括视频指纹提取和匹配两部分,为了验证算法的时间

效率,不失一般性,本文以文献[13]算法时间为基准,即把文献[13]算法时间设为单位“1”,并把其他算法时间和文献[13]算法时间的比值定义为算法时间效率,取 100 个视频作为数据库,这 100 个视频由 20 组不同内容视频组成,每组有一个原始视频和四种攻击下的拷贝视频组成.在每组视频中随意取一个视频作为待测视频,提取视频指纹并进行视频指纹匹配.本文算法和文献[13,14]算法的时间效率比较如图 5 所示.由图 5 可以看出,本文算法时间效率明显优于文献[13,14],原因主要是:文献[13]采用的算法针对所有帧提取特征,没有关键帧提取过程,虽然多维标度算法的速度较快,但是当视频帧数较多时,算法的时间复杂性也变高了,同时,该算法的匹配过程是一种比较耗时的穷尽比较的过程,文献[14]综合采用多维标度和局部线性嵌入构造视频指纹,利用滑动窗进行视频指纹匹配,这种匹配方法仍然需要与数据库中所有视频指纹比对,也是比较耗时的过程,同时此算法中的局部线性嵌入时间复杂性也较高.本文算法采用代表帧特征提取,而且采用双层匹配,通过匹配密钥缩小了匹配范围,从而不需要与所有视频比对,因此时间复杂性低于文献[13,14].随着互联网中视频数量的增多,视频指纹算法的时间效率受到越来越多的重视.

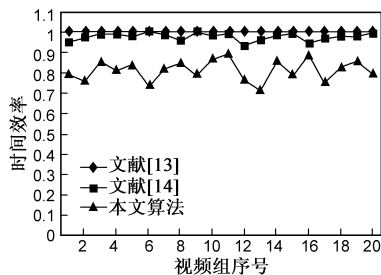


图5 视频检测时间效率

从对不同视频间误码率的测试中可以看到,对于内容上完全不同的视频,误码率大约在 0.5 左右,这就保证了本文算法良好的区分性,对于判别算法区分性好坏的误码率取值可参见文献[4].

对于算法查全率测试,本文通过多次实验研究算

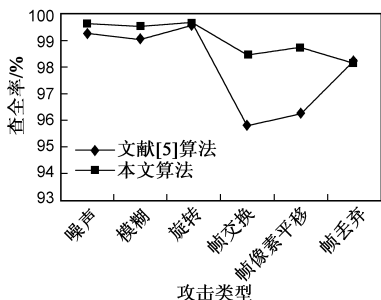


图6 不同攻击下查全率

法的性能,每次随机取五十个视频为样本进行不同的攻击操作(样本中可包含视频数据库中不有的视频),然后在视频库中进行匹配.平均查全率与文献[5]比较如图6所示.从图6可以看出,本文算法在大部分攻击

下的查全率均优于文献[5]算法.本文算法平均各种攻击下的查全率为99.01%,比文献[5]的98.02%的平均查全率有所改进.

表1 各种攻击下的平均误码率及与文献[3,13,14]的比较

攻击类型	攻击后视频与原视频误码率				不同内容视频误码率		
	本文方法	文献[3]	文献[13]	文献[14]	本文方法	文献[13]	文献[14]
高斯噪声	2.19%	9.87%	1.3%	0.95%	50%	49%	48%
模糊	1.21%	2.71%	2.1%	2.3%	47%	49%	49%
旋转	1.84%	2.63%	5.7%	0.68%	49%	48%	48%
帧丢弃	4.07%	6.71%	3.6%	12%	53%	49%	51%
帧像素平移	5.68%	9.13%	7.2%	1.55%	51%	50%	50%
放缩	1.05%	7.26%	5.7%	1.7%	51%	51%	51%
MPEG压缩	6.12%	12.5%	9.3%	7.68%	48%	51%	49%

4 结论

本文提出了一种基于二叉树和随机邻域嵌入的视频指纹方法.该方法首先利用二叉树和归一分割理论选取代表帧,然后把代表帧亮度的DCT变换系数作为高维特征,利用随机邻域嵌入把视频帧投影到低维空间中,并利用低维空间相邻点的距离来构造匹配密钥和视频指纹序列,实现视频认证的目的.实验证明该算法具有较好的鲁棒性和区分性,从而为视频管理提供一个可靠、可行的算法和思路.

本文算法在选取代表帧的过程中,对二叉树的层级和叶节点中选取帧的数目选取固定数值,这虽然减小了计算的复杂性,但代表帧数目应该与视频长短及视频内容有关,如何动态的选择二叉树层级和叶节点的数目,使其与视频个体相关,是将来研究工作的重点.

参考文献

- [1] 梁长垠,李昂,黄继武.基于MPEG压缩的视频认证算法测试平台[J].电子学报,2008,36(12A):133-137.
Liang C Y, Li A, Huang J W. Test platform for video authentication algorithm based on MPEG compression[J]. Acta Electronica Sinica, 2008, 36(12A): 133-137. (in Chinese)
- [2] 徐甲甲,张卫明,俞能海,等.一种基于秘密共享与运动矢量的视频水印算法[J].电子学报,2012,40(1):8-13.
Xu J J, Zhang W M, Xu N H, et al. A video watermarking algorithm based on secret sharing and motion vectors[J]. Acta Electronica Sinica, 2012, 40(1): 8-13. (in Chinese)
- [3] JOLY A, BUISSON O, FRELICT C. Content-based copy retrieval using distortion-based probabilistic similarity search[J]. IEEE Trans on Multimedia, 2007, 9(2): 293-306.

- [4] LEE S, YOO C D. Video fingerprinting based on centroids of gradient orientations [A]. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing [C]. Toulouse: Institute of Electrical and Electronics Engineers Inc, 2006, 401-404.
- [5] Zhou X B, Martin S, Chris B. Perceptual hashing of video content based on differential block similarity [J]. Lecture Notes in Computer Science, 2005, 3802: 80-85.
- [6] Coskun B, Sankur B, Memon N. Spatio-temporal transform-based video hashing [J]. IEEE Transactions on Multimedia, 2006, 8(6): 1190-1208.
- [7] Esmaeili M M, Fatourechi M, Ward R K. A robust and fast video copy detection system using content-based fingerprinting [J]. IEEE Transactions on Information Forensics and Security, 2011, 6(1): 213-226.
- [8] 潘雪峰,李锦涛,张勇东,等.基于视觉感知的时空联合视频复制检测方法[J].计算机学报,2009,32(1):108-114.
Pan X F, Li J T, Zhang Y D, et al. Spatiotemporal video copy detection based on visual perception analyses [J]. Chinese Journal of Computers, 2009, 32(1): 108-114. (in Chinese)
- [9] 张勇东,张冬明,郭俊波,等.压缩域快速视频拷贝检测算法[J].通信学报,2009,30(3):135-140.
Zhang Y D, Zhang D M, Guo J B, et al. Rapid video copy detection on compressed domain [J]. Journal on Communications, 2009, 30(3): 135-140. (in Chinese)
- [10] Wu Z P, Huang Q M, Jiang S Q. Robust copy detection by mining temporal self-similarities [A]. Proceeding of ICME [C]. New York: IEEE Computer Society, 2009. 554-557.
- [11] Zhao Y X, Liu G J, Dai Y W, et al. Robust hashing based on persistent points for video copy detection [A]. Proceeding of International Conference on Computational Intelligence and Security (CIS) [C]. Suzhou: Inst of Elec and Elec Eng, 2008.

305 – 308.

- [12] 赵玉鑫,刘光杰,戴跃伟,等.基于局部排序的视频复制检测[J].计算机辅助设计与图形学学报,2009,21(9):1339 – 1343.
Zhao Y X, Liu G F, Dai Y W, et al. Video copy detection based on local ordinal[J].Journal of Computer-Aided Design & Computer Graphics, 2009, 21(9): 1339 – 1343. (in Chinese)
- [13] Nie X S, Liu J, Sun J D. Robust video hashing for identification based on MDS[A]. Proceeding of Acoustics Speech and Signal Processing (ICASSP)[C]. Dallas: Institute of Electrical and Electronics Engineers Inc, 2010. 1834 – 1837.
- [14] Nie X S, Liu J, Sun J D, et al. Robust video hashing based on double-layer embedding[J]. IEEE Signal Processing Letters, 2011, 18(5): 307 – 310.
- [15] 耿玉亮,须德,冯松鹤.一种快速有效的视频镜头边界检测方法[J].电子学报,2006,34(12):2272 – 2277.
Geng Y L, Xu D, Feng S H. A rapid and effective method for shot boundary detection[J]. Acta Electronica Sinica, 2006, 34(12): 2272 – 2277. (in Chinese)
- [16] Ngo C W, Ma Y U, Zhang H J. Video summarization and scene detection by graph modeling[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(2): 296 – 305.
- [17] Shi J, Malik J. Normalized cuts and image segmentation[J]. IEEE Trans Pattern Anal Machine Intell, 2000, 22(8): 888 – 905.
- [18] Zhang J, Kankanhlli A, Smoliar S W. Automatic partitioning of full motion video[J]. Multimedia Systems, 1993, 1(1): 10 – 28.
- [19] Laurens V D, Hinton G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(Nov): 2579 – 2625.

作者简介



聂秀山 男,1981年3月生,江苏徐州人,博士,副教授,硕士生导师,主要研究方向为视频认证,多媒体内容分析等.

E-mail: niexiushan@163.com



刘 琚 男,1965年4月生,山东临沂人.博士,教授、博士生导师.现为山东大学信息科学与工程学院学术委员会副主任,通信工程系主任,研究方向为多媒体通信等.

E-mail: juliu@sdu.edu.cn