

# 融合时空域变化信息的视频感知哈希算法研究

文振<sup>1</sup>,高金花<sup>2</sup>,朱映映<sup>1</sup>,杜以华<sup>1</sup>,邓良太<sup>1</sup>

(1.深圳大学计算机与软件学院,广东深圳 518060;2.深圳信息职业技术学院,广东深圳 518172)

**摘要:** 感知哈希(Perceptual Hashing)是多媒体数据集到摘要集的单向映射,为多媒体数字内容的标识、检索、认证等应用提供了安全可靠的技术支撑.本文提出一种融合视觉感知及时空域特征的视频感知哈希算法.算法首先对视频序列每一帧进行随机可重叠分块,并计算每个分块以像素为单位的亮度均值,在某一步长下,以同一帧的分块亮度差作为视频帧空域特征,以不同视频帧相同位置的分块亮度差作为时域特征,通过哈希量化得到时空域感知哈希,通过时空域感知哈希融合,最后得到简洁的视频唯一标识——摘要哈希.实验结果表明,该算法表现出较好的鲁棒性与区分性,通过相似度拟合图算法分析,可以实现视频篡改的准确检测及定位.

**关键词:** 视觉感知; 时空域; 感知哈希; 版权保护

**中图分类号:** TN911.23

**文献标识码:** A

**文章编号:** 0372-2112 (2014) 06-1163-05

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

**DOI:** 10.3969/j.issn.0372-2112.2014.06.019

## Video Perceptual Hashing Fusing Spatiotemporal Change Detection

WEN Zhen-kun<sup>1</sup>, GAO Jin-hua<sup>2</sup>, ZHU Ying-ying<sup>1</sup>, DU Yi-hua<sup>1</sup>, DENG Liang-tai<sup>1</sup>

(1. College of Computer and Software, Shenzhen University, Shenzhen, Guangdong 518060, China;

2. Shenzhen Institute of Information Technology, Shenzhen, Guangdong 518172, China)

**Abstract:** Perceptual hashing is a one-directional mapping from multimedia data sets to the abstract sets, which provides a safe and reliable tool for identification, retrieval, and authentication of multimedia digital contents. This paper proposes a video perceptual hashing algorithm fusing visual perception and spatiotemporal features. The algorithm first randomly divides each video frame into a number of overlapping blocks, and then calculates the average brightness (AB) of each block. While spatial features are described by differences of the AB among blocks in the same frame, the features in time domain is represented by the differences of the same block across time. Finally, the two features are quantized and fused to generate the unique id, i.e. abstract hashing for each video. The experimental results show that the features are robust, discriminative, and can effectively detect and accurately locate video tampering using correlation analysis.

**Key words:** visual perception; time airspace; perceptual hash; copyright protection

## 1 引言

视频感知哈希是指视频数据集到感知摘要集的一种单向映射,通过符合人眼感知机理的特征提取方法得到视频内容标识特征序列,然后量化得到摘要哈希.因此,感知哈希除了具有单向性与摘要性之外,还应满足以下三个特征<sup>[1]</sup>:

(1)鲁棒性:对于感知内容相同或相似的多媒体数据,感知哈希能够得到相同或相似的摘要序列.

(2)可区分性:对于感知内容完全不同的多媒体信

息,感知哈希能够得到完全不同或差异巨大的摘要序列.

(3)安全性:算法能否有效的抵抗非法模拟攻击也是重要的衡量指标.

近年来,图像感知哈希函数的研究主要集中在感知哈希的安全性和鲁棒性上,例如王丽娜等提出的应用Gabor滤波器的鲁棒图像感知哈希算法<sup>[2]</sup>;朱国普等提出改进的哈希值随机性定量分析<sup>[3]</sup>;Fouad Khelifi等提出的基于非负矩阵分解的图像感知哈希安全性分析<sup>[4]</sup>;视频感知哈希算法,以 Baris Coskun 的3D-DCT框架为

典型<sup>[5]</sup>,其主要思想是对视频序列归一化预处理后,用DCT变换从三维角度构造视频帧图像 Tube 通道,然后提取感知特征.张惠等人在2008年发表的论文率先将人类视觉系统结合到视频感知哈希算法模型<sup>[6]</sup>.Kim等提出计算视频帧亮度的有序序列作为空间信息,时间信息由连续视频帧之间亮度的变化产生,然后联合视频的时间信息和空间信息作为视频特征<sup>[7]</sup>.

结合当前研究的热点和不足,本文提出一种融合视频时空域感知变化信息的视频感知哈希算法.实验结果显示,算法得到的摘要哈希序列具有更加平衡的鲁棒性能与区分性能,能够为视频内容标识、视频篡改检测提供理论支撑.

## 2 视频感知哈希算法整体框架

- (1) 通过视频帧灰度化得到灰度视频序列.
- (2) 对视频信息进行随机可重叠分块预处理.
- (3) 计算亮度平均值,作为改视频帧分块标识.
- (4) 统计视频帧的空域特征、视频时域特征.
- (5) 采用量化方案对特征矩阵进行哈希量化.

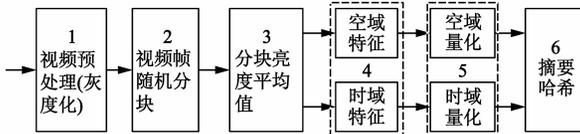


图1 算法整体框架

## 3 随机可重叠分块

将视频每一帧看成由其内部局部区域构成的集合  $I = \cup_{i=1}^M B_i$ , 其中  $M$  表示可能被关注的局部区域的个数. 每个局部区域可以使用一个简单的矩形块  $B_i = I((x_i, y_i), (x_i + w_i, y_i + h_i))$  来表示, 鉴于此, 我们使用一个由随机种子  $seed$  控制的伪随机函数  $R_{seed}$  来模拟视觉选择的随机性.

$$\Psi = \{B_j\}_{j=1}^N \xrightarrow{R_{seed}} \{B_i\}_{i=1}^M = I \quad (1)$$

其中,  $N \ll M$ ,  $\Psi \subset I$  表示特定观察者在特定时刻观察视频时基于视角  $seed$  所选择的感知区域.

## 4 时空域变化信息的视频特征提取

### 4.1 空域特征

由人类视觉系统(HVS)理论可知,人对视频信息的认知过程可以看作一个自顶向下的过程.首先获得视频帧的总体观感,然后进一步落实到由认知目的决定的细节信息上,从而形成个人的主观理解.因此感知特征的提取要重点探索视频空域变化信息与时域变化信息两个方面.

针对视频序列每一帧进行空域特征描述,在随机

种子的控制下产生  $m$  个随机的矩形块.各个分块以互相交叠的方式散布在整个帧图像上,大小同样随机产生.对分块进行编号,通过步长确定亮度差计算的两个分块,然后统计帧的分块亮度均值差作为空域信息.针对长度为  $n$  (帧)的视频序列,空域特征提取过程如图2所示.

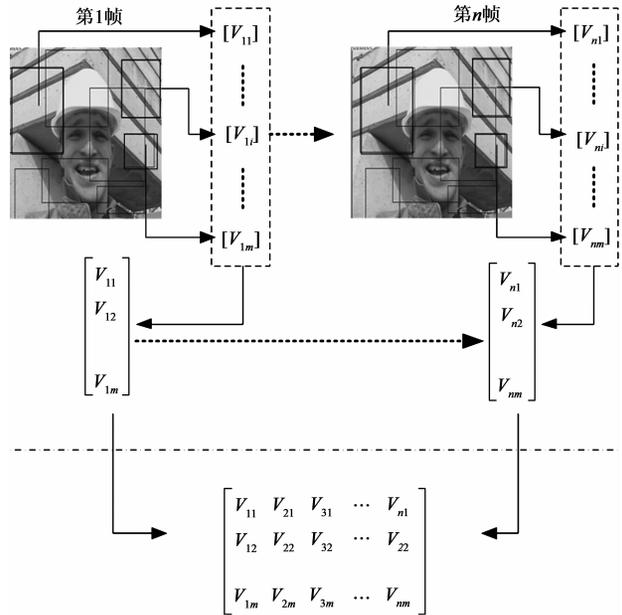


图2 空域特征描述

如图2所示,针对第  $i$  个分块,计算亮度平均值:

$$V_i = \frac{\sum_{j=1}^K \text{gray}(j)}{K} \quad (2)$$

其中,  $K$  表示该分块中像素个数,  $\text{gray}(j)$  表示第  $j$  个像素的亮度值,  $V_i$  表示第  $i$  分块的亮度平均值.

针对视频每一帧,可以得到亮度平均值向量  $[V_1, V_2, V_3, \dots, V_m]$ .在此基础上,确定步长集合  $[S_1, S_2, S_3, \dots, S_p]$ , 其中  $1 \leq S_p \ll m (1 \leq \partial \leq p)$ , 针对步长  $S$  的计算公式如下:

$$d_{(i, i+s)} = V_{i+s} - V_i \quad (3)$$

针对视频长度为  $l$  的视频序列,得到空域特征矩阵.

### 4.2 时域特征

针对视频时域画面变化信息,依然采用上述随机分块策略构建时域管道,分别计算每个视频分块的亮度均值,得到时域分块管道亮度均值为  $[V_{1j}, V_{2j}, V_{3j}, V_{lj}]$ .

时域特征提取过程步骤如下:

针对每个时域管道,构建步长集,针对步长  $T$ , 计算公式如下:

$$d_{t(i, i+T)} = V_{j(i+T)} - V_{jt} \quad (4)$$

通过时域亮度均值差的计算,得到每一个时域管

道的时域信息向量 $[d_{t1}, d_{t2}, d_{t3}, \dots, d_{t\omega}]$ ,其中 $\omega$ 表示最终得到的时域管道亮度均值差的总数.最终我们能够得到时域特征矩阵.

## 5 特征量化及匹配

针对空域特征矩阵,具体量化过程如式(5)所示:

$$S_h(i, j) = \begin{cases} 0, & d_{ij} < 0 \\ 1, & d_{ij} \geq 0 \end{cases} \quad (5)$$

针对时域特征,具体的量化过程如式(6)所示:

$$T_h(i, j) = \begin{cases} 0, & d_{ij} < 0 \\ 1, & d_{ij} \geq 0 \end{cases} \quad (6)$$

通过上述量化方案,能够得到空域信息摘要哈希,时域信息摘要哈希.针对两段视频序列,我们通过摘要哈希的匹配对比达到对具有相同感知内容的视频与不同感知内容的视频进行快速区分,从而实现视频序列的唯一标识.

针对计算所得的原始视频 video 的空域摘要哈希 structHash(video),时域摘要哈希 timeHash(video).待检视频 videop 的空域摘要哈希 structHash(videop),时域摘要哈希 timeHash(videop).二者的空域摘要感知相似度可以描述为:

$$Sdis = \text{structDis} \left( \begin{matrix} \text{structHash}(\text{video}) \\ \text{structHash}(\text{videop}) \end{matrix} \right) \quad (7)$$

时域摘要感知相似度可以描述为:

$$Tdis = \text{timeDis} \left( \begin{matrix} \text{timeHash}(\text{video}) \\ \text{timeHash}(\text{videop}) \end{matrix} \right) \quad (8)$$

通过二者融合,可以将整体的相似度描述为:

$$Dis = \alpha Sdis + (1 - \alpha) Tdis \quad (9)$$

其中, $\alpha \in (0, 1)$ .

## 6 实验结果及分析

文章进行了性能比对实验以及篡改检测实验,实验所用视频序列采用了标准的数据库中的一些原始视频序列: CIF(352 × 288)格式(来源 <http://trace.eas.asu.edu/yuv/index.html>).

为便于对实验数据进行分析,实验结果用误码率(Bit Error Rate, BER)这一指标来描述.误码率是指两个哈希位串在相应位置上比特值的不匹配率<sup>[7]</sup>.

### 6.1 预处理

算法首先进行视频的预处理,对于给定的视频序列,将输入视频帧统一转换到灰度空间,将每一帧重采样为分辨率 256 × 256 的标准化表示,最终得到归一化的标准视频序列.

### 6.2 鲁棒性

鲁棒性是指算法要对各种内容保持操作的视频序

列鲁棒,因此,对于进行了内容保持操作的视频,算法应生成与原视频相近或者相同的哈希位串,即与原视频之间的感知哈希值误码率为 0 或很小.这里我们首先列出一种强度下的内容保持攻击操作项目,如表 1 所示:

表 1 内容保持操作参数设置

攻击操作	参数设置	攻击操作	参数设置
高斯噪声	高斯函数标准方差 0.1	Laplace 锐化	
椒盐噪声	0.2	亮度提升	20%
直方图均衡化	均衡因子 [1 -1]	亮度降低	20%
旋转	逆时针旋转角 +5°	对比度增加	[88 168] - > [0 255]
对比度降低	[0 255] - > [88 168]	中值滤波	[5 5]

foreman 标准视频序列与经过内容保持操作的视频序列进行测试,分别计算其误码率,实验结果如图 3 所示:

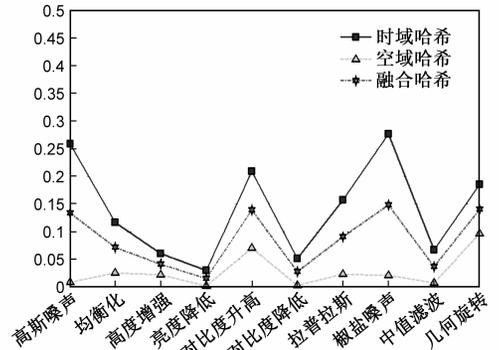


图3 foreman鲁棒性实验

由图 3 可知,空域哈希能够将误码率保持在 0.1 以下,而时域哈希的误码率最大接近 0.3.融合哈希的误码率在二者之间,整体上不会超过 0.15.由此可知,空域特征具有更好的鲁棒性能,融合特征次之,时域特征鲁棒性能最差.同时,空域特征的误码率在 0 ~ 0.1 之间,无法表现出各种操作对视频序列产生的变化,表现为过鲁棒性.因此,融合哈希相对具有更加合理,更好的鲁棒性能.

为了表现算法鲁棒性能的普遍性,实验针对 10 种内容保持操作,每一种类型进行不同视频序列计算,然后计算平均值,每一种攻击类型的误码率平均值结果如表 2 所示.

由表 2 实验结果可知,时域哈希误码率整体上表现为三种哈希中最大的值,空域哈希误码率表现为最小值,融合哈希趋于平衡状态,且融合哈希的误码率保持在 0.05 到 0.3 之间,这一区域更加符合内容保持操作的鲁棒性能.

### 6.3 可区分性

区分性是指对于内容不同的视频序列,其相似度

应该表现出巨大的差距,也就是误码率偏大.根据 Kevin Hamon<sup>[8]</sup>和 Baris Coskun<sup>[9]</sup>的概率分析,在确保区分性时,不同视频序列的误码率均值应该无限接近 50%.利用本文算法对 foreman 视频序列和其他 6 个不同视频进行误码率运算实验结果如图 4 示.

表 2 鲁棒性实验结果(平均值)

攻击类型	误码率 (空域哈希)	误码率 (时域哈希)	误码率 (融合哈希)
(1)高斯噪声	0.0378	0.2950	0.1664
(2)均衡化	0.0406	0.1545	0.0976
(3)亮度增强	0.0408	0.0945	0.0677
(4)亮度减弱	0.0062	0.0592	0.0327
(5)对比度高	0.1265	0.2276	0.1770
(6)对比度降	0.0086	0.0886	0.0486
(7)LapLace	0.0716	0.2269	0.1492
(8)椒盐噪声	0.0504	0.3511	0.2007
(9)中值滤波	0.0326	0.1919	0.1122
(10)几何变换	0.2527	0.3533	0.3030

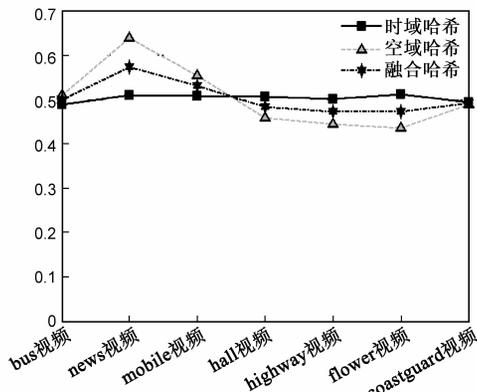


图4 foreman可区分性实验

由图 4 实验结果可知,时空域哈希、融合哈希整体在 0.4 到 0.7 之间.根据上述分析,时域哈希更加无限接近 50%,区分性更好.空域哈希虽然总体上趋于 50%,但是稳定性不够,这也在一定程度上反映了不同视频内容差异程度.从实验结果可知,对于 news 视频整体内容差异最大,融合哈希也在一定程度上反映了这一特性.综上所述,算法整体上具有良好的可区分性能.

为了表现算法区分性能的普遍性,实验针对 6 组视频序列,每一组 50 段视频,同时都与 foreman 视频序列完全不同,通过每一组计算的误码率均值,可以得到下表 3 所示实验结果.

由表 3 实验结果可知,算法整体上表现出优越的可区分性能,融合哈希的误码率保持在 0.45 到 0.55 之间,接近 50%的理想状态.

表 3 区分性实验结果(平均值)

攻击类型	误码率 (空域哈希)	误码率 (时域哈希)	误码率 (融合哈希)
(1)视频分组 1	0.5030	0.5004	0.5017
(2)视频分组 2	0.5266	0.4992	0.5129
(3)视频分组 3	0.4986	0.5008	0.4997
(4)视频分组 4	0.4253	0.4990	0.4662
(5)视频分组 5	0.5039	0.5008	0.5023
(6)视频分组 6	0.5995	0.5019	0.5507

## 6.4 篡改检测

针对当前视频传输过程中经常出现的视频片段替换等非法篡改操作,算法采用哈希序列匹配的策略可以进行有效的检测定位,这里我们通过构建相似度矩阵拟合图的方式给出检测结果.

依据原始视频序列与篡改视频序列的摘要哈希构建相似度矩阵,以左上角为原点,横轴表示原始哈希,竖轴表示篡改哈希,当两段视频完全相同时,矩阵对角线相似度应该为 1,这里我们利用灰度值表达  $[0, 1]$  区间,从而构建拟合图,检测结果如图 5 所示:

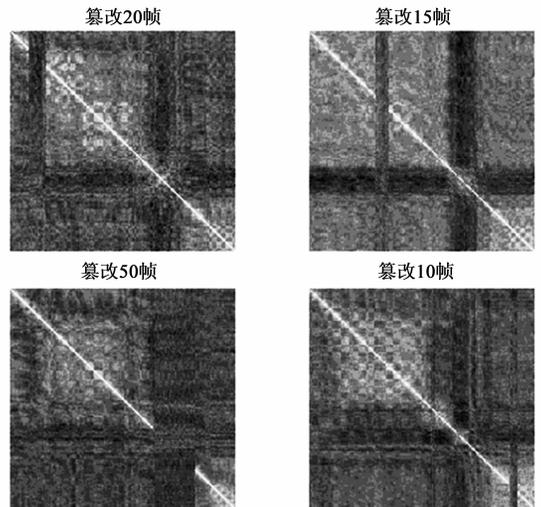


图5 篡改检测拟合图

实验结果如图 5 所示,利用灰度值表示相似度,相似度为 1 表示二者完全相同.由图 5 实验结果可以得出,当视频序列未发生任何篡改操作的时候,对角线灰度值全为 1,只有当某一段发生异常时,该段对应的哈希序列相似度出现下降,出现如图 5 所示的直线间断.由此可知,算法得到的哈希摘要信息能够准确的描述视频信息.并能够实现篡改操作的准确检测.

## 7 结论

本文针提出一种基于时空域亮度变化信息的感知哈希算法.通过不同的量化方案实现二者的融合,实验

结果证明算法具有更加平衡的鲁棒性与区分性能,可以实现篡改的准确检测.

### 参考文献

- [1] 牛夏牧,焦玉华.感知哈希综述[J].电子学报,2008,36(7):1405-1411.  
NIU Xiamu, JIAO Yuhua. An overview of perceptual Hashing [J]. Acta Electronica Sinica, 2008, 36(7): 1405 - 1411. (in Chinese)
- [2] Lina Wang, Xiaqiu Jiang, Donghui Hu. Robust perceptual image Hash using Gabor filters[A]. 2009 International Conference on Multimedia Information Networking and Security[C]. Hubei: MINES, 2009. 53 - 56.
- [3] Guopu Zhu, Jiwu Huang, Sam Kwong, Jianquan Yang. A study on the randomness measure of imageHashing[J]. IEEE Transactions on Information Forensics and Security, 2009, 4(4): 928 - 932.
- [4] Fouad Khelifi. Analysis of the security of perceptual image Hashing based on non-negative matrix factorization [J]. IEEE Signal Processing Letters, 2010, 17(1): 234 - 236.
- [5] Baris Coskun, Bulent Sankur. Spatio-temporal transform based video Hashing[J]. IEEE Transactions on Multimedia, 2006, 8(6): 446 - 458.
- [6] 张慧,张海滨,李琼.基于人类视觉系统的图像感知哈希算法[J].电子学报,2008,36(B12):30-34.  
ZHANG Hui, ZHANG Haibin. Image perceptual Hashing based on human visual system[J]. Acta Electronica Sinica, 2008, 36(B12): 30 - 34. (in Chinese)
- [7] Kim Changick. Spatio temporal sequence matching for efficient video copy detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(1): 127 - 132.
- [8] M K Mih Cak. New iterative geometric methods for robust perceptual image Hashing[A]. Proc of ACM Workshop Security and Privacy in Digital Rights Management[C]. Philadelphia: LNCS, 2001. 13 - 21.

- [9] Kevin Hamon. Histogram-based perceptual hashing for minimally changing video sequence[A]. The Second IEEE International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution [C]. Leeds: IEEE, 2006. 236 - 241.

### 作者简介



文振 ■ 男,1962年3月25日生,广东信宜人,1999年毕业于清华大学计算机科学与技术系,获工学硕士学位,现为深圳大学计算机与软件学院教授,长期从事宽带流媒体技术、视频信息安全技术、视频篡改检测与定位技术。  
E-mail: wenzk@szu.edu.cn



高金花 女,1972年出生江苏泰州,毕业于广东暨南大学,深圳信息职业技术学院讲师,主要研究方向为图像处理技术,视频版权保护技术。  
E-mail: gaojh@szit.com.cn



朱映映(通讯作者) 女,1976年生于湖北武汉,2004年毕业于武汉大学计算机应用专业,博士学位.现为深圳大学计算机与软件学院副教授.主要研究方向包括多媒体信息处理,信息安全等。  
E-mail: zhuyy@szu.edu.cn