基于局部线性嵌入的视频拷贝检测方法

聂秀山[①②] 刘 琚[*①③] 孙建德[①] 秦丰林[①]
[①](山东大学信息科学与工程学院 济南 250100)
[②](山东财政学院计算机信息工程学院 济南 250014)
[③](海信数字多媒体技术国家重点实验室 青岛 266071)

摘 要：该文提出了基于局部线性嵌入(LLE)的视频哈希方法，该方法首先利用一个图模型选取代表帧，然后以四阶累积量作为视频在高维空间的特征并利用 LLE 对视频进行降维，利用视频在 3 维空间中投影点的范数构造视频哈希序列来实现视频拷贝检测。实验证明该方法具有较好的鲁棒性和区分性。

关键词：视频哈希；拷贝检测；鲁棒性；局部线性嵌入
中图分类号：TP391 **文献标识码**：A **文章编号**：1009-5896(2011)05-1030-05
DOI: 10.3724/SP.J.1146.2010.01007

A LLE-based Video Copy Detection Method

Nie Xiu-shan[①②] Liu Ju[①③] Sun Jian-de[①] Qin Feng-lin[①]
[①](*School of Information Science and Engineering*, *Shandong University*, *Jinan* 250100, *China*)
[②](*School of Computer and Information Engineering*, *Shandong University of Finance*, *Jinan* 250014, *China*)
[③](*Hisense State Key Laboratory of Digital Multi-Media Technology Corp.*, *LTD.*, *Qingdao* 266071, *China*)

Abstract: A video hashing based on Locally Linear Embedding (LLE) is proposed in this paper. In this method, some representative frames are first selected based on a graph model, and four-order cumulants are taken as features of video in the high dimensional feature space. Then the video is mapped to a three-dimensional space using LLE, and video hash sequence is generated using the norms of points in the three-dimensional space to detect video copies. Experimental results show that the video hashing has good robustness and discrimination.

Key words: Video hashing; Copy detection; Robustness; Locally Linear Embedding (LLE)

1 引言

在现代社会中，随着多媒体技术和 Internet 的发展，网络视频变的越来越丰富，视频在社会生活和军事领域中的应用越来越多，视频检索在网络中的应用也越来越广泛。与此同时，互联网上对于海量视频的管理却是缺乏规划和统一性，常常引起一些诸如知识产权等的纠纷。作为视频检索的一个分支，基于内容视频拷贝检测(CBVCD)被提出并成为解决上述问题的主要方法。

关于 CBVCD 的研究是从 20 世纪 90 年代末开始的，Indyk 等人[1]利用视频镜头边缘生成时域上的视频指纹来检测互联网上的盗版视频，较早地开展了网络 CBVCD 的探索。Joly 等人[2]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签名向量。这些方法的研究都在某种程度上忽略了视频在时域上的变化特性，因此，一些研究者结合了视频的时空域 3 维特性以及运动特性，提出视频哈希的概念，以视频哈希的相似性衡量作为进行 CBVCD 的主要方式。Coskun 等人[3]通过对视频的 3 维 DCT 变换来获取视频在时空域上的频率特征来构成视频哈希。这些研究充分利用了视频区别于图像的时空特征，达到了性能更优的检索效果。

国内的研究者在 CBVCD 的研究上虽然起步稍晚，但也取得了很多优秀的研究成果。其中具有代表性的：中国科学院计算所李锦涛教授[4]带领的前瞻研究实验室利用视频在时空域上的视觉特征进行视频拷贝检测，张勇东等人[5]较早地开展了基于压缩域的视频拷贝检测的研究；中国科学院黄庆明教授[6]带领的研究组则将自相似矩阵和视觉特征串相结合，进行视频拷贝检测；南京理工大学戴跃伟教授[7]带领的研究组致力于提取视频中最具稳定性的特征点，通过它们构造视频哈希来实现视频复制检测。

2010-09-14 收到，2010-12-16 改回
国家 973 计划项目(2009CB320905)和国家自然科学基金(60872024，60970048，61001180)资助课题
*通信作者：刘琚 juliu@sdu.edu.cn

他们都取得了较好的研究成果。

随着视频的内容越来越丰富，容量越来越大，对于拷贝检测技术，如何简洁有效地浏览和表示视频数据成为首要考虑的问题。有必要把视频数据映射到一个低维空间上进行表示，流形学习就是解决此问题的一个思路。本文就是基于与一种经典的流形学习方法——局部线性嵌入(LLE)来生成视频哈希，从而快速有效地提取视频特征进行视频拷贝检测。

2　基于 LLE 的视频哈希生成方案

整个视频拷贝检测系统共有 3 个主要部分组成：图模型、代表帧选取、视频哈希生成。具体流程见图 1。

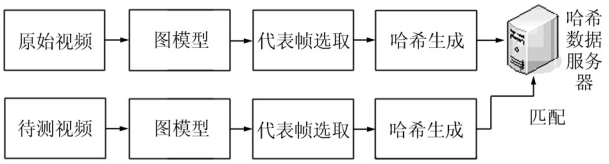


图 1 视频拷贝检测系统

2.1 图模型

本文为选取能代表视频内容的代表帧，把视频等价为一个无向的权重图，其中，每一帧作为图的一个点，两点之间边的权重 $w(i,j)$ 定义如下：

$$w(i,j)=\exp\left\{\frac{-k\cdot \mathrm{sim}(i,j)}{|f_j-f_i|}\right\} \tag{1}$$

$$\mathrm{sim}(i,j)=\max\min_{u\in P}\{H_i(u),H_j(u)\} \tag{2}$$

其中 $\mathrm{sim}(i,j)$ 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 等级上的归一化直方图。从式(1)可以看出，权重的计算充分考虑了视频的时空因素，两帧在时间轴上的距离越远，两帧之间距离越大，同时，两帧的亮度相似值越大，两帧之间的距离就越小。这与实际情况是相吻合的。

对于无向图 G^s，若 $w(i,j)$ 小于给定的阈值，则图 G^s 中对应两点有边相连，其权重值为 $w(i,j)$。反之，两点不相连，其权重值为 0。本文阈值取为所有权重的平均值。由此得到的无向权重图，称之为该视频的时空图。

2.2 代表帧选取

众所周知，视频含有数目众多的帧，若以全部的视频帧作为局部线性嵌入的输入，计算量非常大，为了节省计算量，本文选取视频的代表帧作为输入。所谓代表帧，即可以代表视频中不同镜头内容帧的集合，代表帧的选取要尽可能地分散在不同的镜头里面，因此，选取代表帧的过程，等价于在视频时空图中寻找无不相连的顶点的过程，这是图论中的寻找图的独立集的问题，最大独立集问题是一个 NPC 问题，因此，本文提出了一个寻找唯一而非最大独立集的算法来解决代表帧的选取问题。具体过程如下。

(1)由视频的时空图 G^s，建立一个关联矩阵 $\boldsymbol{R}=\{r_{ij}\}_{N\times N}$ 如下，N 为视频帧的数目。

$$r_{ij}=\begin{cases}1, & w_{ij}>0\\ 0, & w_{ij}=0\end{cases} \tag{3}$$

(2)对矩阵 $\boldsymbol{R}$ 每一行(列)求和，此和称为行(列)标号对应点的关联数，把关联数最大第 i 行(列)和第 i 列(行)的元素全部变为零。

(3)按照(2)中所述方法处理，直到矩阵 $\boldsymbol{R}$ 的行(列)的关联数最大为 1 为止。

换句话说，这些关联数为 1 的行(列)标号所对应的无向图 G^s 中的点只与自身相关联，此时就得到了图 G^s 的一个独立集，根据视频帧与点的对应关系，即得到了视频的一组代表帧。

2.3 视频哈希生成

2.3.1 局部线性嵌入映射　局部线性嵌入算法[8]是一个典型的流形学习算法。它假设数据点 $x_i\in\boldsymbol{R}^m$ 和它的近邻分布在流形的一个局部线性区域，其基本思想是认为能最佳重构高维空间中的数据点的权值能把流形的局部几何信息从高维空间携带到低维空间。

在局部线性嵌入的算法流程中，一个非常重要的步骤就是高维数据点邻居的选取，根据不同的应用可以选择不同的度量标准来选取邻居，对于视频拷贝检测来说，最重要的原则就是对一些非恶意修改的强鲁棒性。本文以帧间块的 DCT 系数差值作为邻居选取的度量标准，具体实现的算法如下：

(1)对代表帧每一帧进行 8×8 分块，并对每块的亮度进行分块 DCT 变换，取每块 DCT 系数的中间 30%的系数并相加，构成向量 $\{P_n^k:1\le n\le m\}$，若两帧之间所有块亮度 DCT 系数的差值的平均值小于某个阈值 r，即 $d_{kt}=\frac{1}{m}\sum_{n=1}^{m}|P_n^k-P_n^t|\le r$，则两帧互为邻居，其中阈值取为由所有帧之间差值 d_{kt} 构成序列的均值。

(2)对于每帧 f_k，得到邻居数目 N_k，为计算方便，取所有帧一个公共的邻居数目 K，即 $K=\min\{N_k\,|\,1\le k\le N\}$。

2.3.2 累积量 视频是由连续的帧组成，每帧是一幅图像，假设每帧有 $w\times h$ 个像素，则此帧存在于以相应点像素值为坐标的 $w\times h$ 维空间中，但是对于视频的拷贝检测来说，像素的鲁棒性太弱，因此，必须寻找具有强鲁棒性的特征空间作为局部线性嵌入的高维输入，因为高阶累计量对噪声不敏感，因此本文选取四阶累计量作为视频高维空间的特征。设信号 $X(n)$ 的四阶累积量是 C_{4X}，则有

$$\begin{aligned}C_{4X}(k,l,m)&=E\{X(n)X(n+k)X(n+l)X(n+m)\}\\&\quad-C_{2X}(k)C_{2X}(l-m)-C_{2X}(l)\\&\quad\cdot C_{2X}(k-m)-C_{2X}(m)C_{2X}(k-l)\end{aligned}\tag{4}$$

对于不改变视频内容的攻击可以建模成一高斯过程，而高阶累积量具有去高斯性，因此选择高阶累积量作为视频高维空间的特征是具有很强鲁棒性的，本文取四阶累积量作为高维特征，四阶累积量是 3 维的，为计算方便和得到一个简洁紧凑的高维特征，因此在本文令 $l=m=0$。

对于含有 $w\times h$ 个像素的每一帧，首先用 zigzag 的方式拉伸为 1 维向量 $\boldsymbol{f}=\{f_1,f_2,\cdots,f_M\},M=w\times h$。然后按照式(4)求其四阶累积量，根据累计量的定义，结合本文实际应用，其中 $k\in[-w\times h,w\times h]$。本文采用 Matlab 高阶累计量工具箱中的函数实现累积量的计算。得到累计量系数的数目可能会非常大，包含很多冗余信息。为减少计算量，本文对每帧的累计量系数序列进行 DCT 变换，取包括直流系数在内前 40 个低频系数作为特征，因为低频系数代表了信号的主要能量。因此，视频所在高维特征空间的维数为 40，每个坐标值的大小即相应的累积量 DCT 系数的值。

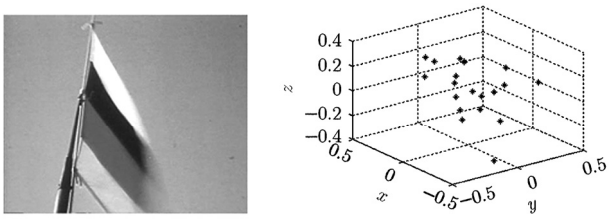
2.3.3 视频的局部线性嵌入投影 视频每帧可以看作以累积量的 DCT 系数为坐标的高维空间中的点，利用局部线性嵌入，可把视频投影到一个 3 维的空间上，具体过程如下：

(1)选取给定视频的代表帧，并为每个代表帧寻找邻居。

(2)对于每个代表帧，计算其亮度系数的四阶累计量并进行 DCT 变换，取前 40 个较大的 DCT 系数，定义为 $\{D_k^O:1\le k\le N\}$，N 是关键帧的数目。并组成矩阵 $\boldsymbol{D}_{40\times N}:\boldsymbol{D}=\{D_k^O:1\le k\le N\}$。

(3)利用局部线性嵌入的方法，把视频各帧投影到 3 维空间中，得到点列 $\boldsymbol{v}=\{v_i\}_{1\times N}$，计算每个点的 F- 范数 $||v_i||_F$，则得到一个范数序列 $\boldsymbol{v}^o=\{v_i^o\mid v_i^o=||v_i||_F\}1\le i\le N$，并用来生成视频哈希序列。

视频“indi010.mpg”帧截图及其在 3 维空间中的投影如图 2 所示。



(a) 视频片段 (b) 在 3 维空间的投影

图 2 视频片段“indi010.mpg”帧截图及其在 3 维空间中的投影轨迹

2.3.4 视频哈希的生成 鲁棒的哈希值是整个拷贝检测的关键，本文采用如下的方法。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 $\boldsymbol{p}=\{p_k\}$ 作为密钥，根据式(5)，生成哈希序列 $\boldsymbol{h}=\{h_k\}$。

$$h_k=\begin{cases}1, & |v_k^o\cdot p_k|\ge \mathrm{Th}\\0, & |v_k^o\cdot p_k|<\mathrm{Th}\end{cases},\quad 1\le k\le N\tag{5}$$

其中 Th 为阈值，计算方法如下：

$$\mathrm{Th}=\mathrm{median}(|v_k^o\cdot p_k|),\quad 1\le k\le N\tag{6}$$

3 仿真实验结果和实验分析

本文的实验视频来自于文献[9]的视频数据库，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。把该方法应用从视频数据库中随机抽取的 50 个视频，并计算在各种攻击下的误码率平均值。本文设定一个阈值 $t=0.2$，若待测视频与原始视频的哈希序列相比误码率小于 t，则说明待测视频是原始视频的一个拷贝，阈值的选取参考文献[10]。把得到的实验结果与文献[3]比较，实验比较结果见表 1。本文算法在 AWGN、帧旋转、帧平移和帧丢弃 4 种攻击下的误码率如图 3 所示。

(2)视频检测的精确率测试。本文从视频数据库中选取 10 个视频作为测试视频，并对测试视频进行

表 1 各种攻击下的平均误码率及与文献[3]的比较

攻击方式	误码率(BER)			不同视频间的误码率		
	本文方法	文献[3]中方法		本文方法	文献[3]中方法	
		DCT 方法	RBT 方法		DCT 方法	RBT 方法
AWGN	0.0249	0.029	0.017	0.50	0.5	0.49
帧模糊	0.0121	0.02	0.042	0.47	0.5	0.49
帧旋转	0.0184	0.13	0.22	0.49	0.5	0.48
帧丢弃	0.0407	0.042	0.058	0.53	0.5	0.49
帧平移	0.0168	0.15	0.14	0.51	0.5	0.50
码率改变	0.123	0.043	0.03	0.51	0.51	0.49
放缩	0	0	0	0.51	0.5	0.51

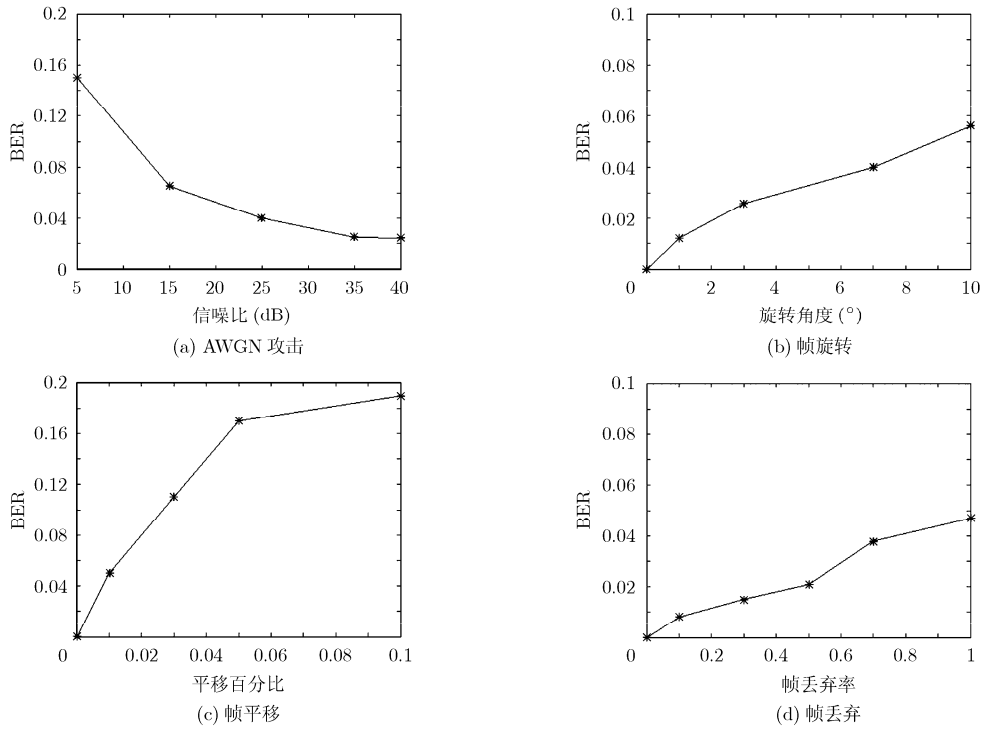


图 3 视频在 AWGN、帧旋转、帧平移和帧丢弃 4 种攻击下的平均误码率曲线

攻击，实验显示仍然能从视频库中搜索到相应的原始视频，实验结果见表 2。

由表 1 可以看出，本文算法几乎在所有给定的攻击条件下，误码率都小于文献[3]中的方法，虽然在帧平移的条件下，性能比文献[3]稍差但仍远小于阈值，故仍可以检测出视频的拷贝。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。图 3 显示了视频在 4 种攻击下的 BER 曲线，可以看出，本文的算法具有较高的鲁棒性。

表 2 视频检测结果

待测视频 (*.mpg)	攻击方式	最小误码率	原始视频 (*.mpg)	是否匹配
anni001	模糊	0.016	anni001	Y
indi001	AWGN	0.023	indi001	Y
boro2_010	帧平移	0.017	boro2_010	Y
moon001	帧旋转	0.024	moon001	Y
win001001	模糊	0.027	win001001	Y
not_in_db	无	0.055	empty	Y
not_in_db	无	0.52	empty	Y
BOR11_005	放缩	0.046	BOR11_005	Y
UGS02_002	码率改变	0.095	UGS02_002	Y
BOR11_002	帧丢弃	0.053	BOR11_002	Y

在表 2 中，第 1 列表示待测视频，第 2 列是待测视频遭受的攻击，第 3 列是待测视频与数据库中某个视频的最小误码率值，第 4 列是数据库中对应于最小误码率的原始视频，最后一列表示待测视频是否找到匹配视频，若找到，则为“Y”，否则为“N”。“not_in_db”表示不在数据库中，由表 2 可以看出，对于待测的 10 个视频，全部匹配正确。

4 结论

本文提出了一种基于局部嵌入映射的视频哈希方法来进行视频拷贝检测。该方法首先利用视频的时空图选取代表帧，然后利用局部嵌入映射把视频代表帧投影到低维空间中，并利用低维空间点的范数来构成哈希序列来进行视频拷贝检测。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。

参 考 文 献

[1] Indyk P, Iyengar G, and Shivakumar N. Finding pirated video sequences on the internet[R]. Technical report, Stanford University, 1999.

[2] Joly A, Buisson O, and Frelicot C. Content-based copy retrieval using distortion-based probabilistic similarity search[J]. *IEEE Transactions on Multimedia*, 2007, 9(2): 293–306.

[3] Coskun B, Sanku B, and Memon N. Spatio-temporal

transform-based video hashing[J]. *IEEE Transactions on Multimedia*, 2006, 8(6): 1190–1208.

[4] 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 Xue-feng, Li Jin-tao, and Zhang Yong-dong. Spatiotemporal video copy detection based on visual perception analyses[J]. *Chinese Journal of Computers*, 2009, 32(1): 108–114.

[5] 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 Yong-dong, Zhang Dong-ming, and Guo Jun-bo, *et al*.. Rapid video copy detection on compressed domain[J]. *Journal on Communications*, 2009, 30(3): 135–140.

[6] Wu Z P, Huan Q M, and Jiang S Q. Robust copy detection by mining temporal self-similarities[C]. ICME 2009, New York City, 2009: 554–557.

[7] 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 Yu-xin, Liu Guang-jie, and Dai Yao-wei, *et al*.. Video copy detection based on local ordinal[J]. *Journal of Computer -Aided Design & Computer Graphics*, 2009, 21(9): 1339–1343.

[8] Roweis S T and Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. *Science*, 2000, 290: 2323–2326.

[9] The origin of the video dataset is MUSCLE-VCD-2007, http://wwwrocq.inria.fr/imedia/civr-bench/index.html.2007.

[10] Job O, Ton K, and Jaap H. Visual hashing of digital video: applications and techniques[C]. Proceedings of SPIE: The International Society for Optical Engineering, 2001, Vol. 4472: 121–131.

聂秀山： 男，1981 年生，讲师，博士，研究方向为视频分析、多媒体信息安全.

刘 琚： 男，1965 年生，教授，博士生导师，研究方向为多媒体信息处理、通信与信号处理.

孙建德： 男，1978 年生，副教授，硕士生导师，研究方向为信息隐藏、多媒体信息处理.

秦丰林： 男，1978 年生，工程师，博士，研究方向为网络信息处理.