

中图分类号: TN911.73 文献标志码: A 文章编号: 1006-8961(2011)10-1883-07

论文索引信息: 欧阳杰, 高金花, 文振焜, 张盟, 刘朋飞, 杜以华. 融合 HVS 计算模型的视频感知哈希算法研究[J]. 中国图象图形学报, 2011, 16(10): 1883-1889

融合 HVS 计算模型的视频感知哈希算法研究

欧阳杰¹⁾, 高金花²⁾, 文振焜¹⁾, 张盟²⁾, 刘朋飞¹⁾, 杜以华¹⁾

¹⁾(深圳大学计算机与软件学院, 深圳 518060) ²⁾(深圳信息职业技术学院, 深圳 518001)

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。目前关于感知哈希算法的研究主要集中在不断提高其鲁棒性和安全性上,忽略了人的主要视觉感知特性,导致了算法的过鲁棒性问题。将人类视觉系统可计算模型融入视频感知哈希算法框架中,用模拟人眼感受野特征提取特性的Cortex变换进行通道分解,并使用时-空域对比度敏感函数、眼球移动函数、亮度适应性调整函数、子带内和子带间对比度掩蔽函数综合计算最小视觉差提取感知特征。在保证较好鲁棒性的前提下,算法中使用扩散分块的机制提高安全性,通过与已有算法之间的比较,结果表明,本文提出的算法在鲁棒性和安全性方面取得了有效折衷,同时也体现了主观感知与客观评测上的一致性。

关键词: 人类视觉系统(HVS); 视频感知哈希; Cortex变换; 扩散分块; 最小视觉差

Video perceptual hashing fuse computational model of human visual system

Ouyang Jie¹⁾, Gao Jinhua²⁾, Wen Zhenkun¹⁾, Zhang Meng²⁾, Liu Pengfei¹⁾, Du Yihua¹⁾

¹⁾(College of Computer Science and Software, Shenzhen University, Shenzhen 518060 China)

²⁾(Shenzhen Institute of Information Technology, Shenzhen 518001 China)

Abstract: Perceptual hashing is a function of mapping from multimedia digital presentations to a perceptual hash value, which provides a secure and reliable technical support in fields such as identification, retrieval, and certification of multimedia content. The current algorithms fail in taking sufficient human visual perceptual factors into consideration. With the improvement of their over-robustness, most of the algorithms can't assure their securities. In this paper, a novel perceptual hashing algorithm is proposed. In order to simulate multi-channel features of the human visual system, a cortex transformation is combined with a computational model of the human visual system, which is designed by jointly considering four visual perceptual factors during the feature extraction stage, such as spatio-temporal contrast sensitivity function, eye movement, lightness adaptation, and intra-band and inter-band masking. Additionally, a diffusion mechanism is introduced into the preprocessing stage. The results suggest our proposed method could achieve better trade-offs between robust and secure resilient to various content-preserving manipulations, and also reflects the uniformity between subjective perception and objective evaluation.

Key words: human visual system (HVS); video perceptual hashing; Cortex transform; diffusion blocking; just noticeable difference

收稿日期: 2010-05-21; 修回日期: 2010-11-29

基金项目: 广东省自然科学基金自由申请项目(9151806001000022); 深圳市基础科研项目(JC200903120115A, 0015533011100512097); 深圳市公共技术服务平台项目(0015533054100524069); 深圳市科技研发资金重大产业技术公共计划项目(0015533021101018033)。

第一作者简介: 欧阳杰(1976—)男。深圳大学计算机与软件学院硕士研究生, 主要研究方向为数据挖掘、视频信息安全。

E-mail: jieouyang@yahoo.cn。

0 引言

感知哈希通过对多媒体感知内容和信息的提取,形成一个简短的数字摘要,通过摘要的匹配运算为多媒体信息的内容识别、认证和篡改检测提供技术支撑。与传统密码学哈希函数不同,感知哈希除了满足单向性和摘要性等性质外,还应具备感知鲁棒性、可区分性和安全性等性质^[1]。

1)感知鲁棒性 经过内容保持操作,感知内容相同的不同多媒体数字表示应映射为同一哈希值或者哈希值在一个指定的阈值内变化。换言之,对于两个多媒体数字内容,只要从感知上没有改变,其哈希序列应该相同或者相似。

2)可区分性 又称抗碰撞性,对于感知内容不同的多媒体数字信息,其映射的感知哈希值相差很大甚至完全不同。

3)安全性 安全性从 3 个方面度量,即单向性、区分性和随机性,算法中主要关注区分性。

目前图像感知哈希算法的研究成果较为显著,而特征选取较单一,主要集中在追求更好的鲁棒性。直到 2009 年,以张惠等人发表的文献[2]中,加入了人类视觉系统的一些特性,但也只考虑了图像中的亮度掩蔽特性。

然而,近年来在视频感知哈希领域鲜有比较开创性的研究成果发表,并且现有算法中都是基于帧与帧的信号处理方法。例如, Oostveen 等人在文献[3]中,对视频序列每帧的亮度分量随机分块,用 2×2 Haar 滤波器对视频序列进行时-空域滤波,首次在视频感知哈希算法中引入了时域上的同步因素; Zhou 等人在视频帧非重叠分块的基础上,根据帧间时域连续性定义不同相邻帧的不同块之间的相似性计算模型,并由此量化成哈希值^[4]; Venkatesan 在图像小波域中得到图像哈希值^[5]。2006 年, Coskun 首次对视频序列整体性做 3D-DCT (离散余弦变换),依次从 3 个维度作 DCT 后提取感知特征,并引入了随机置乱 DCT 基的方法保证算法的安全性,算法效率较高,其不足之处是 3D-DCT 缺乏严格的数学模型,且算法中包含的人类感知因素较少^[6]。

针对当前感知哈希模型对于人类感知因素的匮乏,我们对人类视觉系统进行了深入研究。研究表明,人类视觉系统试图通过对视觉信息从人眼到大

脑的各个信息处理阶段的模拟,来建模多媒体信息的视觉感知特性,而视频序列的感知距离度量相似性符合人类对感知的主观度量,使得人眼敏感的视觉内容在感知哈希编码中占有更重要地位,能够反映更多的人眼视觉特性。

综合上述分析,本文将人类视觉系统(HVS)计算模型引入视频感知哈希算法中,探讨如何在视频感知哈希中融入更多的感知因素,使得人的主观感知与客观评测(鲁棒性和区分性的评测)相一致,具有非常深远的意义。

1 基于 HVS 的视频感知哈希算法

本文从人类视觉系统计算模型出发,结合扩散分块机制,提出一个基于人类视觉系统和 Cortex 变换的视频感知哈希算法(以下简称 Cortex-JND 算法),该算法通过 Cortex 变换对 HVS 的某些底层特性进行建模,并引入人眼感知的主要视觉特性因素:时域-空域对比度敏感函数(S-T CSF)、眼球移动(eye movement)函数、亮度适应性调整函数和对比度掩蔽函数,由此模型计算能被人眼察觉的最小视觉差 JND 单位,将 JND 映射为哈希序列,用明斯基和的误差合并方法来分析哈希序列,并计算感知距离。

图 1 描述了基于 HVS 和 Cortex 变换的视频感知哈希算法模块图,包括 6 个主要模块。

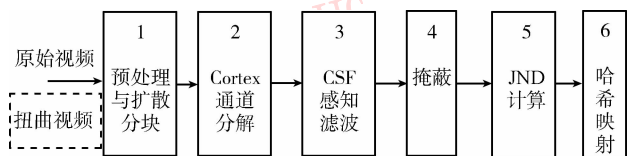


图 1 基于 HVS 与 Cortex 变换的算法模块图

Fig. 1 Algorithm diagram based on HVS and cortex transform

1)预处理扩散分块模块 视频序列归一化,帧图像扩散分块等;

2)通道分解模块 该模块主要通过人眼的多通道以及频率与方向选择特性,通过对输入视频帧进行 Cortex 变换,将其分解成 6 个径向频率和方向的子带;

3)CSF 感知滤波模块 根据 CSF 函数的带通和低通特性,利用线性滤波处理来模拟对比度敏感函数效应,达到模拟人眼对不同空间频率信号的感知效果;

4) 掩蔽模块 计算每个子带的对比度掩蔽阈值, 并计算各个子带的掩蔽阈值;

5) JND 计算 将各个影响因素综合, 计算 JND 值;

6) 哈希映射 根据 JND 和分解的主要系数, 映射成哈希值。

1.1 预处理与扩散分块

首先将原始视频序列在时域上平滑滤波后重采样, 得到长度为 T ($T = 64$) 的标准长度视频序列 $V_1(w, h, T) = V_1(w, h, 64)$, 再用 2-D 的 Gaussian 滤波器 ($h(m, n) = \exp((m^2 + n^2)/10) / \sqrt{10\pi}$) 对做空域平滑后重采样, 得到归一化的标准视频序列: $V_{\text{norm}}(256, 256, 64) = V_1(w, h, T) \times h(m, n)$ 。

用随机密钥 K 将归一化的标准视频序列分成 N 个随机可重叠分块, 分块大小为 8×8 的倍数, N 则为 256 的倍数。

由于置乱 (confusion) 与扩散 (diffusion) 被广泛应用到密码学的安全性评测上, 从理论上讲, 扩散与置乱的结合可以很好地满足感知哈希算法的鲁棒性、区分性和安全性的折衷, 即在置乱的基础上, 将视频帧的一个感知单元与哈希序列形成一对多的关系, 本文中, 为了使得算法较低程度的降低鲁棒性, 使用 1 个感知单元对应于哈希值的 2 个比特的扩散机制。

1.2 Cortex 通道分解

由 Watson 建立后经 Daly 改进的 Cortex 变换是以基本的 DFT 变换为基础, 将图像从不同的空间频率和方向角度分解成一系列的子带, 变换后每一个输出被认为是对人类视觉皮层 (cortex) 的一系列刺激响应, cortex 变换由两个滤波器组成: 径向频带滤波器 (Dom 滤波器) 和方向滤波器 (Fan 滤波器)^[7] 可表示为

$$\text{Cortex}_{k,l}(\alpha, \theta) = \begin{cases} \text{Dom}_k(\alpha) \cdot \text{Fan}_l(\theta) & k = 1, \dots, 4 \\ & l = 1, \dots, 6 \\ \text{base}(\alpha) = e^{-(\alpha/2\sigma^2)} & k = 5 \end{cases} \quad (1)$$

由于 DFT 与 DCT 的对称性对应关系, 利用 DFT 与 DCT 的转换公式可以将 Cortex 变换映射到 DCT 域中, 最后可以得到 Cortex 通道分解的直流分量和子带系数。

1.3 CSF 感知滤波

时-空域对比度敏感函数首次由 Kelly^[8] 建立,

后陆续经 Dally^[9] 改进, 加入了眼球移动 V 等补偿因素对空域对比度敏感函数进行了扩展和修正。本文结合 Cortex 通道分解后子带系数, 并加入视频序列中对象块时域运动因素, 构成时-空域对比度敏感函数 $t_{\text{st_csf}}(\alpha, V)$ 的计算

$$t_{\text{st_csf}}(\alpha, V) = c_0(k_1 + k_2 | \log(\varepsilon \cdot V/3 |^3) \cdot V \cdot \alpha^2 \cdot \exp(-c_1 \cdot \alpha \cdot (\varepsilon \cdot V + 2)/k_3) \quad (2)$$

式中, k_1, k_2, k_3 为常数, 分别取值 6.1, 7.3 和 23; ε 为常数, $\varepsilon = 1.7$; c_0, c_1 为常数, 用于控制 S-T CSF 的带宽和间隙; α 为空间频率 (周期/度); V 即对象在视网膜成像的相对速度 ($^\circ$ /s)。

每个分块为一个基本的感知单元, 感知单元的 S-T CSF 计算如下:

$$t_{\text{st_csf}}(n, i, j, t) = c_0(k_1 + k_2 | \log(\varepsilon \cdot V(n, t)/3 |^3) \cdot V(n, t) \cdot (2\pi\rho_{i,j})^2 \cdot \exp(-c_1 \cdot (2\pi\rho_{i,j}) \cdot (\varepsilon \cdot V(n, t) + 2)/k_3) \quad (3)$$

式中, i, j 为分块维数下标, $\rho_{i,j}$ 子带空间频率, 单位为周期/ $^\circ$; $V(n, t)$ 为视网膜相对移动速度。该结果为视觉扭曲阈值函数的倒数。

1.4 掩蔽函数

掩蔽包括子带间和子带内对比度掩蔽, 分别用子带间 $a_{\text{inter}}(n, t)$ 和子带内 $a_{\text{intra}}(n, i, j, t)$ 两个函数来计算, 分别用来模拟人眼对不同空间频率和纹理区域掩蔽的感知程度, 纹理区域的划分采用与文献 [10] 相同方法, 计算如下

$$a_{\text{inter}}(n, t) = \begin{cases} 1 + \frac{(E_{\text{texture}}(n, t) - \xi_1)}{(2\xi_2 - \xi_1)}\delta_1 & \text{纹理区域} \\ \delta_1 & \text{边缘区域} \\ \delta_2 & \text{边缘区域} \\ 1 & \text{平坦区域} \end{cases} \quad (4)$$

$L + M > 400$
 $L + M \geq 400$

$$a_{\text{intra}}(n, i, j, t) = \begin{cases} 1 & (i, j) \in LF \cup MF \text{ (平坦和边缘区域)} \\ \max\left\{1, \left[\frac{\text{Cortex}(n, i, j, t)}{T(n, i, j, t) \cdot a_{\text{Lum}}(n, t)}\right]^\varepsilon\right\} & \text{其他} \end{cases} \quad (5)$$

$\xi_1 = 290, \xi_2 = 900, \delta_1 = 1.125, \delta_2 = 1.25$ 为常数, ε 为常数, $\varepsilon = 0.36$ 。

1.5 JND 计算

Just-Noticeable Difference (JND) 是指人眼能感

知的最大可视化阈值, 低于该阈值的任何变化都不能被 HVS 所检测^[10], 基本视觉扭曲阈值函数为 CSF 函数的倒数, 将其扩展至灰度空间

$$T_{st_csf}(n, i, j, t) = \frac{1}{t_{st_csf}(n, i, j, t)} \cdot \frac{G}{\phi_i \phi_j (L_{max} - L_{min})} \cdot \frac{1}{r + (1 - r) \cdot \cos^2 \theta_{i,j}} \quad (6)$$

将式(4)—(6)相乘, 再计算亮度适应性调整函数 $a_{Lum}(n, t)$, 得到 JND 的计算公式

$$JND(n, i, j, t) = T_{st_csf}(n, i, j, t) \cdot a_{intra}(n, i, j, t) \cdot a_{inter}(n, t) \cdot a_{Lum}(n, t) \quad (7)$$

$a_{Lum}(n, t)$ 计算参考文献[11]中的 γ 校正方法。

1.6 哈希映射

提取每帧中每个分块 Cortex 分解投影后的直流分量 and 对应 JND 的均值, 得到两个分量序列

直流分量序列 $\overline{DC_1}, \overline{DC_2}, \dots, \overline{DC_N}$ (N 为分块数)

JND 序列 $\overline{JND_1}, \overline{JND_2}, \dots, \overline{JND_N}$ (N 为分块数)

本文中取 $N = 1\ 024$, 对一个标准视频序列, 共有 2 个长度为 $256 \cdot N$ 的实向量。由感知特征向量计算帧的分块感知误差

$$d[n, t] = \frac{(\overline{DC_n}(o) - \overline{DC_n}(d))}{\overline{DC_n}(o)} \quad (8)$$

$$jnd[n, t] = \frac{(\overline{JND_n}(o) - \overline{JND_n}(d))}{\overline{JND_n}(o)} \quad (9)$$

式中, $d[n, t]$ 表示第 t 帧的第 n 块, o 表示原始视频帧, d 表示受到攻击的测试视频帧, $d[n, t]$ 为

分块感知误差, $jnd[n, t]$ 为分块 JND 感知误差。用闵可夫斯基 (Minkowski) 求和公式合并 $d[n, t]$ 和 $jnd[n, t]$ 计算整个视频序列的感知距离

$$D(V(o), V(d)) = \left| \sum_t |d(t) - jnd(t)|^\beta \right|^{1/\beta} \quad (10)$$

式中, $V(o), V(d)$ 分别为原始视频序列和扭曲的测试视频序列, 实验中 $\beta = 4$ 。

为了方便与其他文献中算法比较, 算法用中值量化方法将视频帧映射为海明距离, 与式(10)感知距离计算的区别是海明距离的计算是先在帧内用闵可夫斯基求和公式求和, 再用中值量化为哈希值。

2 实验结果与分析

实验中, 为了比较扩散分块与无扩散分块算法的性能, 下文中用 Cortex-JND (扩散分块) 和 Cortex-JND (无扩散) 分别加以标记。实验中采用了一些标准数据库的原始视频: QCIF 和 CIF 格式 (主要来源 <http://trace.eas.asu.edu/yuv/index.html>), 共有 55 个视频序列 (QCIF 格式 22 个, CIF 格式 30 个)。实验中, 引入结构相似度 (SSIM) 和峰值信噪比 (PSNR) 两个客观评价指标来衡量对不同攻击强度的客观感知评价。实验中包含 14 种攻击类型的内容保存操作, 每种攻击类型随攻击强度而变化, 如表 1 所示。

表 1 一种攻击强度下的实验项目描述

Tab. 1 The description of modifications and their parameters

攻击操作	参数描述	客观指标	攻击操作	参数描述	客观指标
1) 高斯模糊	高斯滤波核方差 $\sigma^2 = 55$	SSIM = 0.70 PSNR = 19.88	8) 剪切	每帧剪切 剪切因子: 0.2	无
2) 高斯噪声	高斯噪声函数标准方差 $\sigma = 110$	SSIM = 0.04 PSNR = 9.01	9) 旋转	逆时针旋转角 +3°	无
3) 中值滤波	滤波平滑窗口 (5, 5)	SSIM = 0.89 PSNR = 25.68	10) 直方图均衡化	均衡因子 [1 - 1]	SSIM = 0.58 PSNR = 18.41
4) 亮度增强	像素值 = $r \times Lum_{frame}$ ($r = 80\%$)	SSIM = 0.68 PSNR = 8.83	11) 帧内循环移动	每帧行列循环移动: 移动行和列的 3%	SSIM = 0.78 PSNR = 25.56
5) 亮度减弱	像素值 = $r \times Lum_{frame}$ ($r = -80\%$)	SSIM = 0.379 PSNR = 6.53	12) 随机帧替换	随机替换 25 帧/s	SSIM = 0.84 PSNR = 30.22
6) 对比度增加	亮度值 [88, 168] 映射为 [0, 255]	SSIM = 0.46 PSNR = 13.08	13) 帧内容篡改	每帧内容替换窗口: [64 64]	SSIM = 0.68 PSNR = 29.11
7) 对比度降低	亮度值 [0, 255] 线性映射为 [88, 168]	SSIM = 0.73 PSNR = 17.67	14) MPEG4 压缩	divX MPEG4 压缩 压缩率为 5bit/s	SSIM = 0.74 PSNR = 27.83

本文算法都在 Windows XP + MATLAB 2009a 平台上用 MATLAB 语言实现,并将算法与现有的一些引用率最高的几篇文献中算法比较。

2.1 鲁棒性实验及分析

当用感知距离、海明距离来度量时,其基本的感知阈值期望值都是 0.5(或 $T/2$),直观上容易理解,两个感知相同或感知不同的视频序列的感知距离或归一化的海明距离期望值分别应该小于 0.5 和在 0.5 附近。从数学角度,无论在哈希算法中是否包含量化压缩阶段,由概率论分析,其实验数据集中的视频序列的哈希值总体上一定满足统一分布,如典型的二项分布和高斯分布。由此可以计算出感知相同和感知不同的基本阈值是 0.5,此点在文献 [5,11] 中做了详细分析。

1)用 Foreman 序列与受到 14 种不同攻击的视频序列做鲁棒性对比,实验结果见表 2。

表 2 Foreman 序列感知距离结果

Tab. 2 Perception distance of Foreman sequence

攻击类型	参数	Cortex-JND	Cortex-JND
		(无扩散) 感知距离	(扩散分块) 感知距离
1) 高斯模糊	高斯滤波 $\sigma^2 = 30$	0.021 0	0.222 6
2) 高斯噪声	$\sigma^2 = 110$	0.088 1	0.226 8
3) 中值滤波	[5 5]	0.006 1	0.038 1
4) 亮度增强	80%	0.401 0	0.358 6
5) 亮度减弱	80%	0.552 4	0.552 5
6) 对比度增加	[88 168] → [0 255]	0.491 3	0.547 1
7) 对比度降低	[0 255] → [88 168]	0.332 6	0.333 4
8) 剪切	20%	0.461 7	0.461 7
9) 帧旋转	7°	0.503 8	0.503 4
10) 直方图均衡化	均衡化因子 [1 -1]	0.501 7	0.635 2
11) 帧内循环移动	7%	0.203 8	0.303 8
12) 随机帧替换	1 s(25 帧)	0.256 7	0.456 1
13) 帧内容篡改	窗口 [64 64]	0.023 8	0.204 0
14) MPEG4 压缩	5 kbit/s	0.099 4	0.098 5

从表 2 对单个视频的鲁棒性实验中发现,当攻击强度达到一定的程度时,一些项目的感知距离已接近或者超过 0.5,此时表明算法针对此种攻击强度的攻击操作的鲁棒性降低。而

对于高斯模糊、叠加高斯噪声等项目,感知距离保持在 0.1 以下,鲁棒性较好;其他项目的感知距离保持在 0.1 ~ 0.4 之间,针对 Foreman 序列的各种不同攻击,算法具有符合人类视觉的鲁棒性性能。

2)为了更客观检测本文算法表现的鲁棒性,分别对不同的内容保持操作进行了相关实验,这里仅以高斯噪声攻击为例,利用 Foreman 序列在不同攻击强度下的视频质量,相应的感知距离与攻击强度的关系如图 2 所示,为比较方便,包含了算法 [2] 性能曲线。

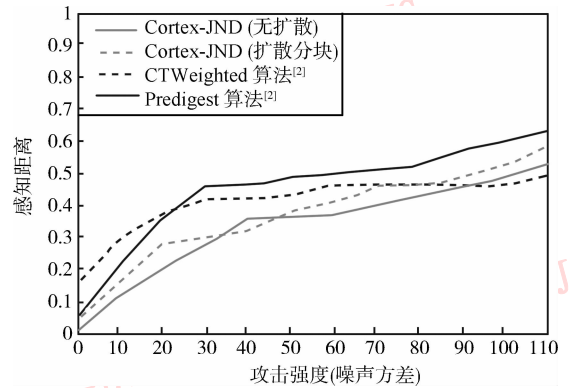


图 2 不同算法在高斯噪声攻击下的感知距离比较
Fig. 2 Comparison of different algorithms in AWGN

图 2 比较直观地比较了各种算法在高斯噪声攻击下的整体性能,4 种算法都表现在 $\sigma^2 = 80$ 附近,感知距离接近于 0.5,比较符合人眼的视觉特性。在噪声强度 $\sigma^2 = 85$ 时,视频帧的原始信号几乎被噪声淹没,不可辨认。同时,从图 2 可以看到,本文算法的感知距离曲线整体都在其他两种算法以下,反映出本文算法较其他两种算法具有更好的鲁棒性性能。

3)为了探测算法的整体性能,用 55 个原始视频序列,在一种攻击强度下,对 14 种攻击类型进行鲁棒性实验,用感知距离度量相似性,对每一种攻击类型,计算所有视频序列的平均值,得到的实验结果如表 3 所示。

分析表 3 结果,针对同一攻击强度的不同攻击,算法整体表现出良好的鲁棒性。同时也说明了基于人类视觉特性提取感知特征的有效性,在表 3 显示的攻击强度下,14 种攻击操作的平均感知距离最大值在 0.5 附近,超过 0.5 的项目只有少数几个,而大部分项目的感知距离小于 0.4。

表 3 Cortex-JND 算法的组内测试感知距离平均值

Tab.3 Mean intra-Hash of Cortex-JND(perceptual distance)

攻击类型	参数	Cortex-JND	Cortex-JND
		(无扩散) 感知距离	(扩散分块) 感知距离
1) 高斯模糊	高斯滤波 $\sigma^2 = 30$	0.139 4	0.146 7
2) 高斯噪声	$\sigma^2 = 110$	0.371 4	0.386 8
3) 中值滤波	[5 5]	0.011 2	0.028 2
4) 亮度增强	80%	0.487 8	0.425 4
5) 亮度减弱	80%	0.534 7	0.559 3
6) 对比度增加	[88 168]→[0 255]	0.356 9	0.401 6
7) 对比度降低	[0 255]→[88 168]	0.385 0	0.371 4
8) 剪切	20%	0.374 9	0.480 4
9) 帧旋转	7°	0.402 1	0.429 2
10) 直方图均衡化	均衡化因子[1-1]	0.502 6	0.510 7
11) 帧内循环移动	7%	0.215 0	0.267 4
12) 随机帧替换	1s(25 帧)	0.249 5	0.277 0
13) 帧内容篡改	窗口[64 64]	0.027 5	0.075 8
14) MPEG4 压缩	5kbit/s	0.128 5	0.177 3

表 4 Cortex-JND 算法的组内组间测试海明距离平均值

Tab.4 Mean intra-Hash and inter-Hash of Cortex-JND

(Hamming distance)

攻击类型 (参数同表 3)	Cortex-JND 组内测试 海明距离均值		Cortex-JND 组间测试 海明距离均值	
	无扩散	扩散分块	无扩散	扩散分块
	1) 高斯模糊	4.46	7.85	33.20
2) 高斯噪声	7.82	9.37	31.31	32.00
3) 中值滤波	3.11	8.78	31.01	32.48
4) 亮度增强	12.40	15.34	30.66	32.45
5) 亮度减弱	17.10	21.61	35.28	32.83
6) 对比度增加	11.42	14.57	31.86	33.97
7) 对比度降低	12.32	16.39	32.05	32.24
8) 剪切	3.35	3.87	34.22	32.66
9) 帧旋转	17.13	17.67	33.79	33.45
10) 直方图均衡化	6.88	9.65	34.80	33.24
11) 帧内循环移动	6.85	8.67	32.40	33.72
12) 随机帧替换	4.12	5.79	32.04	28.06
13) 帧内容篡改	7.16	9.21	32.50	30.52
14) MPEG4 压缩	7.982	9.93	32.05	33.38

2.2 区分性实验及分析

区分性主要对感知哈希值的抗碰撞性分析,为了方便与其他算法比较,用海明距离来度量感知哈希值的相似性。针对 55 个视频序列的统计分两种情况:组内测试和组间测试。组内测试表示同一个视频与不同攻击操作类型在同种攻击强度下的海明距离,理论上该距离越小越好;组间测试实验中用数据集中的 30 个 CIF 视频序列两两匹配,得到 435 个匹配结果。根据前文分析,理论上组间测试的海明距离应该小于或等于 $T/2$ 。针对一种攻击强度(攻击强度同表 3)的实验结果见表 4。

分析表 4 的实验结果,整体上各种攻击操作的组内和组间海明距离虽然比文献[6]中的算法略高一些(文献[6]中组内和组间均值分别小于 11.29 和 32.05),但还是保证了算法的鲁棒性和区分性。其主要原因是 Cortex-JND 算法中融入了更多的人类视觉感知因素,导致鲁棒性受影响。

统计组间测试的匹配实验结果,可以近似地拟合为期望值 $\mu = 32.740 3$,方差 $\sigma^2 = 3.906 9$ 的高斯分布 $N(\mu, \sigma^2)$ 。如图 3 所示。

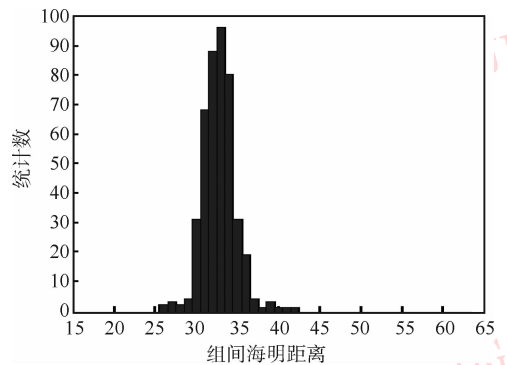


图 3 组间测试海明距离直方图分布

Fig.3 Histogram of inter-Hash(Hamming distance)

由实验数据,选择 $T_1 = 26.00$,可以计算哈希值的碰撞率为

$$P_c = 1 - \int_{T_1}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 - \frac{1}{2} \operatorname{erfc}\left(\frac{T_1 - \mu}{\sqrt{2}\sigma}\right) = 3.248 2 \times 10^{-4}$$

从上式计算结果可以看出,算法生成的哈希值,

冲突概率很小,为 10^{-4} 级,以较高的概率保证视频哈希值的唯一性,即保证了感知不同内容视频序列之间哈希值的较好区分性。

3 结 论

实验表明,提出的算法在 14 种攻击操作下总体上满足鲁棒性和区分性的要求,能够抵抗大多数攻击造成的合理失真,而对于人眼比较敏感的攻击操作,如亮度和对比度变化、几何操作、直方图均衡化等,由于算法融入了更多的人类视觉感知特性,在提高人的主观感知时降低了算法的鲁棒性,并保证了算法的区分性。

总体来说,算法在鲁棒性和区分性上取得了有效折衷,可通过对算法的进一步分析改进,为视频信号的篡改检测、版权保护、内容认证提供技术支持。

参考文献 (References)

- [1] Niu Xiamu, Jiao Yuhua. An overview of perceptual hashing[J]. Chinese of Journal Electronics, 2008, 36(7):1405-1411. [牛夏牧, 焦玉华. 感知哈希综述[J]. 电子学报, 2008, 36(7): 1405-1411.]
- [2] Zhang Hui, Zhang Haibin, Li Qiong. Image perceptual Hashing based on human visual system [J]. Chinese of Journal Electronics, 2008, 12(12A):30-34. [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, 12(12A):30-34.]
- [3] Oostveen J, Kalker T, Haitsma J. Visual Hashing of digital video: applications and techniques [J]. SPIE Applications of Digital Image Processing XXIV, SPIE Press, 2001, 4472: 121-131.
- [4] Zhou X, Schmucker M, Brown C L. Perceptual Hashing of video content based on differential block similarity [C]// Proceedings of 2005 International Conference on Computational Intelligence and Security. Xi'an: Xidian University Press, 2005: 3802.
- [5] Venkatesan R, Koon S M, Jakubowski M H. Robust image Hashing [C]// Proceedings of IEEE International Conference on Image Processing (ICIP). Vancouver: IEEE Press, 2000, (3): 664-666.
- [6] Coskun B, Sankur B, Memon N. Spatio-temporal transform based video Hashing [J]. IEEE Transactions on Multimedia, 2006, 8(6):1190-1208.
- [7] Watson A B. The Cortex transform: rapid computation of simulated neural images [J]. Computer Vision Graphics and Image Processing, 1986, 39(1): 311-327.
- [8] Kelly D H. Motion and vision: II. stabilized spatio-temporal threshold surface [J]. Journal of the Optical Society of America A, 1979, 69(3):1340-1349.
- [9] Daly S. Engineering observations from spatiotemporal and spatiotemporal visual models [C]// Processing SPIE. San Jose, CA, USA: SPIE Press, 1998:180-191.
- [10] Wei Zhenyu, King N N. Spatio-temporal just noticeable distortions profile for grey scale image/video in DCT domain [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2009, 19(3):337-346.
- [11] Zhu Guopu, Huang Juwu, Kwong Sam, et al. Fragility analysis of adaptive quantization-based image Hashing [J]. IEEE Transactions on Information Forensics and Security, 2010, 5(1):133-144.