



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院遥感与数字地球研究所  
中国图象图形学学会  
北京应用物理与计算数学研究所

# 中国图象图形学报

2013  
08  
VOL.18

ISSN1006-8961  
CN11-3758/TB



北魏



隋



初唐

数字文化遗产研究 P968

## 综述

尺度不变特征变换算子综述

刘立, 詹茵茵, 罗扬, 刘朝晖, 彭复员 ..... 885

## 图像处理和编码

改进的H.264帧间模式选择算法

任克强, 张旭光, 罗会兰 ..... 893

均质雾天下摄像机动态标定的车速估计

宋洪军, 陈阳舟, 郜园园 ..... 901

基于人脸五官特征的空域差错掩盖算法

张江鑫, 谢晋, 邝万坤 ..... 913

采用圆谐-傅里叶矩的图像区域复制粘贴篡改检测

秦娟, 李峰, 向凌云, 殷蓓蓓 ..... 919

## 图像分析和识别

基于视频感知哈希的视频篡改检测与多粒度定位

朱映映, 文振焜, 杜以华, 邓良太 ..... 924

视觉注意模型的道路监控视频关键帧提取

刘云鹏, 张三元, 王仁芳, 张引 ..... 933

格拉斯曼流形上的半监督判别分析

姜伟, 陆瑶, 杨炳儒 ..... 944

精神疲劳实时监测中多面部特征时序分类模型

陈云华, 张灵, 丁伍洋, 严明玉 ..... 953

图像分类中的概率乘积核函数

杨赛, 赵春霞 ..... 961

轮廓整体结构约束的壁画图像相似性度量

唐大伟, 鲁东明, 杨冰, 许端清 ..... 968

应用色彩空间聚类方法实现道路建模

向宸薇, 王拓, 于舰 ..... 976

## 图像理解和计算机视觉

杂乱背景和摄像机移动下的时空兴趣点检测

刘长红, 陈勇, 王明文 ..... 982

形状和颜色混合不变量在图像检索中的应用

公明, 曹伟国, 李华 ..... 990

## 计算机图形学

结合统计分类和边缘特征的最优视图提取

潘翔, 陈敖, 章国栋 ..... 1004

## 医学图像处理

局部熵驱动下的脑MR图像分割与偏移场恢复耦合模型

张建伟, 杨红, 陈允杰, 方林, 詹天明 ..... 1011

改进的模糊聚类亚实质肺结节3维分割

李翠芳, 聂生东, 王远军, 孙希文, 郑斌 ..... 1019

## 遥感图像处理

高光谱海岸带区域分割的活动轮廓模型

王相海, 金弋博 ..... 1031

改进的基于非局部均值的极化SAR相干斑抑制

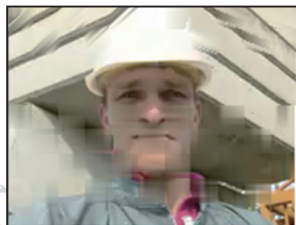
赵忠民, 赵拥军, 牛朝阳 ..... 1038

小波变换和稀疏表示相结合的遥感图像融合

刘婷, 程建 ..... 1045

《中国图象图形学报》论文摘要写作要求

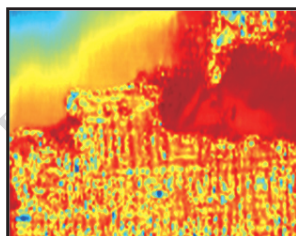
1054



基于人脸五官特征的空域差错掩盖算法 (第913页)



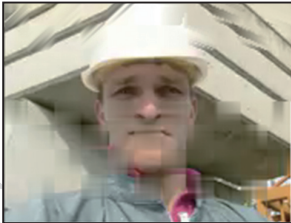
应用色彩空间聚类方法实现道路建模 (第976页)



改进的基于非局部均值的极化SAR相干斑抑制 (第1038页)

# CONTENTS

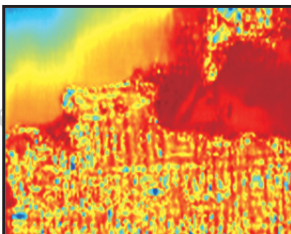
## JOURNAL OF IMAGE AND GRAPHICS



Spatial error concealment algorithm based on the facial features ( P913 )



Road modeling using color spatial clustering ( P976 )



Improved polarimetric SAR speckle filter based on non-local means ( P1038 )

### Review

#### Summarization of the scale invariant feature transform

Liu Li , Zhan Yinyin, Luo Yang, Liu Chaohui, Peng Fuyuan ..... 885

### Image Processing and Coding

#### Improved inter frame mode selection algorithm for H.264

Ren Keqiang, Zhang Xuguang, Luo Huilan ..... 893

#### Vehicle velocity estimation based on camera calibration under homogenous fog

Song Hongjun, Chen Yangzhou, Gao Yuanyuan ..... 901

#### Spatial error concealment algorithm based on the facial features

Zhang Jiangxin, Xie Jin, Kuang Wankun ..... 913

#### Detection of image region copy-move forgery using radial harmonic Fourier moments

Qin Juan, Li Feng, Xiang Lingyun, Yin Changming ..... 919

### Image Analysis and Recognition

#### Video forgery detection and multi-granularity location based on video perceptual hashing

Zhu Yingying, Wen Zhenkun, Du Yihua, Deng Liangtai ..... 924

#### Key frame extraction based on the visual attention model for lane surveillance video

Liu Yunpeng, Zhang Sanyuan, Wang Renfang, Zhang Yin ..... 933

#### Semi-supervised discriminant analysis on Grassmannian manifold

Jiang Wei, Lu Yao, Yang Bingru ..... 944

#### Time-series classification model based on multiple facial feature for real-time mental fatigue monitoring

Chen Yunhua, Zhang Ling, Ding Wuyang, Yan Mingyu ..... 953

#### Probability product kernels in image classification

Yang Sai, Zhao Chunxia ..... 961

#### Similarity metrics between mural images with constraints of the overall structure of contours

Tang Dawei, Lu Dongming, Yang Bing, Xu Duanqing ..... 968

#### Road modeling using color spatial clustering

Xiang Chenwei, Wang Tuo, Yu Jian ..... 976

### Image Understanding and Computer Vision

#### Spatio-temporal interest point detection in cluttered backgrounds with camera movements

Liu Changhong, Chen Yong, Wang Mingwen ..... 982

#### Application of combined shape and color invariants to image retrieval

Gong Ming, Cao Weiguo, Li Hua ..... 990

### Computer Graphics

#### Best view selection based on statistical classification and edge features

Pan Xiang, Chen Ao, Zhang Guodong ..... 1004

### Medical Image Processing

#### Brain MR image segmentation and bias correction model based on local entropy

Zhang Jianwei, Yang Hong, Chen Yunjie, Fang Ling, Zhan Tianming ..... 1011

#### Segmentation of sub-solid pulmonary nodules based on improved fuzzy

#### C-means clustering

Li Cuifang, Nie Shengdong, Wang Yuanjun, Sun Xiwen, Zheng Bin ..... 1019

### Remote Sensing Image Processing

#### The active contour model for segmentation of coastal hyperspectral remote sensing image

Wang Xianghai, Yin Yibo ..... 1031

#### Improved polarimetric SAR speckle filter based on non-local means

Zhao Zhongmin, Zhao Yongjun, Niu Chaoyang ..... 1038

#### Remote sensing image fusion with wavelet transform and sparse representation

Liu Ting, Cheng Jian ..... 1045

中图法分类号: TN911.73 文献标识码: A 文章编号: 1006-8961(2013)08-0924-09

论文引用格式: 朱映映, 文振焜, 杜以华, 邓良太. 基于视频感知哈希的视频篡改检测与多粒度定位[J]. 中国图象图形学报, 2013, 18(8): 924-932. [DOI:10.11843/jig.20130806]

## 基于视频感知哈希的视频篡改检测与多粒度定位

朱映映, 文振焜, 杜以华, 邓良太

深圳大学计算机与软件学院, 深圳 518060

**摘要:** 为了对被篡改过的视频进行准确快速的篡改检测与定位, 引入人类视觉可计算模型, 提出一种多层次、多粒度的视频篡改快速检测与定位算法。采用随机分块采样技术, 提取视频结构感知特征及视频图像时域感知特征, 利用哈希理论的单向摘要特性量化感知特征, 获取视频摘要哈希。通过应用相似度矩阵进行多粒度、多层次篡改部位检测与定位。实验结果表明, 相似度拟合图能够体现视频篡改攻击强度和攻击部位, 算法表现出更好的篡改检测准确率与定位精确度。

**关键词:** 视频篡改; 视频感知哈希; 多粒度定位; 感知特征

### Video forgery detection and multi-granularity location based on video perceptual hashing

Zhu Yingying, Wen Zhenkun, Du Yihua, Deng Liangtai

College of Computer Science and Software, Shenzhen University, Shenzhen 518060, China

**Abstract:** To fast and accurately detect videos that were tampered, the human visual model is introduced to our algorithm. A multi-level and multi-granularity algorithm to detect and locate video tampering is presented in this paper. The random block sampling technology is used, and video structure perceptual features and time-domain perceptual features of video images are extracted. Then, the unidirectional abstract of the hash theory is used to quantify perceptual features, and the video abstract hash value is obtained. The similarity matrix is applied to give a multi-level and multi-granularity detection and location for tampered data. Experimental results show that similarity fitting diagram can reflect the attack power and the attack site of video tampering. The proposed algorithm shows better precision and positioning accuracy.

**Key words:** video tampering; video perceptual hashing; multi-granularity location; perceptual features

## 0 引言

由于网络技术、多媒体技术的迅猛发展, 视频传播日益迅速, 视频数据呈爆炸性增长。同时伴随如

Adobe Premiere 等视频编辑软件的广泛使用, 用户对数字视频的修改和编辑变得更加容易。这使得视频信息的安全问题日益突出。对于视频数据, 目前需要亟待解决的问题是如何认证数字视频的真实性和完整性, 保护视频数据的版权, 以及快速准确地检

收稿日期: 2012-08-10; 修回日期: 2012-12-25

基金项目: 国家自然科学基金项目(61170326); 广东省自然科学基金项目(9151806001000022); 深圳市基础研究项目(JC200903120115A, 0015533011100512097); 深圳市公共技术服务项目(0015533054100524069)

第一作者简介: 朱映映(1976—), 女, 副教授, 2004年于武汉大学获计算机应用专业博士学位, 主要研究方向为多媒体信息处理, 多媒体信息安全。E-mail: coco211@126.com

测并定位可能发生的视频篡改。

视频篡改主要是指借助一些视频编辑工具或者软件,对视频的内容、格式等进行修改,导致其原始信息完整性、真实性被破坏。广义上,任何对视频的编辑都属于视频篡改;狭义上,视频篡改主要是指那些通过修改视频达到一定违规甚至非法目的的行为<sup>[1]</sup>。近几年出现的视频篡改检测方法,主要有数字水印、数字签名和数字视频被动认证技术。对于数字水印技术,现有的大部分视频数据都没有事先将水印嵌入。即使在某些视频中嵌入了水印,也或多或少会影响视频的原始信息。而且,对于剪切、旋转等常规的视频图像编辑操作,嵌入水印信息的鲁棒性一般较难满足,水印算法很难在提取和嵌入之间进行同步。而感知哈希技术因为继承了传统密码学哈希的单向性、抗碰撞性和摘要性等特点,可以采用与密码学哈希技术相似的方式来认证视频内容,对视频篡改进行检测和定位。相对于数字水印技术,一方面,感知哈希不用修改原始的视频数据,所需的资源更少,其应用模式灵活方便;另一方面,感知哈希用较为简短的哈希值来表达视频图像内容特征,因此,还能够作为数字水印技术的一个扩充,将哈希值作为水印信息嵌入到视频数据中。由于数字视频数据自身结构较为复杂,数字视频技术难度较大,视频被动认证技术的研究起步较晚,仍然有很多的困难没有克服。

## 1 研究现状

近些年,国内外数字多媒体取证技术主要围绕数字图像,对数字视频的篡改检测与定位研究才刚刚起步。Lin<sup>[2]</sup>提出了一种基于视频水印的篡改检测方法;Altun<sup>[3]</sup>设计了一种能容忍轻微压缩的分等级的半脆弱性水印的框架,并实现了对篡改的定位;文献[4-5]分别提出了非常有效的视频水印和视频Hash的篡改检测方法,一般都能较好地检测出视频的篡改,但是数字签名会增加视频的传输数据量,而且这两种方法都需要在传输端设定相应的仪器用来嵌入视频水印或数字签名,增大了视频的传输延时。以美国Dartmouth大学的Farid研究团队为首,率先开展数字视频被动取证的研究,并取得一些阶段性的成果<sup>[5-7]</sup>。Wang与Farid<sup>[6]</sup>对MPEG格式视频序列进行分析,基于视频是一个连续的内容,相邻帧间具有非常大的相关性特性,利用篡改前后预测残差

的周期性发生这一特性,对视频删除帧或者插入帧的篡改操作进行检测,但是当删除或者插入GOP(group of pictures)的整数倍帧数时,则无法检测。进而,Wang等人在视频篡改检测中将图像双压缩篡改检测方法进行延伸,利用视频双重MPEG压缩过程中双量化存在的特征来判断篡改情况<sup>[7]</sup>,它的局限在于无法检测第二次压缩质量低于第一次压缩质量的视频。Zhang等人<sup>[8]</sup>根据图像模糊区域的非一致性来探测一种移除Logo的视频篡改,对视频帧进行分块处理,同时利用小波分析计算各分块的模糊值,通过建立高斯混合模型(GMM)分析各分块模糊值分布情况来检测视频篡改。但该方法对于视频压缩本身产生的低质量模糊视频效果不佳。Kobayashi等人<sup>[9]</sup>提出了一种通过噪声特征从指定场景里检测可疑区域的算法,通过计算得到平均时间内每个像素的噪声情况,对每个像素以高度准确率拟合离群点的估计值。秦运龙等人<sup>[10]</sup>根据监控视频篡改时留下的痕迹特征以及帧间内容的连续性,通过统计视频帧间预测的运动矢量来检测恶意视频篡改。此方法通过计算滑动窗口内平均运动矢量幅度均值的距离与各P帧运动矢量的幅度的差值,检测篡改并且定位。Lee等人<sup>[11]</sup>通过验证设备痕迹和噪声的一致性,给出了在静止场景下,根据噪声特征进行视频可疑区域检测的方法。基本思想是视频伪造区域像素的噪声级函数(NLF)与其他区域的不一致,可以通过最大后验估计区分。此方法引入一种概率模型,给出了NLF对控制每个像素噪声特征的影响。其优点在于对于静止场景视频的取证精度高,可以实现像素级的认证。缺点是没有考虑视频压缩本身是一种噪声源的影响。而且,该方法对亮度和对比度调节等可能严重影响噪声特征的篡改伪造效果。

近年来,随着视频感知哈希的兴起,以及视频感知哈希值具有独立于视频数据、无需嵌入码流便能实现视频安全性认证的巨大优势,国内外研究人员进行了基于视频感知哈希算法的视频篡改检测的尝试。Oostveen等人<sup>[12]</sup>对视频序列每帧的亮度分量随机分块,用 $2 \times 2$  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<sup>[13]</sup>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,该文献提出了基于视频感知哈希的篡改检测方法。然而,视频感知哈希是一种不可逆的数字摘要

技术,无法实现从感知哈希值到视频画面的重构。Coskun 等人<sup>[14]</sup>首次在整体视频序列做 3 维离散余弦变换,依次从 3 个维度作离散余弦变换(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。因此,当前基于视频感知哈希的篡改定位面临着诸多挑战和重大发展瓶颈。

感知哈希是通过研究人类视觉系统的信息处理和感知过程,基于人类视觉系统计算模型与哈希理论相融合,构建视频感知摘要集,从而实现视频的唯一标识。因此,利用视觉感知哈希理论分析及处理海量媒体并获取感知特征,利用简短摘要哈希的对比匹配,可以实现视频篡改的准确检测与定位。

结合视频感知哈希技术,引入人类视觉可计算模型,提出一种多层次、多粒度的视频篡改快速检测与定位算法。在降低复杂程度的基础上,以自顶向下的感知过程进行视频特征提取。针对视频篡改检测与多粒度定位等关键技术提出解决方案,实现视频篡改的准确检测定位。

## 2 结构哈希与时域哈希构造

为了确保算法的普遍性与安全性,首先对输入视频进行灰度化预处理,在此基础上,采用随机可重叠分块的方式对视频每一帧画面进行分块操作。将视频每一帧看成由其内部局部区域构成的集合  $I = \bigcup_{i=1}^M B_i$ ,其中  $M$  表示局部区域的个数。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。由于人的视觉选择的随机性,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(如图 1 所示)。从而视



图 1 视觉选择随机性示意图  
Fig. 1 Visual random choice

频观察者在视频帧所有可能被关注的局部区域上的一个随机采样的过程可以描述为随机可重叠的分块,即

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

式中,  $N \ll M$ ,  $\psi \subset I$  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。在随机种子的控制下产生  $m$  个随机的矩形块  $B_1, B_2, \dots, B_m$ 。针对第  $i$  个分块,计算亮度平均值

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

式中,  $K$  表示该分块中像素个数,  $\text{gray}(j)$  表示第  $j$  个像素的亮度值。

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  $S_\theta \geq 1, 1 \leq \theta \leq p$ ,从步长集合中随机选择多个步长作为计算亮度均值差的分块步长,针对步长  $S$  的计算公式为

$$d(i, i+S) = V_{i+S} - V_i \quad (3)$$

通过亮度均值差的计算,可以得到每一帧的结构特征向量  $[d_1 d_2 d_3 d_4 \dots d_q]$ ,其中  $q$  表示最终计算亮度差值的总数,针对视频长度为  $l$  的视频序列,其特征信息可以描述为

$$\begin{bmatrix} d_{11} & d_{12} & d_{13} & \dots & d_{1q} \\ d_{21} & d_{22} & d_{23} & \dots & d_{2q} \\ \vdots & \vdots & \vdots & & \vdots \\ d_{l1} & d_{l2} & d_{l3} & \dots & d_{lq} \end{bmatrix}$$

为了体现视频的时域特性以及便于后面计算,将上述矩阵进行转置,最终得到结构特征信息为

$$\begin{bmatrix} d_{11} & d_{12} & d_{13} & \dots & d_{1l} \\ d_{21} & d_{22} & d_{23} & \dots & d_{2l} \\ \vdots & \vdots & \vdots & & \vdots \\ d_{q1} & d_{q2} & d_{q3} & \dots & d_{ql} \end{bmatrix}$$

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,每一帧分块情况为  $B_1, B_2, \dots, B_m$ ,以分块  $B_2$  为基准的视频时域管道可以描述为  $[B_{12}, B_{22}, B_{32}, \dots, B_{l2}]$ ,其中  $B_{ij}$  表示时域上第  $i$  帧的第  $j$  个分块,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  $[V_{1j}, V_{2j}, V_{3j}, \dots, V_{lj}]$ 。

针对每个时域管道,构建步长集  $[T_1, T_2, T_3, \dots, T_p]$  其中  $T_\beta \geq 1, 1 \leq \beta \leq p$ ,从步长集合中随机选择多个步长作为计算亮度均值差的分块步长。针对步长

$T_{\text{step}}$ , 计算公式为

$$dt(i, i + T_{\text{step}}) = V_{j(i+T)} - V_{ji} \quad (4)$$

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  $[dt_1, dt_2, dt_3, \dots, dt_\omega]$ , 其中  $\omega$  表示最终得到的时域管道亮度均值差的总数。最终能够得到时域特征矩阵为

$$\begin{bmatrix} dt_{11} & dt_{12} & dt_{13} & \cdots & dt_{1\omega} \\ dt_{21} & dt_{22} & dt_{23} & \cdots & dt_{2\omega} \\ \vdots & \vdots & \vdots & & \vdots \\ dt_{m1} & dt_{m2} & dt_{m3} & \cdots & dt_{m\omega} \end{bmatrix}$$

为了使得算法具有更好的时间复杂度与空间复杂度, 针对结构信息与时域信息均采用量化哈希的方式得到 01 位串, 结构哈希与时域哈希的产生情况为

$$Sh(Th) = \begin{cases} 0 & d_{ij}(dt_{ij}) < 0 \\ 1 & d_{ij}(dt_{ij}) \geq 0 \end{cases} \quad (5)$$

式中,  $i, j$  分别表示行与列。

对于给定视频片段  $vp$  与  $vq$ ,  $vp$  为原始视频段,  $vq$  为待检测的篡改视频。分别计算其结构哈希与时域哈希

$$\begin{aligned} Sp &= Hs(vp) \\ Sq &= Hs(vq) \\ Tp &= Ht(vp) \\ Tq &= Ht(vq) \end{aligned} \quad (6)$$

式中,  $Hs$  表示结构哈希计算过程,  $Ht$  表示时域哈希计算过程。

为了体现视频内容的标识性能, 采用相似度来表述视频内容的差异程度, 首先计算结构特征相似度

$$S = S_{\text{dis}}(Sp, Sq) \quad (7)$$

再计算时域特征相似度

$$T = T_{\text{dis}}(Tp, Tq) \quad (8)$$

接着分别对  $T, S$  进行归一化。当  $T \in (0.5, 1)$  时, 则考虑视频  $vq$  可能在时域上对视频  $vp$  有篡改。在后续的步骤中进行进一步的检测与定位。当  $T \in (0, 0.5]$  时, 表明  $vp$  与  $vq$  在时域上有超过 50% 的视频内容有差异, 可将  $vp$  与  $vq$  看作两个不同的视频。若  $T = 1$ , 则认为  $vp$  与  $vq$  两视频在时域上是相同的。同理, 可对结构特征相似度作相同的处理。

视频整体相似度为

$$Dis = \alpha S_{\text{dis}} + (1 - \alpha) T_{\text{dis}} \quad (9)$$

式中,  $0 < \alpha < 1$ ,  $S_{\text{dis}}$  表示结构相似度计算过程,  $T_{\text{dis}}$  表示时域相似度计算过程。 $\alpha$  的取值见后续算法步骤。

### 3 篡改检测与多粒度多层次定位

针对各种篡改行为, 采用多粒度的相似度矩阵检测定位策略。首先将原始视频与篡改视频感知哈希值进行相同的分段 (分段大小随机确定, 多数以粒度为参考目标) 处理, 在此基础上构建相似度矩阵。以原始视频为 X 轴, 以篡改视频为 Y 轴, 对于每个分段计算二者相似度 (海明距离等)。这里, 对角线反映二者对应的整体匹配程度, 通过计算相似度矩阵对角线相似度达到整体检测的目的。在没有被篡改的情况下, 理论上相似度矩阵会在对角线处达到最大值。算法整体框架见图 2。

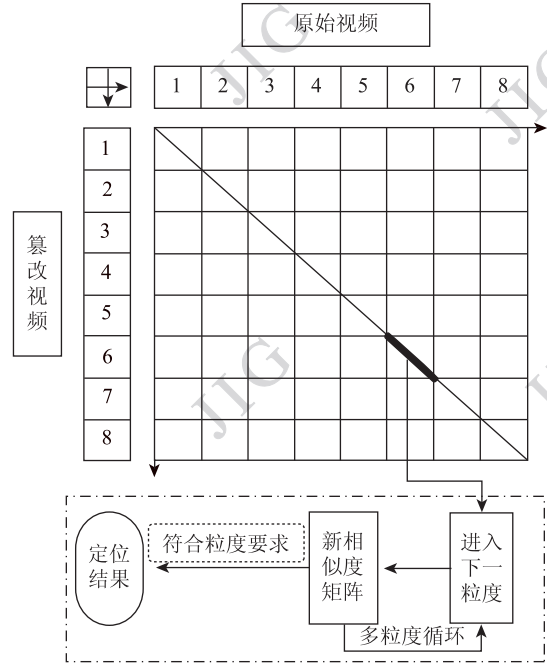


图 2 篡改检测与多粒度定位框图

Fig. 2 The framework of video forgery detection and multi-granularity location

篡改检测与多粒度定位的算法步骤如下:

1) 首先针对原始视频序列、篡改待检视频序列进行同样的分段预处理, 同时视频摘要哈希采用相同的分段策略, 即  $[c_1, c_2, c_3, \dots, c_n]$ ,  $[c_{h1}, c_{h2}, c_{h3}, \dots, c_{hn}]$ ,  $c_i$  表示视频片段的分段,  $c_{hi}$  表示相应哈希序列的分段。

2) 利用哈希序列分段构建相似度矩阵, 即

$$\begin{bmatrix} CC & c_{h1} & c_{h2} & c_{h3} & \cdots & c_{hn} \\ c'_{h1} & \ddots & & & & \\ c'_{h2} & & \ddots & & & \\ c'_{h3} & & & \ddots & & \\ \vdots & & & & \ddots & \\ c'_{hn} & & & & & \ddots \end{bmatrix}$$

其中,  $c_{hi}$  代表原始哈希分段序列,  $c'_{hi}$  代表篡改待检视频哈希分段序列, 利用二者构建的相似度矩阵, 重点分析对角线上的哈希匹配, 利用匹配值  $S$ 、 $T$ 、 $Dis$  度量二者的相似程度。

3) 分别对  $T$ 、 $S$  进行归一化。当  $T \in (0.5, 1)$  且  $S = 1$  时, 表明视频发生了时域篡改, 此时设  $\alpha = 0$ ; 当  $S \in (0.5, 1)$  且  $T = 1$  时, 则视频发生了空域篡改, 此时设  $\alpha = 1$ ; 当  $T \in (0.5, 1)$  且  $S \in (0.5, 1)$  时, 表明视频既发生了时域篡改, 也发生了空域篡改, 此时设  $\alpha = 0.5$ 。

4) 根据相似度矩阵对角线相似度匹配原则, 通过感知距离确定疑似篡改片段, 重点分析疑似片段, 再次进行视频哈希分段, 即

$$\begin{bmatrix} CC & c_{h1} & c_{h2} & c_{h3} & \cdots & c_{hn} \\ c_{h1} & \ddots & & & & \\ c_{h2} & & \ddots & & & \\ c_{h3} & & & \ddots & & \\ \vdots & & & & \ddots & \\ c_{hn} & & & & & \ddots \end{bmatrix} \rightarrow \begin{bmatrix} c_1 & c_2 & c_3 & \cdots & c_n \\ c_{h1} & c_{h2} & c_{h3} & \cdots & c_{hn} \end{bmatrix}$$

5) 根据给定粒度要求, 对特征哈希进行不同粒度的分段形式, 构建不同粒度的相似度矩阵, 最终实现相关粒度下的精确篡改定位。

当某一个分段的相似度跌入篡改范围, 便将其归入怀疑对象, 将该段可疑对象的特征哈希值提取, 进行下一层次的分析。具体算法模型如图 3 所示。

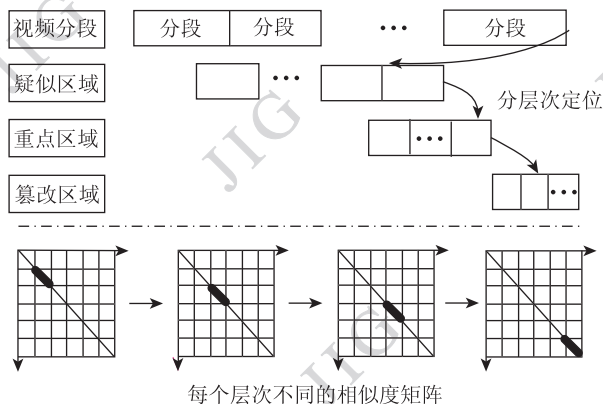
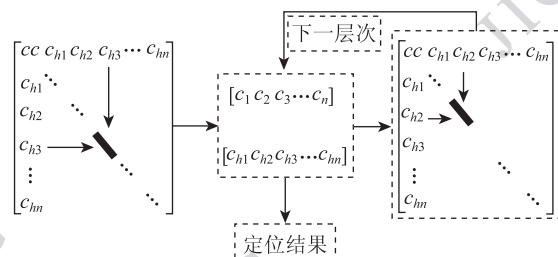


图 3 多层次定位

Fig. 3 Multi-granularity location

根据相似度矩阵对角线匹配原则, 通过相似度确定疑似篡改片段, 重点分析疑似片段, 提取疑似片段, 重复多粒度检测步骤进行不同层次的多粒度定位, 即



通过相似度矩阵多粒度、多层次篡改检测与定位算法, 完全符合人类识别视频信息由整体到局部, 由局部到重点的感知过程, 能够充分体现感知哈希技术在篡改检测定位过程中对于感知模型的具体描述。

## 4 实验结果与分析

采用美国国家自然科学基金资助下, 由美国亚利桑那州立大学提供的一系列标准视频数据库中的原始视频序列 (<http://trace.eas.asu.edu/yuv/index.html>) 约 10 G 左右的 CIF (352 × 288) 格式的视频文件作为实验视频数据。在 Intel Xeon 2.5 GHz, 4 GB 内存的 PC 机上使用 Matlab2009a 对算法进行编程实现。

为便于对实验数据进行分析, 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, 纵轴表示篡改哈希, 当两段视频完全相同时, 矩阵对角线相似度应该为 1, 利用灰度值表达 [0, 1] 区间, 从而构建拟合图。

相似度通过误码率 (BER) 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。对于每一次实验, 首先计算哈希值误码率, 然后用 1 减去误码率即为二者的相似度度量结果。主要针对两种篡改操作进行实验: 空域替换和时域替换。

### 4.1 视频空域篡改检测

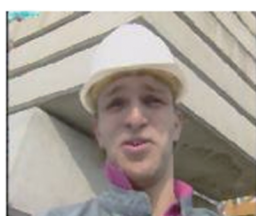
在空域篡改实验中, 首先给出感知哈希值对于篡改的准确区分以及利用相似度拟合图表达视频篡改攻击强度的大小情况, 就给定标准视频 Foreman, 就其中片段进行椒盐噪声处理, 采用噪声密度作为攻击强度的描述指标, 原始视频见图 4。

实验对视频片段进行 4 种不同强度的椒盐噪声处理, 处理后的视频帧如图 5 所示。

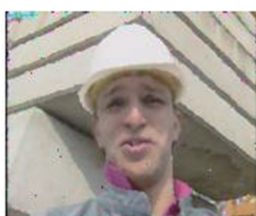


图4 原始视频帧

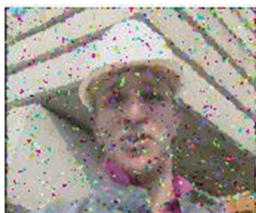
Fig. 4 The original video frame



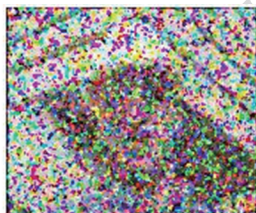
(a) 噪声密度 0.2



(b) 噪声密度 0.4



(c) 噪声密度 0.6



(d) 噪声密度 0.8

图5 不同攻击强度的视频处理结果

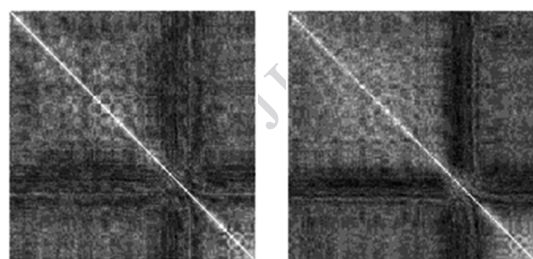
Fig. 5 Video processing results for different attack intensity

针对图5中不同强度的噪声处理情况,构建相似度矩阵,并且以相似度拟合图的形式表现实验结果,如图6所示。

由图5和图6的实验结果可知,视频在某一片段处受到了椒盐噪声攻击,通过相似度拟合图的灰度值变化可以发现,攻击强度不断加大。图6中攻击强度为0.2、0.4时,对角线灰度值变化微小,对应图5可以发现这两种攻击强度下的视频变化微小。图6中攻击强度为0.6、0.8时,相似度矩阵对角线灰度值变化较大,对应了如图5(c)(d)所示的视频巨大变化。实验结果符合人眼感知过程。

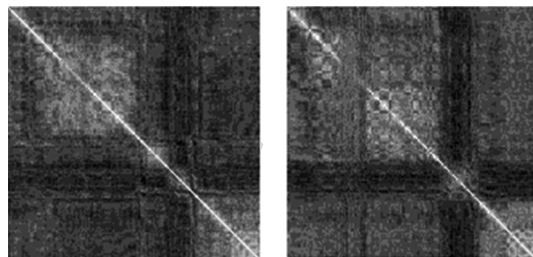
#### 4.2 时域片段替换检测与定位

在时域篡改实验中,采用完全不同的视频片段在不同位置对原始视频序列进行替换,通过原始视频与篡改视频摘要哈希的多粒度分段匹配,多层次定位,可以实现篡改位置的准确定位。



(a) 噪声密度 0.2

(b) 噪声密度 0.4



(c) 噪声密度 0.6

(d) 噪声密度 0.8

图6 不同噪声密度下的相似度拟合图

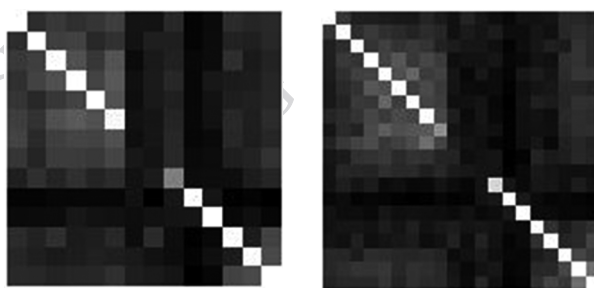
Fig. 6 Similarity fitting figures of different noise density

标准视频序列 Foreman ( $352 \times 288 \times 280$ ), 利用内容不同的视频序列 Bus ( $352 \times 288 \times 50$ ) 在第121帧处进行替换得到视频序列 ForemanInsert, 该视频第121—170帧处发生篡改替换, 首先给出多粒度定位相似度拟合图, 如图7所示。



(a) 分成7段

(b) 分成10段



(c) 分成14段

(d) 分成20段

图7 多粒度定位相似度拟合图

Fig. 7 Similarity fitting figures of multi-granularity location

图7(a)~(d)将视频分为7、10、14、20不同的粒度下进行视频篡改检测与定位,根据不同的精度

可以适当调整粒度要求,算法能够准确检测出视频是否发生篡改,不同的分段分别将篡改部位定位于不同的区间大小中,分段越多,绘制相似度拟合图的复杂度也就越高,故而结合不同粒度的定位分析,能够满足不同应用需求。

为了实现篡改部位的精确定位,采用多层次相似度计算的方法,对不同层次采用不同的分段策略。实验结果如图 8 所示。

由图 8(a)~(d)的结果可知,篡改部位在第 3 层次的分析上便可以进行精确定位,由第 1 层次、第 2 层次的分段计算确定了可疑片段,将可疑片段提取出来在第 3 层次进行多粒度分析,发现不同的分段情况下该段视频均表现为异常,由此可以断定此整段视频是非法替换的视频片段。

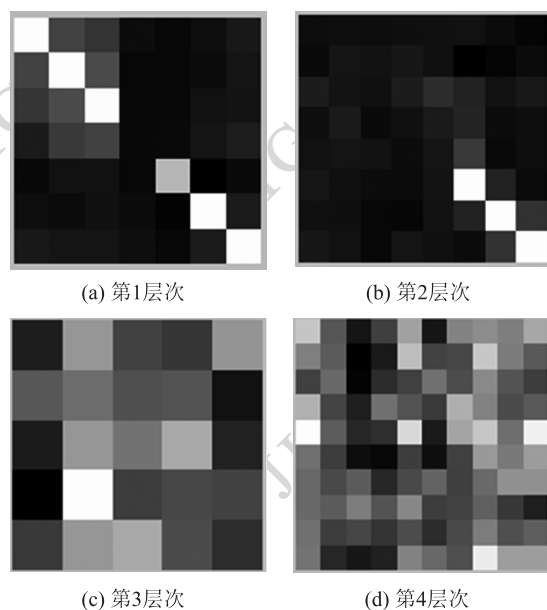


图 8 多层次相似度计算

Fig. 8 Multi-level similarity computation

### 4.3 时域顺序更改检测与定位

时域顺序更改是指视频片段在时间轴上发生了前后替换异常,帧位 200~300 的视频片段与帧位为 1~100 的视频片段进行了位置对换,从而使得视频内容发生改变。实验采用多粒度篡改检测定位的方法可以实现替换位置的准确定位。

采用标准视频序列 Foreman (352 × 288 × 280), 将视频第 21 帧后 50 帧长度的视频序列与第 201 帧开始后 50 帧长度的视频片段进行时域上位置对换,相似度拟合图实验结果如图 9 所示。

通过相似度矩阵的比对可以确定视频帧交替的

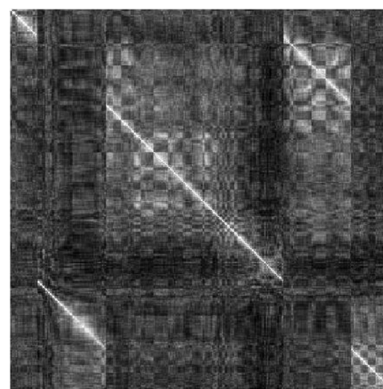


图 9 时域替换相似度拟合图

Fig. 9 Similarity fitting figure of time-domain replacement

位置,篡改视频的 21~70 帧、201~250 帧视频片段与原始视频的 201~250 帧、21~70 帧视频片段相似度最高,由此可以确定时域替换位置,实验结果显示,算法能够正确确定篡改位置。

### 4.4 查准率与查全率

采用查准率和查全率对算法进行性能评价。首先针对视频整体相似度(式(12))中  $\alpha$  的取值对查准率和查全率的影响进行实验。接着将本文算法与文献中的方法在相同的数据集上进行比较。

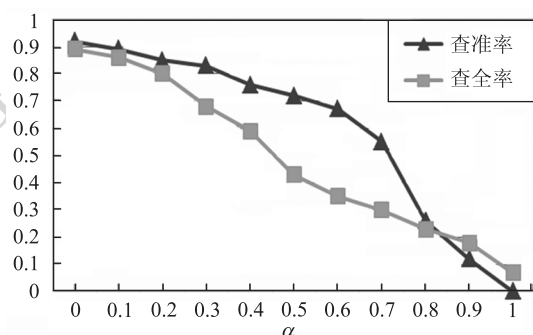
#### 1) $\alpha$ 取值的影响

分别对视频数据库中的 5 个视频片段: Bridge (Far)、Claire、Elephants Dream、Foreman 和 Highway 进行椒盐噪声处理(空域篡改)和视频帧顺序更改篡改(时域篡改)实验。 $\alpha$  选取 11 个值,分别从 0 到 1,步长  $\Delta\alpha = 0.1$  进行实验。实验结果如图 10 所示,纵轴为对 5 个视频分别进行时域篡改和空域篡改时的查准率平均值和查全率平均值。

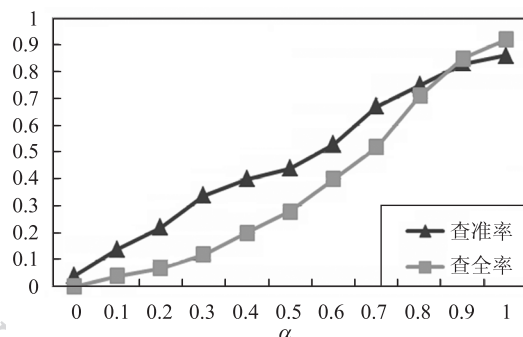
根据图 10 的实验结果,当视频的篡改仅为时域篡改时, $\alpha$  与查准率和查全率的关系为递减关系。这是因为时域特征相似度  $T$  起主导作用,影响因子  $\alpha$  越小越好,如图 10(a)所示。反之,当对视频的篡改仅为空域篡改时,因为结构特征相似度  $S$  起主导作用,影响因子  $\alpha$  越大越好。 $\alpha$  与查准率和查全率的关系为递增关系。如图 10(b)所示。当对视频的篡改,部分为时域篡改,部分为空域篡改时, $\alpha$  与查准率和查全率的关系为随机事件。

#### 2) 本文算法与其他方法比较

分别对视频数据库中的另 5 个视频片段: Akiyo(片段 1)、Big Bug Bunny(片段 2)、Bridge (Close)(片段 3)、Bus(片段 4)和 Foreman(片段 5)进行椒盐噪声处理和帧替换篡改实验。图 11



(a) 时域篡改



(b) 空域篡改

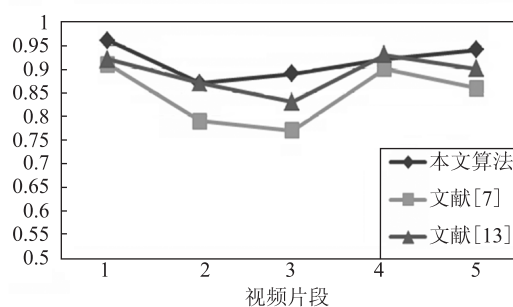
图 10  $\alpha$  取值对查准率与查全率的影响Fig. 10 The influence of  $\alpha$  to precision and recall

(a)一(d)为本文算法与文献[7]、文献[13]、文献[9]方法在时域、空域篡改实验中的查准率和查全率的比较结果。

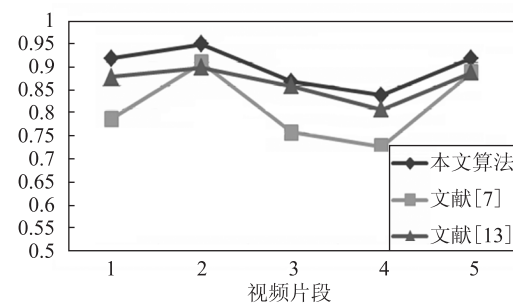
根据图 11 的结果,本文算法、文献[9]和文献[13]对 Bus 视频进行空域篡改检测时,查全率较低。这是因为 Bus 视频的视频帧内容相对于其他视频更为复杂,视频在篡改区域的量化噪声掩盖了本身的差异所导致的。从查准率与查全率来看,本文算法基本能正确检测出异源篡改位置,表现出较好的检测效果。

## 5 结 论

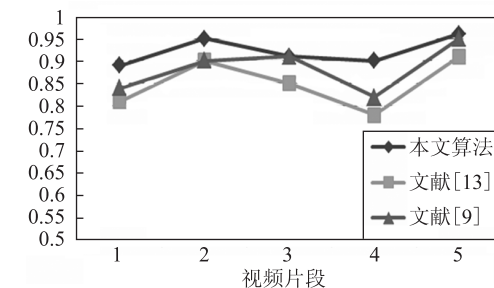
视频感知哈希通过视频感知内容和信息的提取,形成一个简短的数字摘要,通过摘要的匹配运算为视频信息的内容识别、认证和篡改检测提供技术支撑。结合视频时空域感知变化信息提取视频特征,在此基础上进行哈希量化。针对当前视频中普遍出现的篡改攻击,利用相似度矩阵进行多粒度、多层次篡改部位检测与定位。实验结果显示,相似度拟合图能够体现视频篡改攻击强度和攻击部位,对于时域、空域替换篡改均具有良好的效果。



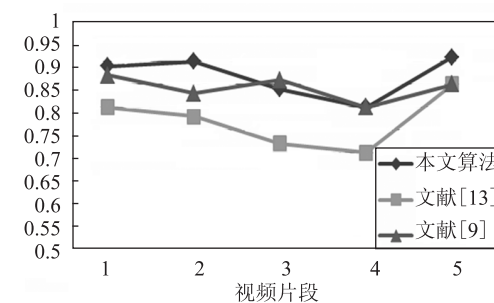
(a) 时域查准率



(b) 时域查全率



(c) 空域查准率



(d) 空域查全率

图 11 本文算法与其他方法查全率与查准率的比较结果

Fig. 11 The comparison results of precision and recall

## 参考文献 (References)

- [1] Tang M Z. Research on video forgery passive detection based on special-temporal feature [D]. Shanghai: Shanghai Jiao Tong University, 2012. [谭茂洲. 基于时空特征的视频篡改被动检测技术研究[D]. 上海:上海交通大学,2012.]
- [2] Lin Y R, Huang H Y, Hsu W H. An embedded watermark technique in video for copyright protection [C]//Proceedings of the

- 18th Int. Conf. on Pattern Recognition. Piscataway, NJ: Institute of Electrical and Electronics Engineers Inc. 2006: 795-798. [DOI: 10.1109/ICPR.2006.244]
- [3] Altun O, Sharma G, Celik M, et al. A set theoretic framework for watermarking and its application to semifragile tamper detection [J]. IEEE Transactions on Information, Forensics and Security, 2006, 1(4): 479-492. [DOI: 10.1109/TIFS.2006.885018]
- [4] De Roover C, De Vleeschouwer C, Lefebvre F, et al. Robust video hashing based on radial projections of key frames [J]. IEEE Transactions on Signal Processing, 2005, 53(10): 4020-4037. [DOI: 10.1109/TSP.2005.855414]
- [5] Wang W H, Farid H. Exposing digital forgeries in color filter array interpolated images [C]//Proc. of MM&Sec'07. New York: Association for Computing Machinery, 2007: 715-728.
- [6] Wang W H, Farid H. Exposing digital forgeries in video by detecting double MPEG compression [C]//Proceedings of the Multimedia and Security Workshop Geneva. Switzerland: ACM, 2006: 37-47.
- [7] Wang W H, Farid H. Exposing digital forgeries in video by detecting double quantization [C]//Proceedings of the 11th ACM Multimedia and Security Workshop. Princeton: ACM, 2009: 39-48.
- [8] Zhang J, Su Y T. Detecting logo-removal forgery by inconsistencies of blur [C]//Proceedings of 2009 International Conference on Industrial Mechatronics and Automation. Singapore: IEEE Industrial Electronics Society, 2009: 32-36. [DOI: 10.1109/ICIMA.2009.5156553]
- [9] Kobayashi M, Okabe T, Sato Y. Detecting forgery from static-scene video based on inconsistency in noise level functions [J]. IEEE Transactions on Information Forensics and Security, 2010, 5(4): 883-892. [DOI: 10.1109/TIFS.2010.2074194]
- [10] Qin Y L, Sun G L, Zhang X P. Exposing digital forgeries in video via motion vectors [J]. Journal of Computer Research and Development, 2009, 41(sup.): 227-233. [秦运龙, 孙广玲, 张新鹏. 利用运动矢量进行视频篡改检测[J]. 计算机研究与发展, 2009, 46(增刊): 227-233.]
- [11] Lee J W, Lee M J, Oh T W, et al. Screenshot identification using combing artifact from interlaced video [C]//Proceedings of the 12th ACM Workshop on Multimedia and Security. Princeton: ACM Multimedia and Security Workshop, 2010: 653-660.
- [12] Oostveen J, Kalker T, Haitsma J. Visual hashing of digital video: applications and techniques [J]. SPIE Applications of Digital Image Processing XXIV, 2001, 4472: 121-131.
- [13] 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: 3802-3809.
- [14] Coskun B, Sankur B, Memon N. Spatio-temporal transform based video hashing [J]. IEEE Transactions on Multimedia, 2006, 8(6): 1190-1280. [DOI: 10.1109/TMM.2006.884614]