

面向公路监控视频关键帧的感知哈希认证算法

丁凯孟^{1,2}, 张浩正¹

(1. 金陵科技学院网络与通信工程学院, 江苏 南京 211169;
2. 江苏省数据科学与智慧软件重点实验室, 江苏 南京 211169)

摘要: 针对公路监控视频关键帧的完整性认证问题, 提出一种感知哈希认证算法。首先, 采用帧间差分法提取公路监控视频的关键帧; 然后, 使用 YOLO(you only look once) 算法检测关键帧的主要特征对象(道路车辆), 并对检测的特征对象进行融合; 最后, 分别提取特征对象融合结果和关键帧的低频信息, 作为感知特征生成关键帧的感知哈希序列。在关键帧的完整性认证过程中, 通过比对两组关键帧的哈希序列, 可以对视频可能存在的篡改进行检测。实验结果表明, 感知哈希认证算法具有良好的鲁棒性与可区分性, 可以有效地对公路监控视频关键帧进行完整性认证。

关键词: 感知哈希认证算法; YOLO 算法; 关键帧; DCT

中图分类号: TP391

文献标识码: A

文章编号: 1672-755X(2021)02-0001-06

Perceptual Hash Authentication Algorithm for Key Frame of Highway Surveillance Video

DING Kai-meng^{1,2}, ZHANG Hao-zheng¹

(1. Jinling Institute of Technology, Nanjing 211169, China;
2. Jiangsu Key Laboratory of Data Science and Smart Software, Nanjing 211169, China)

Abstract: Aiming at the integrity authentication of key frame of highway surveillance video, this paper proposes a perceptual Hash authentication algorithm. Firstly, the key frames of highway surveillance video are extracted by inter-frame difference method. Then, the main feature objects(road vehicles) of the key frame are detected by using you only look once(YOLO) algorithm. Finally, low frequency information of the fusion result and key frame are extracted as perceptual features to generate perceptual Hash sequence of key frames. In the process of key frame integrity authentication, the possible tampering of video can be detected by comparing the Hash sequences of two groups of key frames. The experimental results show that the perceptual Hash authentication algorithm has good robustness and distinguishability, and can effectively authenticate the integrity of key frames in highway surveillance video.

Key words: perceptual Hash authentication algorithm; YOLO algorithm; key frame; DCT

作为监控视频的重要分支, 公路视频监控技术在交通领域发挥着非常重要的作用, 是一种重要的公共安全保障技术。随着多媒体应用技术和视频数据处理技术的发展, 监控视频的编辑与修改难度大大降低, 使得监控视频在传输与存储过程中极易面临各种有意或者无意的“篡改”^[1]。被篡改的公路监控视频不但

收稿日期: 2021-01-14

基金项目: 国家自然科学基金(41801303); 江苏省自然科学基金(BK200170116); 金陵科技学院博士科研启动基金(jit-b-201520); 金陵科技学院校级科研基金孵化项目(jit-fhxm-201604); 2019年度江苏省高校青蓝工程资助项目

作者简介: 丁凯孟(1985—), 男, 山西长治人, 副教授, 博士, 主要从事深度学习、数据安全等研究。

不能发挥应有的作用,还可能带来严重的负面影响。这不仅损害了监控视频的可靠性和真实性,还对交通管理、案件侦破等造成影响。因此,对于公路监控视频的完整性认证技术的研究具有重要的科研意义与社会意义。

本文针对公路监控视频关键帧的完整性认证问题,结合感知哈希与深度学习理论,提出面向公路监控视频关键帧完整性认证的感知哈希认证算法,以实现公路监控视频关键帧的完整性认证。

1 理论背景

1.1 面向监控视频的认证算法概述

目前,监控视频完整性认证的主要技术途径大致分为以下三类:1)基于数字水印的认证算法^[2]。这一类算法一般通过水印嵌入技术将水印信息嵌入至视频中,并在接收端提取嵌入的水印信息,进而判断视频数据是否存在篡改。但是,基于数字水印的认证算法需要在视频中嵌入水印信息,会对视频造成一定程度的修改,损坏原有视频的信息完整性。2)基于感知哈希的认证算法。视频感知哈希认证算法作为感知哈希的一类,实质上是一种基于视频数据的单项映射算法^[3],可以生成每个视频帧的哈希值,并通过密钥保障感知哈希序列自身的安全性。视频感知哈希认证算法不需要对视频进行修改,不会破坏视频自身的完整性。但是,现有的视频感知哈希认证算法一般没有顾及公路监控视频的特点,在公路监控视频的完整性认证中存在不足之处。3)基于深度学习的哈希认证算法。这一类算法本质上是感知哈希的一个分支,将深度学习理论与感知哈希相结合,并将其应用到视频数据的认证中。例如,基于 3D 卷积神经网络的哈希认证算法把 3D 卷积神经网络与哈希学习方法进行结合,通过 3D 卷积网络对视频数据进行时间特征和空间特征的提取与融合,再使用哈希学习方法对融合后的视频时空特征进行量化编码,最后得到视频的哈希二值码^[4]。但是,基于深度学习的哈希认证算法同样未能顾及公路监控视频的应用环境与特点,如公路监控视频主要涉及车辆等信息,而道路和路边环境涉及得较少。

1.2 感知哈希

感知哈希,实质上是一种基于感知内容的单项映射^[5-7],可以将具有相同内容的视频、图像等映射成为一组固定长度的二进制字符串,同时满足人们对于安全性与鲁棒性的要求。因此,感知哈希被广泛应用于图像、视频、音频等多媒体数据的检索、认证、篡改检测等领域。应用于多媒体数据完整性认证的感知哈希算法,也被称作“感知哈希认证算法”。感知哈希除拥有单向性与摘要性等特征外,还通常具备如下特征:1)鲁棒性。指具有相同内容的不同多媒体对象所映射出的感知哈希序列相同或者相似。2)安全性。指无法从所提取的感知哈希序列中得到能够反映多媒体对象的有效信息。3)唯一性。指内容不同的多媒体对象所映射出的感知哈希序列应产生较大的差异,每个序列只对应唯一的多媒体对象。

本文在分析公路监控视频应用环境与特点的基础上,结合经典的深度学习模型 YOLO 算法,研究面向公路监控视频关键帧的感知哈希认证算法。

1.3 YOLO 算法

公路监控视频关键帧中最具价值的往往是车辆等物体的相关信息。本文提出的感知哈希认证算法针对监控视频的这一特点,通过基于深度学习的目标检测算法来提取公路监控车辆的车辆等信息,对车辆等信息进行更为严格的完整性认证。

近年来,基于深度学习模型的目标检测算法取得了长足发展,目前大致可以分为两类:一类是两级检测算法,如基于候选框的 R-CNN(regions with convolutional neural networks,R-CNN)系算法(主要包括 R-CNN 算法、Fast R-CNN 算法、Faster R-CNN 算法)^[8-9],该类算法准确度普遍较高,但运行速度较慢。另一类是单级检测算法如 YOLO 算法^[10],仅使用一个 CNN 网络便可以直接预测不同目标的类别与位置,这一类算法速度快,但准确性相对较低。本文考虑到公路监控视频认证的实时性,结合深度学习模型 YOLO 算法实现公路监控视频的目标检测。

1.4 视频关键帧提取

视频关键帧^[11-12]用于描述一段视频的关键图像,还能够反映出一个镜头的关键内容。公路监控视频关键帧能够清楚地反映出在镜头内的道路车辆信息以及路面异常信息的图像帧^[13]。关键帧的使用可以

大幅度减少视频索引与监控视频认证的工作量,提高认证效率。目前有如下几种关键帧的提取方法:1)基于镜头边界提取关键帧^[14]。该类算法通过将一段视频分割成一系列的镜头之后,将每个镜头的第一帧或者最后一帧作为镜头的关键帧。但是,每个镜头的首帧与尾帧并不能完全反映出镜头的主要内容,所提取的关键帧可能不具有整体代表性。2)基于视频聚类提取关键帧^[15]。此类算法一般通过聚类将镜头中的帧序列分配到各个簇,在每个簇内选取能够代表这类信息的帧作为关键帧。该算法的计算量较小,提取的关键帧也可以较好地代表视频的关键内容。3)基于视频帧间差分提取关键帧^[16]。该类算法逐帧地将视频数据转换为若干张图片数据,将每两帧图像进行差分,利用平均像素强度来衡量两帧图像的变化大小,提取平均像素强度变化较大的帧画面作为关键帧。

由于公路监控视频的摄像头安装位置相对固定,不需要进行镜头分割。所以,本文综合公路监控视频认证内容的单一性与固定性,采用基于视频帧间差分的方法提取监控视频的关键帧。

2 算法流程与认证过程

针对公路监控视频涉及的主要内容为车辆等信息,而道路和路边环境涉及较少这一特点,本文提出了一种面向公路监控视频关键帧的感知哈希认证算法,对监控视频关键帧的车辆等重要信息进行更为严格的认证。

2.1 算法流程

本文算法流程如图 1 所示,主要包括关键帧的提取、特征图像的检测与提取、哈希序列的生成等步骤。

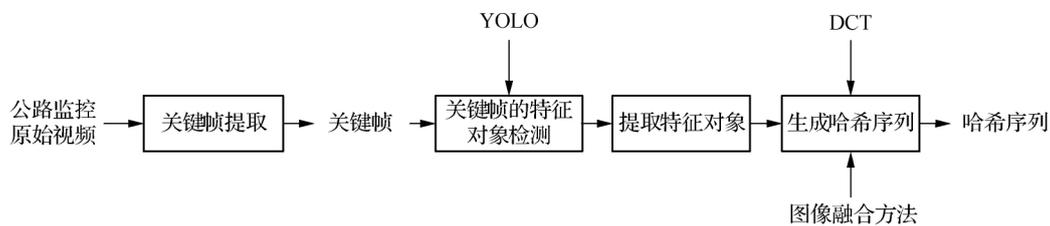


图 1 感知哈希认证算法流程

2.1.1 关键帧中特征对象的检测与提取

本文以车辆作为特征对象,采用 YOLO v3 模型实现视频关键帧(记为 KF)特征对象的检测,并通过对框内物体置信度的判断,完成特征对象的提取,得到 $T(T \geq 0)$ 个特征对象图像,并分别记为 O' ($T \geq t \geq 0$)。

2.1.2 特征对象图像的融合

对关键帧的特征对象图像 O' 进行融合,得到特征对象的融合结果,记为 FI 。在特征对象图像的融合过程中,本文采用平均加权的方式,即:

$$FI = \frac{1}{T}O^1 + \frac{1}{T}O^2 + \dots + \frac{1}{T}O^t \quad (1)$$

由公式(1)可知,融合过程中每个特征对象图像 O' 的权重相同,且权重之和为 1。

2.1.3 感知特征的提取

感知特征的提取是感知哈希算法的关键。常用的感知特征提取方法包括基于离散小波变换(discrete wavelet transform, DWT)的方法^[17-18]、基于特征点的方法^[19-20]、基于图像统计特征的方法^[21]、基于离散余弦变换(discrete cosine transform, DCT)的方法^[22-24]等。其中,基于 DCT 的方法是目前应用最为广泛的特征提取方法,具有较好的篡改敏感性、摘要性、鲁棒性、唯一性等优点^[22]。

为了使算法具有更好的通用性,本文采用了基于 DCT 的方法提取关键帧 KF 和融合结果 FI 的感知特征。基于 DCT 方法提取感知特征的具体流程为:1)将图像大小归一化为 64×64 像素,以减小 DCT 的计算量;2)归一化之后的图像均匀划分为 8×8 大小的矩阵,并对每个矩阵进行 DCT 分解。

设单通道图像(二维矩阵)的大小为 $M \times N$,则对应的 DCT 变换可以表示为:

$$F(u, v) = c(u)c(v) \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f(i, j) \cos \left[\frac{(i+0.5)\pi u}{M} \right] \cos \left[\frac{(j+0.5)\pi v}{N} \right]$$

$$c(u) = \begin{cases} \sqrt{\frac{1}{M}}, & u = 0 \\ \sqrt{\frac{2}{M}}, & u \neq 0 \end{cases}, \quad c(v) = \begin{cases} \sqrt{\frac{1}{N}}, & v = 0 \\ \sqrt{\frac{2}{N}}, & v \neq 0 \end{cases} \quad (2)$$

其中, M 和 N 分别表示二维矩阵的行坐标和列坐标; $i, u=0, 1, 2, \dots, M-1; j, v=0, 1, 2, \dots, N-1; f(i, j)$ 为原始信号; $F(u, v)$ 为 DCT 变换后的系数; $c(u), c(v)$ 为补偿系数。通过 DCT 变换得到正交矩阵, 该正交矩阵的左上角呈现出图像的低频信息。

2.1.4 哈希序列生成

对提取原始关键帧和融合结果的低频信息, 分别进行量化, 得到 2 个 64 bit 的子序列, 分别记为 PH_k 和 PH_f 。关键帧的感知哈希序列 PH 就是 2 个子序列的串联, 即

$$PH = PH_k \parallel PH_f \quad (3)$$

2.2 完整性认证过程

本文算法的完整性认证流程如图 2 所示。采用 2.1 节的算法流程生成待认证视频关键帧的感知哈希序列, 并与原始关键帧的感知哈希序列进行比对, 从而实现关键帧的完整性认证。

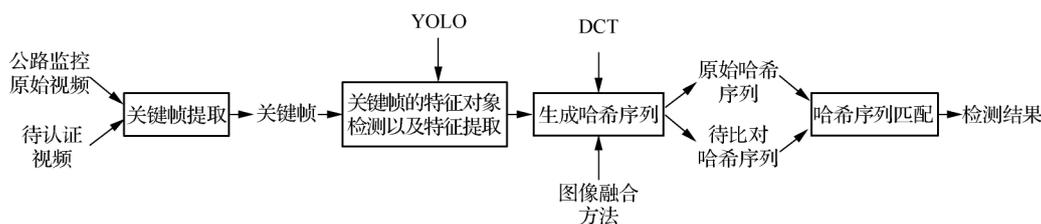


图 2 本文算法完整性认证流程

3 实验与分析

为了验证本文算法的有效性, 本文以南京某高校校园内的道路监控视频作为实验数据来进行算法测试。选取的道路监控视频时长 30 s, 分辨率为 1280×720 像素。

实验硬件平台为 1.60 GHz 主频的 CPU 和 NVIDIA GeForce MX150; 软件平台为 Windows 10。采用 Python 3.7 作为开发语言, 并基于 OpenCV 实现部分功能。

如本文 2.1 节所述, 基于 DCT 的感知哈希算法是较为常见的, 也是比较有代表性的感知哈希算法。所以, 本文采用基于 DCT 的感知哈希算法作为对比算法, 并采用相同的软硬件平台, 提取公路监控视频的关键帧, 结果如图 3 所示。



图 3 本文算法运行过程中提取的部分关键帧

本文算法的特点在于能够对目标对象(车辆)相关的篡改进行更加灵敏的检测, 因此从“车辆相关”和“车辆无关”这两个方面来验证算法的有效性。图 4 是对图 3 中的关键帧进行车辆相关信息篡改后的关键帧, 图 5 是对图 3 中的关键帧进行车辆无关信息篡改后的关键帧。



图 4 车辆相关信息篡改后的关键帧



图 5 车辆无关信息篡改后的关键帧

3.1 本文算法结果

针对图 4 和图 5 所示的篡改,本文算法的实验结果如表 1 所示。该算法生成的感知哈希序列 PH 包含 2 个子序列 PH_k 和 PH_f , PH_k 表示感知哈希序列原始关键帧的子序列, PH_f 表示感知哈希序列关键帧篡改后的子序列。从表 1 可以看出,该算法不仅可以对对象特征(车辆)无关信息的篡改关键帧进行准确检测,还能对对象特征(车辆)相关信息的“细微篡改”关键帧进行检测。

表 1 公路监控视频关键帧完整性认证算法结果

关键帧	原始的感知哈希序列(PH_k)	对象特征相关信息篡改后的感知哈希序列(PH_f)	检测结果	对象特征无关信息篡改后的感知哈希序列(PH_f)	检测结果
关键帧 A	CCCC-4C14-7040-0002	CCCC-4C14-7040-0002	识别	E8C8-4C0C-C6C0-0323	识别
	ECF1-1206-8304-2568	E823-1286-8314-2760		ECF0-0286-8384-2578	
关键帧 B	E4CC-0C0C-A4E0-0001	E4CC-0C0C-A4F0-0001	识别	E4C4-0406-B4F1-0000	识别
	E037-3FD0-8400-2840	E035-2DD0-9444-6040		E037-3FD0-8400-2840	
关键帧 C	E4CC-0C0C-B022-0108	E4CC-0C0C-B062-0108	识别	C4EC-0CC4-F060-010F	识别
	DA20-2880-80C4-0000	E209-08A6-80C8-2281		DA20-2880-80C4-0000	

3.2 算法对比结果

为突出本文算法的特点,这里选取传统的基于 DCT 的感知哈希算法作为对比算法,同样采取上述 3 组关键帧数据作为实验对象进行完整性认证,实验结果如表 2 所示。由表 2 可知,对于特征对象无关信息篡改后的关键帧图像,本文算法与传统基于 DCT 感知哈希算法均能实现关键帧的完整性认证,表明能够检测可能存在的篡改。但对于特征对象相关信息篡改后的关键帧图像,本文算法相对于传统算法具有较大优势,即便特征对象被细微篡改,本文算法也能够进行有效准确的检测,完成监控视频关键帧的完整性认证。

表 2 算法对比结果

算法	特征对象相关信息篡改后的关键帧图像			特征对象无关信息篡改后的关键帧图像		
	关键帧 A	关键帧 B	关键帧 C	关键帧 A	关键帧 B	关键帧 C
本文算法	识别	识别	识别	识别	识别	识别
传统算法	未识别	未识别	识别	识别	识别	识别

4 结 语

本文算法根据公路监控视频的特点,利用深度学习在特征提取方面的优势,使用基于 DCT 提取关键帧与融合结果的低频信息,实现监控视频关键帧的感知哈希认证算法。该算法与传统的感知哈希认证算

法相比,结合了道路监控视频的位置固定化、信息复杂化、实时性强等特点,能够对公路监控视频中的主要特征对象(车辆)进行更为灵敏的篡改检测,具有较强的可区分性。

在算法实现过程中,为了使本文算法具有更好的通用性,采用了基于 DCT 的方法提取关键帧的感知特征。在实际应用中,用户也可以根据实际需求,采用其他的感知特征提取方法(算法流程不变),比如面向视频感知特征提取的深度神经网络,以进一步提升感知哈希算法的认证性能。

参考文献:

- [1] 朱映映,文振焜,杜以华,等.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
- [2] 彭许红,陈威兵,夏明,等.基于可逆水印的监控视频隐私信息保护与隐藏[J].小型微型计算机系统,2014,35(6):1226-1231
- [3] 欧阳杰,高金花,文振,等.融合 HVS 计算模型的感知哈希算法研究[J].中国图象图形学报,2011,16(10):1883-1889
- [4] 刘玉莹,刘宏哲,袁家政,等.基于 3D 卷积神经网络的视频哈希算法[J].计算机应用研究,2020,37(3):887-890
- [5] 牛夏牧,焦玉华.感知哈希综述[J].电子学报,2008,36(7):1405-1411
- [6] 张维克,孔祥维,尤新刚.安全鲁棒的图像感知哈希技术[J].东南大学学报(自然科学版),2007,37(1):188-192
- [7] CHEN C, DING F, ZHANG D. Perceptual hash algorithm-based adaptive GOP selection algorithm for distributed compressive video sensing[J]. IET Image Processing, 2017, 12(2): 210-217
- [8] MINARI G H, SILVA F, PEREIRA D R, et al. Anomalies identification in images from security video cameras using mask R-CNN[J]. IEEE Latin America Transactions, 2020, 18(3): 530-536
- [9] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149
- [10] WU W, LI Q. Machine vision inspection of electrical connectors based on improved YOLO v3[J]. IEEE Access, 2020(8):166184-166196
- [11] 王娟,蒋兴浩,孙锁锋.视频摘要技术综述[J].中国图象图形学报,2014,19(12):1685-1695
- [12] LUO J, PAPIN C, COSTELLO K. Towards extracting semantically meaningful key frames from personal video clips: from humans to computers[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2009, 19(2): 289-301
- [13] 蓝章礼,帅丹,李益才.基于帧间相关性的道路监控视频关键帧提取[J].微电子与计算机,2015,32(5):51-54
- [14] 卜庆凯,胡爱群.一种新的镜头边界检测和静态视频摘要提取方法[J].东南大学学报(自然科学版),2007,37(4):559-564
- [15] 朱映映,周洞汝.一种基于视频聚类的关键帧提取方法[J].计算机工程,2004(4):12-13
- [16] 李亮,罗毅.帧间差分法在视频监控中的应用研究[J].四川理工学院学报(自然科学版),2015,28(6):58-62
- [17] SWAMINATHAN A, MAO Y, WU M. Robust and secure image hashing[J]. IEEE Transactions on Information Forensics and Security, 2006, 1(2): 215-230
- [18] QIN C, CHANG C C, TSOU P L. Robust image hashing using non-uniform sampling in discrete fourier domain[J]. Digital Signal Processing, 2013, 23(2): 578-585
- [19] YANG L, CHEN Q, TIAN J, et al. Content based image hashing using companding and gray code[J]. Security and Communication Networks, 2011, 4(12): 1378-1386
- [20] LV X, WANG Z J. Perceptual image hashing based on shape contexts and local feature points[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(3): 1081-1093
- [21] HUANG Z, LIU S. Robustness and discrimination oriented hashing combining texture and invariant vector distance [C]//Proceedings of the 26th ACM international conference on multimedia, Singapore; IEEE Xplore, 2018: 1389-1397
- [22] CHANG L, YAN W G, WANG W D. Research on robust image perceptual hashing technology based on discrete cosine transform[M]. Springer; Berlin Heidelberg, 2012: 799-809
- [23] JIAO Y, JI L, NIU X. Robust speech hashing for content authentication[J]. IEEE Signal Processing Letters, 2009, 16(9): 818-821
- [24] TANG Z, LAO H, ZHANG X, et al. Robust image hashing via DCT and LLE[J]. Computers & Security, 2016, 62: 133-148

(责任编辑:谭彩霞)