

基于短空时变化的鲁棒视频哈希算法

于 晓¹ 聂秀山^{1,3} 马林元² 尹义龙^{1,3}

(山东财经大学计算机科学与技术学院 济南 250014)¹

(山东财经大学实验教学中心 济南 250014)² (山东大学计算机科学与技术学院 济南 250100)³

摘 要 针对互联网相似视频内容检测问题,提出了基于短空时变化的鲁棒视频哈希算法。特征提取和特征量化是该算法的两个关键步骤。在特征提取中,与现有基于时空信息融合的特征提取方法相比,该算法的创新性在于充分利用相邻帧之间局部空域信息的短时变化(简称“短空时变化”)来提取特征。该算法首先构造视频内接球,并以球心为起点对内接球进行划分,获取一系列内接球环,从而捕捉相邻帧的空域信息的短时变化,然后将球环非负矩阵分解系数作为视频内容进行特征表示;在特征量化中,该算法采用改进的曼哈顿量化策略将视频特征映射成二进制的哈希序列,更好地保留了原空间中的近邻关系,提高了量化的准确度。实验结果表明,该算法具有良好的性能。

关键词 视频哈希,时空信息,非负矩阵分解,相近视频检测,曼哈顿哈希

中图分类号 TP181 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.02.014

Robust Video Hashing Algorithm Based on Short-term Spatial Variations

YU Xiao¹ NIE Xiu-shan^{1,3} MA Lin-yuan² YIN Yi-long^{1,3}

(School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China)¹

(Experimental Teaching Center, Shandong University of Finance and Economics, Jinan 250014, China)²

(School of Computer Science and Technology, Shandong University, Jinan 250100, China)³

Abstract A robust video hashing algorithm based on short-term spatial variations was proposed to detect near-duplicate videos in the Internet. Feature extraction and feature quantization are key steps in this algorithm. In the feature extraction phase, compared to the existing feature extraction methods based on temporal and spatial information fusion, the innovation of the proposed algorithm is to make full use of short-time variations of local spatial information between adjacent frames (referred to “short-term spatial variations”). In the proposed algorithm, inscribed spheres of the video are constructed first, and then a series of spherical tori are obtained by partitioning the inscribed spheres with the center of the sphere as the starting point to capture short-term changes in spatial information between adjacent frames. After that, the decomposition coefficients by non-negative matrix factorization of spherical tori are used as the feature representation of the video. In the feature quantization phase, to map the feature representation into binary hash sequences, the optimized Manhattan hashing strategy is adopted which better reserves the neighborhood structure in the original data space, and thus improves the accuracy of quantization. Experiments were carried out on a video dataset to evaluate the performance of the proposed video hashing method. Experimental results show that the proposed algorithm has good performance.

Keywords Video hashing, Spatio-temporal information, Nonnegative matrix factorization, Near-duplicate video detection, Manhattan hashing

1 引言

近年来,随着云计算和大数据技术的飞速发展以及智能

终端的普及,越来越多的用户开始在互联网上发布、搜索、观看和分享视频,从而导致互联网视频数量呈指数级增长^[1-2]。视频内容的丰富性和各种应用服务的普及也带来了一系列的

到稿日期:2017-04-18 返修日期:2017-05-29 本文受国家自然科学基金项目(61671274),中国博士后科学基金项目(2016M592190),山东省高等学校科技计划项目(J17KB161),山东省高等学校优势学科人才团队培育计划资助。

于 晓(1981—),女,硕士,讲师,主要研究方向为机器学习与人工智能等;聂秀山(1981—),男,博士,教授,CCF 会员,主要研究方向为机器学习及其应用、多媒体信息处理,E-mail:nixsh@sdufe.edu.cn(通信作者);马林元(1980—),男,硕士,主要研究方向为信息管理与信息系统;尹义龙(1972—),男,博士,教授,主要研究方向为机器学习及其应用。

问题,例如,同一内容会有多个视频版本在网上发布,部分用户重复上传互联网上已有的视频内容或非法修改后再上传,这些问题都会降低视频搜索的效率,同时在某种程度也侵害了视频原创者的版权。因此,如何从大量视频中识别并去除相似视频进而提高检索的效率,已成为当前多媒体内容分析领域的研究热点。相似视频检测在版权保护^[3]、视频监控^[4-5]、视频搜索重排序^[6]、视频推荐^[3]等方面有着广泛的应用前景^[7]。鲁棒视频哈希技术正是解决相似视频内容检查的核心技术。鲁棒哈希技术首先提取视频特征,构造视频内容的低维表示形式,然后对内容特征进行量化,以二值的哈希序列代表视频,通过对比视频哈希的相似度来判断两个视频是否为相似视频。

早期的视频哈希算法大多是基于视频帧特征的,但是视频的本质是一种时间序列,因此,近年来基于视频时间和空间信息融合的方法逐渐成为研究的热点。但是,现有算法在利用时空信息的过程中,大多使用的是帧与帧之间平行的时间变化,而视频时空信息的变化蕴含更丰富的信息。为充分利用视频空域和时域信息的变化,使用视频内接球环来捕捉相邻帧之间局部信息的短时变化(本文称之为“短空时变化”),如图 1 所示。“短空时变化”同时考虑了视频的空域和时域变化,而这种变化正是视频内容结构的重要体现。在特征量化阶段,传统的视频哈希法在特征量化阶段大多采取了单比特的量化策略,即首先设定一个阈值,若某特征维度小于该阈值,则该维度哈希值为“0”,反之“1”。此类量化策略容易破坏原始向量空间的邻域结构。为克服这一缺点,Kong 等人^[8]提出了一种曼哈顿量化策略,其基于曼哈顿距离来度量哈希序列的距离,但是该量化策略在编码时忽略了类别编号本身的距离关系。本文基于曼哈顿量化策略,对 k 均值算法得到的类别编号进行了优化调整,提高了视频哈希检测的精度。

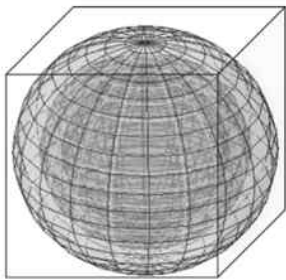


图 1 将视频立方体划分为球环的示意图

Fig. 1 Schematic diagram of dividing video cube into ball ring

2 相关工作

视频哈希一般先提取视频特征,构造视频内容的低维表示形式;然后对内容特征进行量化,以二值的哈希序列表示视频内容,通过对比视频哈希序列的相似度来判断两个视频内容的相似性。视频哈希序列具备两个特征:鲁棒性和区分性^[9]。鲁棒性指的是,具有相似内容的视频在遭受非内容改变的攻击(如几何变换、噪声、对比度等)时,视频哈希序列比

较相似。区分性指的是,哈希序列能够区分不同内容的视频,即不同内容的视频,其哈希序列也不同。

在视频特征提取方面,早期的研究大部分是基于视频帧特征的。这类方法把视频的每一帧看成图像,对图像进行哈希,再把图像的哈希结果连接到一起,其中包含时间信息。Roover 等人^[10]提出了基于放射投影的哈希方法(简称为 Rash),在沿着中心点的一系列线上计算像素值的变化情况,然后用离散余弦变换来得到最终的哈希值。Rash 对旋转攻击的鲁棒性较好。Lee 等人^[9,11-12]提出了基于面向梯度中心点的视频哈希方法 CGO,其主要思想是利用相邻像素值的差异来提取方向信息,将视频中的每一帧分成若干块,把每块像素的角度信息结合起来形成一个均值中心点,再结合从每一帧中提取出来的中心点信息形成最终的视频哈希。针对视频压缩、加噪声的情形,该方法的鲁棒性比较好。但是,由于 CGO 哈希依赖于局部角度信息,在类似于旋转这种几何攻击情形下,CGO 哈希的性能一般。

基于视频帧特征的哈希方法计算比较简单,但是在受到类似帧率改变这种时间攻击时容易造成哈希值的不唯一性。Chen^[13]提出了一种基于时间序数测量的视频序列匹配方法,将每一帧划分为网格,把对应网格沿着时间序列进行排序,描述出了视频时间变化的全局信息和局部信息。该方法对长视频比较有效,对短视频效果不佳,因为短视频没有足够的时间信息用于与其他视频进行区分。

近年来,人们开始研究融合视频的时域信息和空域信息来构造视频哈希。其中较有代表性的方法是基于三维的离散余弦变换法 3-D DCT^[14],该方法对视频序列的亮度信号进行三维 DCT 变换,选择低频系数来形成哈希向量。Esmaili 等人^[15-16]从视频序列中提取信息形成时间的代表性图像(TI-RIs),然后对这些图像进行离散余弦变换来得到哈希值。这些时间-空间信息法虽然能提高抵抗时间攻击时鲁棒性,但是在面对几何攻击时性能一般。Li 等人^[17]提出了基于张量的低秩逼近的方法(简称为 LRTA),该方法把视频片段建模成三维张量,使用多线性子空间投影来构造哈希编码,在投影中通过分析平行因子来进行张量的低秩逼近,并通过把视频划分成随机选择的重叠子立方体,在哈希函数中引入了随机性。Li 还提出了两阶段的哈希方法^[18]:在第一阶段,使用动态时间规整算法来进行自动同步,并基于流哈希对视频进行比较;在第二阶段,使用距离提升机制将动态时间规整和流哈希提取的信息融合在一起。

但是,现有算法在对时空信息的利用过程中并没有充分考虑视频空间和时间局部信息的同时变化,而这种场景和信息的局部变化正是视频内容结构的重要体现。因此,本文利用视频内接球环来捕捉视频相邻帧之间局部信息的短时空域信息变化情况,并构造时空图来表示视频特征。

在特征量化阶段,大部分的视频哈希算法都采取了单比特策略。典型的做法是选取一个阈值,如将特征值的中位数^[14,16,19]设为阈值,若特征分量小于阈值,则该分量记为“0”,

反之记为“1”^[20],并使用海明距离来进行相似度的衡量。但此类量化策略具有一定的局限性,即当两个相似的特征分量刚好位于阈值两侧附近时,会使得相似的视频特征分量被量化为不同的比特。此外,海明距离在某种程度上会破坏原始向量空间的邻域结构。为此,Kong等人^[8]提出了曼哈顿量化策略,该策略从一定程度上解决了海明距离的局限性。曼哈顿哈希算法的主要思想是使用 k 均值算法来对每个投影后的维度聚类,使用多位的自然二进制码对每一维度进行编码,并基于曼哈顿距离来进行相似度的比较。但是曼哈顿量化策略直接使用聚类后得到的类别编号来进行编码和比较,没有考虑类别编号本身的距离关系。本文对曼哈顿哈希算法进行了改进,对类别编号进行了优化,使得编码更好地体现了原向量空间的邻域结构。实验结果表明,改进方法的性能要优于曼哈顿哈希法。

3 基于短空时变化的鲁棒视频哈希算法

基于短空时变化的鲁棒视频哈希算法主要分为两部分:特征提取和特征量化,具体流程如图2所示。特征提取部分包括预处理、球环分割和非负矩阵分解。首先对视频进行预处理,从视频中采集出所需要的帧,得到标准化的视频;接着对标准化的视频进行球环分割,并将球环投影为一幅时空图;然后使用非负矩阵分解对投影得到的时空图进行降维,用分解后得到的系数代表视频的特征。特征量化部分主要是使用改进后的曼哈顿哈希算法 OMH 对降维后的内容进行哈希,使用得到二进制的哈希码来表示视频,并使用二进制的哈希序列进行视频的匹配。

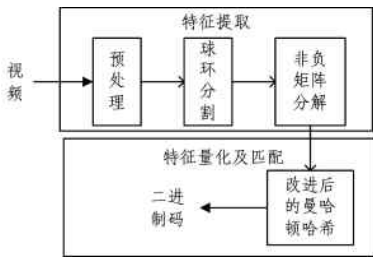


图2 算法框架图

Fig. 2 Framework of the proposed method

3.1 基于短空时变化的特征提取

视频相邻帧的局部空域信息变化隐含了丰富的视频内容信息,本文利用视频内接球环来获取这类变化信息,具体方法如下。

3.1.1 视频内切球环构造

视频内切球环的构造已在本项目组前期的工作^[21]中进行了详细的分析,为更好地说明本文算法,本部分首先对内切球环的构造过程进行简要说明。

对于给定的视频片段,可以将视频分割成若干个立方体,并找出每个立方体的内切球,然后把内切球按体积大小平均分为 n 个球环。将每个球环中所有点的亮度值投影成一幅图像的一列,最终每个立方体都被投影为一幅时空图,如图3所示。

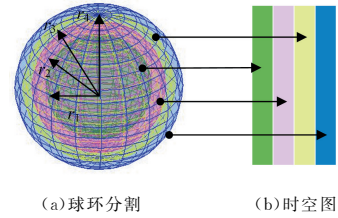


图3 时空图构建示意图

Fig. 3 Structure diagram of spatio-temporal graph

设视频立方体的大小为 $D \times D \times D$,球环的数量为 n , U_k 表示第 k 个球环的像素点集合, Y_k 是 U_k 中像素点所对应的亮度值, r_k 表示第 k 个以内切球中心点为球心的球半径。

球环可以通过球的半径以及像素点和内切球中心点的距离计算得到。 r_n 表示内切球的半径,也就是 $D/2$,根据式(1)可算出内切球的体积,根据式(2)可计算出每个球环的体积 A 。

$$V = \frac{4}{3} \pi r_n^3 \quad (1)$$

$$A = \lfloor V/n \rfloor \quad (2)$$

第一个球环(实际上是个小球)的半径 r_1 可通过式(3)计算:

$$r_1 = \sqrt[3]{\frac{3A}{4\pi}} \quad (3)$$

根据式(4),在已知 r_{k-1} 的情况下可计算出 r_k 。

$$r_k = \sqrt[3]{\frac{3A}{4\pi} + r_{k-1}^3} \quad (4)$$

其中, $1 < k \leq n$ 。

设 $p(x, y, z)$ 表示视频第 y 行、第 x 列、第 z 帧的像素值, (x_c, y_c, z_c) 表示内切球中心点的位置。如果 D 是偶数, $x_c = y_c = z_c = D/2 + 0.5$;如果 D 是奇数, $x_c = y_c = z_c = (D+1)/2$ 。像素点 $p(x, y, z)$ 与中心点 (x_c, y_c, z_c) 的距离 d 的计算公式如下:

$$d = \sqrt{(x-x_c)^2 + (y-y_c)^2 + (z-z_c)^2} \quad (5)$$

在计算出所有的半径以及每个像素点和内切球中心点的距离后,对距离和半径进行比较,若该像素点离中心点的距离在 r_{k-1} 和 r_k ($1 < k \leq n$)之间,则该像素点位于第 k 个球环,并根据式(6)和式(7)将该像素点存入 U_k 中。

$$U_1 = \{p(x, y, z) \mid d \leq r_1\} \quad (6)$$

$$U_k = \{p(x, y, z) \mid r_{k-1} < d \leq r_k\} \quad (7)$$

识别出第 k 个球环内的像素点后,将该球环内所有像素点的亮度值存到列向量 S_k 中,并对其进行升序排序,从而得到时空图矩阵 S ,如式(8)所示:

$$S = [S_1, S_2, \dots, S_n] \quad (8)$$

3.1.2 非负矩阵分解(NMF)

视频中每个立方体得到的时空图矩阵 S 是高维度矩阵。为了提高算法的效率,使用NMF对 S 进行降维处理。

非负矩阵分解是一种降低维度的有效方法。对任一非负矩阵 $S = \{S_{i,j}\}_{n \times m}$,可以将其分解成非负矩阵 $W \in R^{n \times r}$ 和 $H \in R^{r \times m}$,满足公式 $S \approx WH$,其中 W 是基矩阵, H 是系数矩阵。时空图矩阵 S 经过非负矩阵分解后,用系数矩阵 H 代替 S ,视频中的每个立方体都用 H 表示,将每个立方体的 H 进

行连接,一个视频用一个实数向量 B 来表示。第 i 个视频的特征矩阵表示如式(9)所示:

$$B_i = (b_1, b_2, \dots, b_m)^T \quad (9)$$

3.2 特征量化

特征量化,即哈希码构造,是视频哈希的重要步骤。传统的单比特或双比特编码及海明距离度量具有一定的局限性,Kong 等人^[8]提出的曼哈顿哈希编码的思路在一定程度上突破了这一局限性。曼哈顿哈希算法的主要步骤如下(详细内容请参考文献[8]):

- 1) 选择一个整数 q ;
- 2) 选择已有投影方法或设计新的投影方法对原空间的样本进行降维处理,得到实数表示;
- 3) 使用 k 均值法聚类,得到 2^q 个类别,并根据中心点得到类别的阈值;
- 4) 基于阈值对实数进行分类,并量化成 q 位的二进制表示。量化后的两个二进制数 x 和 y 之间的距离定义为:

$$d_m(x, y) = \sum_{i=1}^q d_d(x_i - y_i) \quad (10)$$

其中, d_d 表示两个二进制数的十进制距离。

曼哈顿哈希算法直接利用聚类后的分类结果进行二进制编码。在计算距离时使用了式(10),即二进制数的十进制距离。因为二进制编码是由 k 均值算法聚类时得到的类别编号决定的,所以类别编号会影响到距离的准确性。由于聚类时得到的类别编号是随机分配的,因此类别编号的距离无法体现出类和类之间的距离。本文对曼哈顿哈希算法进行了改进,主要是对类别编号策略进行了优化,使得聚类后距离远的类所对应的类别编号的距离较大,距离较近的类之间的类别编号的距离较小。

类别编号优化的主要思想如下:

- 1) 根据聚类后得到的中心点位置矩阵 C , 得到分类数量和原类别的编号;
- 2) 计算 C 中任意两个中心点的距离,并找出其中距离最大的两个中心点;
- 3) 将这两个中心点的新类别编号分别设为 0 和新类别编号的最大值,并记录原类别编号和新类别编号的对应关系;
- 4) 从新类别编号 0 出发,在尚未分配新类别编号的中心点中找到离它距离最近的中心点,并将该中心点的新编号设为 1;
- 5) 再从 1 出发,按照步骤 4) 中的方法找到新编号 2, 依次类推,直到分配完所有的新类别编号为止。

改进后的曼哈顿哈希算法见算法 1。

算法 1 改进后的曼哈顿哈希算法 (Optimized Manhattan Hashing, OMH)

输入: 正整数 q , 原空间样本降维后得到的实数矩阵训练集 B_{Train} 和测试集 B_{Test}

输出: 数据的二进制表示 KB_{TestB}

1. 使用 k 均值法对 B_{Train} 中的向量进行聚类,将其分成 2^q 个类别;
2. 根据聚类后得到的中心点矩阵获得阈值;
3. 基于阈值,对 B_{Test} 数据进行分类;

4. 使用类别编号优化算法对原类别编号进行优化;
5. 将 B_{Test} 数据分类结果转换为新类别编号的分类结果 KB_{TestB} ;
6. 将 KB_{TestB} 中的数据量化成 q 位的二进制表示 KB_{TestB} 。

视频经过球环分割、非负矩阵分解和改进后的曼哈顿哈希算法,最终得到一串二进制的哈希序列。利用式(10)计算距离,并对这些距离进行比较,计算出视频之间的距离。

以 $q=3$ 为例,每个类别都可以用一个 3 位的二进制数表示,一个视频就可以利用二进制数的哈希向量来表示。若视频哈希向量 $KB_1^T = (001, 110, 010)$, $KB_2^T = (011, 010, 001)$, 则 KB_1 就可以用 001110010 来表示, KB_2 用 011010001 来表示。

$$\begin{aligned} d_m(KB_1, KB_2) &= d_m(001110010, 011010001) \\ &= d_d(001, 011) + d_d(110, 010) + \\ &\quad d_d(010, 001) \\ &= 2 + 4 + 1 = 7 \end{aligned}$$

即两个视频哈希向量 KB_1 和 KB_2 之间的距离为 7。

4 实验与分析

为验证所提算法的性能,将其与现有视频哈希算法的性能进行了比较;进一步地,为体现本文对曼哈顿哈希算法的改进,在现有数据集上,将其与原有曼哈顿哈希算法进行了性能比较。实验性能用漏检率 (Miss probability) 和虚警率 (False alarm probability) 的 ROC 曲线来表示。

实验所用的所有视频数据均是从 CC_WEB_VIDEO (video.cs.cityu.edu.hk/webvideo) 和 OV (www.openvideo.org) 数据集中得到的。原始视频共有 1181 个,我们采用相关软件工具对原始视频进行几何变换、后期处理等攻击。在特征量化阶段,本实验使用 945 个视频进行训练,用其余的 236 个视频进行测试,参数设置如表 1 所列。

表 1 实验参数设置

Table 1		Experimental parameter setting
序号	攻击类型	参数设置
a	加白噪声	均值为 0, 方差为 0.005
b	加标志	尺寸为 10×10
c	随机丢弃帧	从原视频中随机丢弃 6 帧
d	加白噪声+模糊	白噪声的均值为 0, 方差为 0.005; 高斯低通滤波的尺寸为 3×3 , 滤波器的标准值为 0.5
e	加标志+白噪声+模糊	标志的尺寸为 10×10 ; 白噪声的均值为 0, 方差为 0.005; 高斯低通滤波的尺寸为 3×3 , 滤波器的标准值为 0.5
f	画中画+白噪声	插入的图像大小为 10×10 ; 白噪声的均值为 0, 方差为 0.005

4.1 本文算法与现有算法的性能比较

将提出的算法 ST-OMH 与其他基于时空信息融合的视频哈希算法进行了比较,即分别与基于球环分割法 (Spherical Torus-based video hashing, ST)^[21]、基于三维的离散余弦变换法 (3D-Discrete Cosine Transform, 3D-DCT)^[14]、基于张量的低秩逼近法 (Low-Rank Tensor Approximations, LR-TA)^[17] 等方法进行了鲁棒性方面的对比,实验结果如图 4 所示。

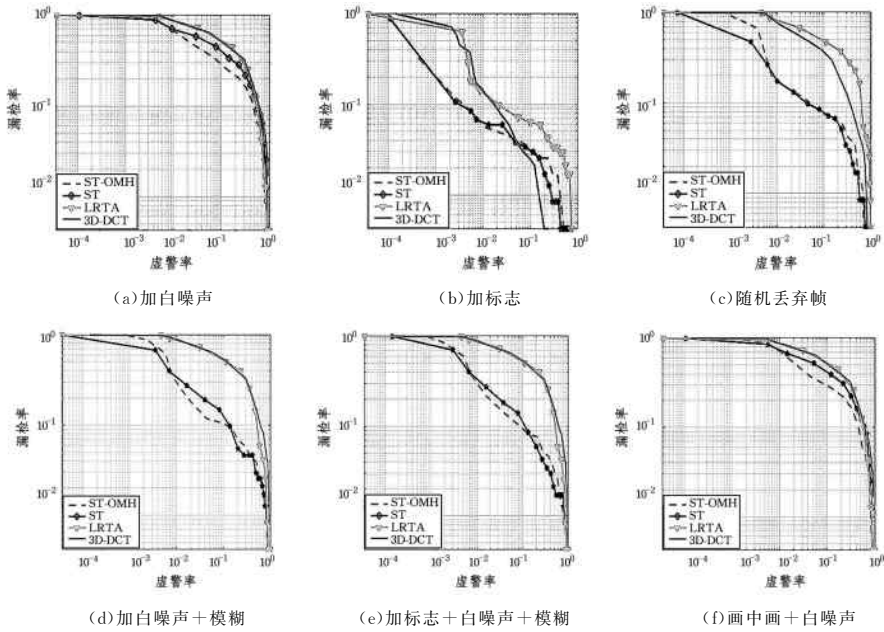


图4 ST-OMH与其他算法的ROC曲线

Fig 4 ROC curves of ST-OMH and other video hashing algorithms

从图4中可以看出,本文算法在大多数攻击下都取得了较好的性能。其中在图4(b)、图4(c)和图4(e)中,ST-OMH和ST的结果相差不多。原因在于,图4(b)和图4(e)中都有加标志的攻击,在攻击时对每一个帧都加了同样的标志,无形中增加了原始视频的相似性,哈希量化使得检索性能的提升被这种相似性抵消了,因此性能相差不多;图4(c)是丢帧攻击,丢帧后利用插值的方式又插入了一些帧,这也无形中增加了本来不同内容视频的相似性。但是,ST-OMH算法采用哈希值来表示视频内容,提高了检索的效率,缩减了存储空间,对大数据环境下的视频检索具有积极的意义。

4.2 改进的哈希量化策略与原曼哈顿哈希算法的性能比较

为了验证OMH算法的性能,将基于曼哈顿哈希的算法ST-MH与基于改进的曼哈顿哈希的算法ST-OMH进行了比较,实验结果的ROC曲线如图5所示。实验结果表明,相

比于基于曼哈顿哈希的算法ST-MH,基于改进的曼哈顿哈希算法ST-OMH的性能有了提升,特别是在图5(b)和图5(e)中,因为改进的曼哈顿哈希算法对类别编号进行了优化,使得k均值算法聚类后得到的类别编号更好地体现了原数据空间的近邻关系,从而提高了哈希的准确率。由于ST算法本身在一些攻击中就会造成一定的漏检率和虚警率,因此图5(c)中的性能对比一般,但在大部分的攻击中,ST-OMH的性能都明显高于ST-MH。进一步,对ST-OMH和ST-MH进行量化比较,结果如表2所列。在表2中,X代表图5中各子图的横轴,即虚警率;Y代表图5中各子图的纵轴,即漏检率。从表2中可以看出,在不同的攻击下,本文提出的基于改进的曼哈顿哈希的算法ST-OMH与基于曼哈顿哈希的算法ST-MH相比,平均性能有了明显提升。同时,对实验结果的标准差进行对比,结果表明本文提出的ST-OMH算法的性能结果更集中。

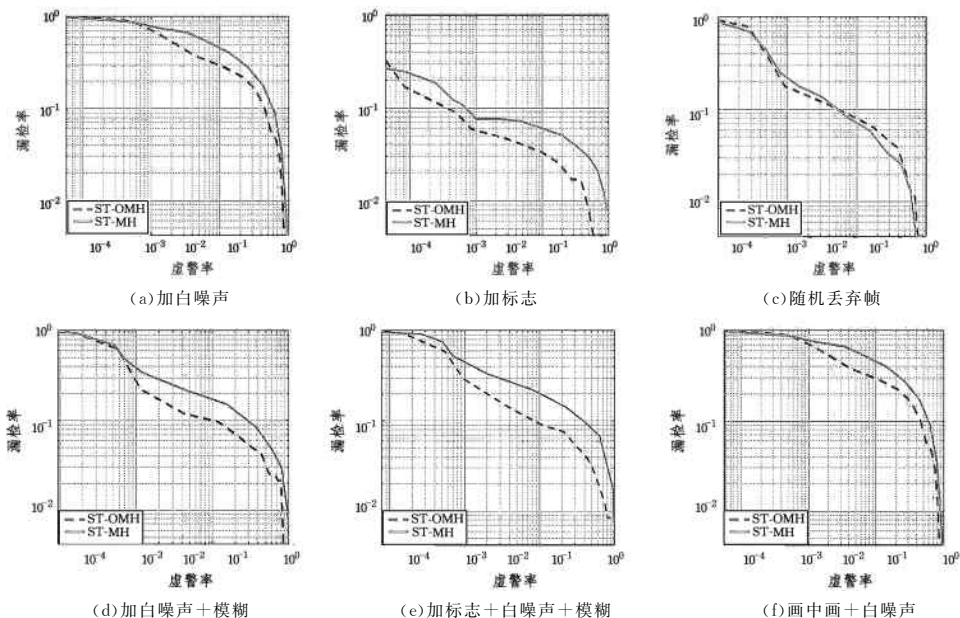


图5 ST-OMH与ST-MH的ROC曲线

Fig 5 ROC curves of ST-OMH and ST-MH

表2 OMH算法的数据分析

Table 2 Data analysis of OMH

攻击序号	算法	均值		标准差	
		X	Y	X	Y
a	ST-OMH	0.8941	0.05343	0.2713	0.1876
	ST-MH	0.8986	0.07148	0.2856	0.2285
b	ST-OMH	0.8993	0.0080903	0.2654	0.03824
	ST-MH	0.8868	0.01377	0.2978	0.04502
c	ST-OMH	0.9010	0.02623	0.2638	0.1256
	ST-MH	0.8880	0.036787	0.2900	0.1489
d	ST-OMH	0.9052	0.03233	0.2585	0.1489
	ST-MH	0.8994	0.05822	0.2868	0.2028
e	ST-OMH	0.9057	0.03267	0.2589	0.1493
	ST-MH	0.9121	0.05051	0.2680	0.1882
f	ST-OMH	0.9066	0.04822	0.2575	0.1817
	ST-MH	0.9029	0.06733	0.2771	0.2182

结束语 本文提出了一种基于短空时信息变化的鲁棒视频哈希算法 ST-OMH,该算法可被用于相似视频检测、版权保护等领域。在特征提取方面,本文算法利用视频内切球环有效地将视频的时空信息融合在一起。在特征量化方面,为提高量化的性能,引入曼哈顿哈希算法,并对曼哈顿哈希算法进行了改进,对编码时使用的类别编号进行了优化,更好地保留了原数据空间的邻域结构。实验结果表明,与现有视频哈希算法以及原曼哈顿哈希量化策略相比,本文算法的性能有了显著提升。在未来的研究中,将进一步在大型视频数据库上对本文算法进行测试,并对算法在特征表示和量化策略上进行改进。

参考文献

- [1] TAN H K, NGO C W, HONG R, et al. Scalable detection of partial near-duplicate videos by visual-temporal consistency[C]// Proceedings of the 17th ACM International Conference on Multimedia. ACM, 2009: 145-154.
- [2] WU X, HAUPTMANN A G, NGO C W. Practical elimination of near-duplicates from web video search[C]// Proceedings of the 15th ACM International Conference on Multimedia. ACM, 2007: 218-227.
- [3] CHERUBINI M, OLIVEIRA R D, OLIVER N. Understanding near-duplicate videos: a user-centric approach[C]// Proceedings of the 17th ACM International Conference on Multimedia. ACM, 2009: 35-44.
- [4] HUANG Z, SHEN H T, SHAO J, et al. Practical online near-duplicate subsequence detection for continuous video streams [J]. IEEE Transactions on Multimedia, 2010, 12(5): 386-398.
- [5] SHEN H T, ZHOU X F, HUANG Z, et al. UQLIPS: a real-time near-duplicate video clip detection system[C]// Proceedings of the 33rd International Conference on Very Large Data Bases. VLDB Endowment, 2007: 1374-1377.
- [6] WU X, HAUPTMANN A G, NGO C W. Practical elimination of near-duplicates from web video search[C]// Proceedings of the 15th ACM International Conference on Multimedia. ACM, 2007: 218-227.
- [7] LIU J J, HUANG Z, CAI H Y, et al. Near-duplicate video retrieval[J]. ACM Computing Surveys, 2013, 45(4): 1-23.
- [8] KONG W H, LI W J, GUO M Y. Manhattan hashing for large-scale image retrieval[C]// Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2012: 45-54.
- [9] LEE S, YOO C D. Robust video fingerprinting for content-based video identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18(7): 983-988.
- [10] ROOVER C D, VLEESCHOUWER C D, LEFÈBVRE F, et al. Robust video hashing based on radial projections of key frames [J]. IEEE Transactions on Signal Processing, 2005, 53(10): 4020-4037.
- [11] LEE S, YOO C D. Video fingerprinting based on centroids of gradient orientations[C]// 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings. IEEE, 2006, 2: II.
- [12] LEE S, YOO C D, KALKER T. Robust video fingerprinting based on symmetric pairwise boosting[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2009, 19(9): 1379-1388.
- [13] CHEN L, STENTIFORD F W M. Video sequence matching based on temporal ordinal measurement[J]. Pattern Recognition Letters, 2008, 29(13): 1824-1831.
- [14] COSKUN B, SANKUR B, MEMON N. Spatio-temporal transform based video hashing[J]. IEEE Transactions on Multimedia, 2006, 8(6): 1190-1208.
- [15] ESMAEILI M M, WARD R K. Robust video hashing based on temporally informative representative images[C]// Proceedings of the 32nd International Conference on Coastal Engineering Consumer Electronics (ICCE 2010). Shanghai, China, 2010: 179-180.
- [16] 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.
- [17] LI M, MONGA V. Robust video hashing via multilinear subspace projections[J]. IEEE Transactions on Image Processing, 2012, 21(10): 4397-4409.
- [18] LI M, MONGA V. Twofold video hashing with automatic synchronization[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(8): 1727-1738.
- [19] SONG J K, YANG Y, HUANG Z, et al. Effective multiple feature hashing for large-scale near-duplicate video retrieval[J]. IEEE Transactions on Multimedia, 2013, 15(8): 1997-2008.
- [20] LI W J, ZHOU Z H. Learning to hash for big data: current status and future trends[J]. Chinese Science Bulletin, 2015, 60(5/6): 485-490. (in Chinese)
李武军, 周志华. 大数据哈希学习: 现状与趋势[J]. 科学通报, 2015, 60(5/6): 485-490.
- [21] NIE X S, CHAI Y E, LIU J, et al. Spherical torus-based video hashing for near-duplicate video detection[J]. Science China Information Sciences, 2016, 59(5): 1-3.