



面向内容网络环境的视频内容标识方案

陈勋¹, 韩帅^{1,2}, 刘姗姗¹, 阳小龙¹

(1. 北京科技大学, 北京 100083; 2. 中国洛阳电子装备试验中心, 河南 洛阳 471000)

摘要: 视频指纹技术是当前解决海量视频数据标识问题的一种有效方案。首先阐述了内容网络的研究现状以及存在的主要问题, 尤其在内容标识方面的问题; 然后从内容网络中的视频内容标识需求出发, 提出视频内容标识方案的设计原则, 并深入分析已有方案的原理和缺点; 最后在解析视频指纹技术概念与原理之后, 设计并实现一种基于视频指纹技术的视频内容标识方案, 其特点为: 将时域和频域两种类型的视频指纹技术进行融合, 在缩短了视频标识生成时间的同时, 提高了整体系统的顽健性。

关键词: 内容网络; 内容标识; 视频处理; 视频指纹

中图分类号: TN919.85

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2018182

Video content identification scheme in content oriented network environment

CHEN Xun¹, HAN Shuai^{1,2}, LIU Shanshan¹, YANG Xiaolong¹

1. University of Science and Technology Beijing, Beijing 100083, China

2. Luoyang Electronic Equipment Test Center of China, Luoyang 471000, China

Abstract: Video fingerprinting technology is a video identification technology proposed for the identification of massive video data. Firstly, the research status of the content network and the main problems that exist were elaborated, especially the problems of content identification. Then design principles of the video content identification scheme were proposed from the video content identification requirements in the content network, and the principles and disadvantages of the existing solutions were analyzed in depth. Finally, a video content identification scheme based on video fingerprinting technology was designed and implemented after analyzing the concept and principle of video fingerprinting technology. The feature was: the two types of video fingerprinting technology in the time domain and frequency domain were integrated. While shortening the time for generating video tags, it also improved the robustness of the overall system.

Key words: content network, content identification, video processing, video fingerprint

1 引言

随着互联网技术的进步和终端处理能力的

增强, 相对于传统的文字、语音和图像等内容形式, 视频业务已然从多样的网络内容中脱颖而出, 并成为内容网络中数据流量的主要贡献者。

收稿日期: 2018-04-08; 修回日期: 2018-05-06

通信作者: 阳小龙, yangxl@ustb.edu.cn



根据思科 (Cisco) 公司预计, 到 2019 年, 全球互联网视频的数据流量将达到 10^5 EB/月, 相比 2014 年的 25 EB/月有相当显著的增长。与此同时, 4G、Wi-Fi、小型基站的普及和 5G 等新技术的出现, 使得移动互联网呈现高速发展态势, 其业务由传统的单一化形式向多元化方向发展, 如移动视频、即时通信、移动社交网络等。华为在 2017 年世界移动通信大会 (MWC) 上的报告显示, 预计到 2021 年, 移动网络中的视频流量占比将达到 58%。但是爆发式的增长也带来了种种问题, 例如: 视频种类繁多, 不易于管理, 视频格式与用户终端设备不匹配, 无线带宽波动影响用户体验等, 这些问题都给视频内容的高效分发和高体验质量共享带来了一定的挑战。

目前, 用户仍然是通过传统网页或者相关视频软件来获取所需要的视频资源。这种获取视频内容的方式, 在当前僵化的网络传输协议下, 容易受到一定的限制。比如: 浏览器通过 URL (uniform resource locator) 来获取相关网页视频信息时, 视频资源将会被特定的协议、主机号以及相关文件名所绑定, 导致资源不能够被灵活地分享和调度。同时, 由于硬件设备技术的不断发展, 用户终端的种类日益繁多, 使视频资源在这种异构网络形态下的传播变得困难, 从而出现用户获取的视频内容在编码格式、呈现方式等方面与其终端设备能力不匹配的情况, 这将影响用户的观看体验。虽然当前网络中同一内容视频内容普遍存在多个不同的版本 (如格式、分辨率等), 由此以不同版本为不同用户提供视频内容服务, 但是这极大地提高了视频内容管理和分发难度。

当前有一种以内容为中心的新型网络架构, 在解决以连接为中心的传统互联网架构与当前以信息或内容共享为中心的网络服务模式冲突的同时, 也解决了视频数据分发与服务的问题。其中 CNN (content-centric networking, 内容中心网络) 和 NDN (named data networking, 命名数据网络) 是

两个著名的内容网络体系结构^[1-2]。它们的共同点在于: 都试图将内容和地址分离, 其中内容可以是任意形式的数据 (包括视频、文本和音乐等)。这些内容都需要被标识, 才能被用户所使用。但在 CCN/NDN 新型网络架构中, 所有的内容都被同样处理, 而没有将视频资源进行特殊对待, 提取的标识特征都是基于底层字节信息。因此, 这导致同一内容的视频资源由于编码的不同方式从而产生不同标识特征, 不利于最终的分发与管理。所以, 需要一个实质内容感知的视频特征提取方案, 能够给同一内容的视频资源提供相同的特征标识。对此, 视频指纹技术很好地解决了上述问题, 并在海量视频数据分类与管理中取得了一定的成果。视频指纹作为视频资源管理保护的新手段, 就像人类指纹一样可以唯一标识身份。同时视频指纹也是数字视频内容的精简数字化表示形式, 通过对视频进行分析、提取、计算处理而形成的一个唯一标识符。视频指纹应用广泛, 可用于视频检索、视频认证、监视视频广播以及视频过滤, 当然也能够应用于视频标识管理方面。本文设计并实现了一种基于时-频域视频指纹技术的视频标识方案, 利用视频指纹技术在解决视频标识管理问题的同时, 也为视频分发与分类提供一定的便利。

2 内容网络

2.1 技术特点

随着互联网应用的不断发展变化, 网络中的资源内容日渐繁多, 基于 TCP/IP 的现有互联网也逐渐暴露出许多不适应。当前, 互联网上主要暴露的问题有: 不安全、移动性差、可靠性差和灵活性差。为了解决这些问题, 国内外很多研究机构都进行了相关研究, 并提出了对应的解决方案。内容中心网络作为一种新型的网络体系架构, 其中 CCN/NDN 是最典型的解决方案之一。因此本文以 CCN/NDN 为代表阐述内容网络的技术特点。

CCN 是 Van Jacobson 在 2009 年提出的, 在 CCN 的基础上, 对 CCN 及其涉及的关键技术展开研究^[3]。因此, CCN 与 NDN 本质上是同一种网络架构。CCN/NDN 的主要思想是将信息对象作为构建网络的基础, 分离信息的位置信息与内容识别, 通过内容名字而不是主机 IP 地址获取数据, 从而实现高效、可靠的信息分发。其目的是开发一个可以适应当前通信模式的新型互联网架构, 从以“where”为中心的架构转化为以用户和应用所关心的“what”为中心的架构^[4]。综合参考文献[5-8], CCN/NDN 与当前互联网架构相比有如下特点。

(1) 体系结构

虽然 CCN/NDN 体系结构的外形和当今 TCP/IP 网络很相似, 都是沙漏模型, 如图 1 所示^[5], 但其不同是在“瘦腰”处用内容块 (content chunk) 代替了 IP 地址。从网络的角度看, 就是用对数据命名代替了对实体接口地址的命名。另外, 网络中内建存储功能, 用来缓存经过的数据, 用以缩短其他用户访问同样数据的响应时间, 同时可大大减少网络中的流量。

(2) 缓存策略

CCN/NDN 与传统网络的另一个不同点是, 它充分利用了网络设备存储容量大、存储代价小的优势, 而适当地运用缓存机制, 从而使网络性能得

到进一步的提升。CCN/NDN 目前有很多研究集中在缓存策略的设计上, 包括缓存替换策略、放置决定策略、管理技术等。现有大部分研究主要分成两类: 通过建立数学模型或仿真实验的方法, 评估 CCN/NDN 现有缓存策略性能; 提出新型缓存策略来提升网络性能。根据参考文献[9], CCN/NDN 缓存具有透明性、普遍性以及细颗粒度等特点。

(3) 安全性

传统网络是通过主机到主机的链接进行数据传输的, 因此在安全方面存在很大的缺陷。在 CCN/NDN 中, 安全机制是针对信息内容本身的, 对于每个信息对象, 由发布者进行签名或加密, 然后接受者对签名进行认证或者解密, 从而判断信息的有效性。

(4) 命名机制

在 CCN/NDN 体系中, 整个网络的需求是内容, 而不是主机, 因此网络不再关心内容存储在哪里, 而仅关心内容本身。它从根本上改变了 IP 分组的封装结构和寻址方式, IP 报文的分组头不再以 IP 地址作为标识, 而是以内容名称作为标识。内容的标识名其实就是 CCN 的地址, 这种标识名是层次化的, 通常由几个部分组成, 通过这种层次化的地址, 可以利用类似于 URL 前缀机制来迅速

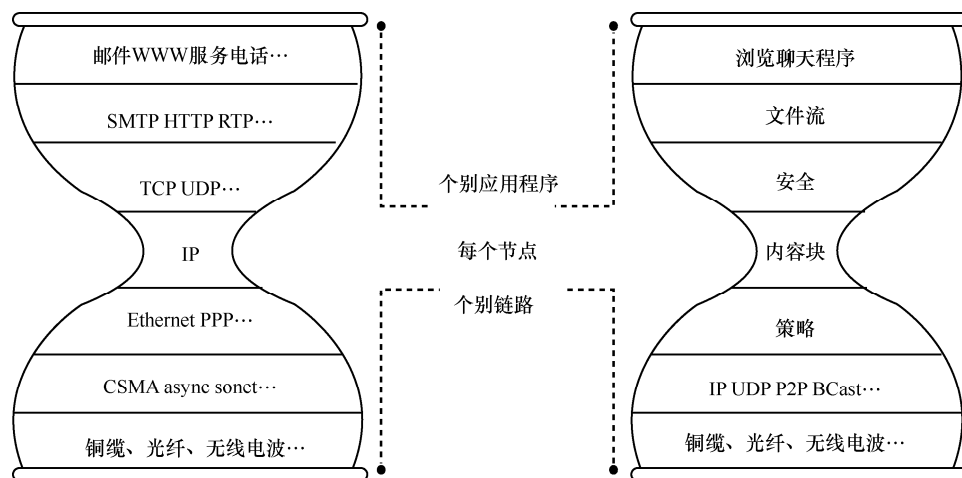


图 1 IP 和 CCN 的比较^[5]



定位所需要的信息。如“/cctv.cn/videos/WidgetA.mpg/_v(timestamp)/_s1”，其中“cctv.cn”是全网可识别名称，“videos”是内容类型，“WidgetA.mpg”是内容名称，“_v(timestamp)”是版本时间等信息，“_s1”是分段等信息^[10-11]。可以看出，内容名可以分为内容名称和分片名称两个部分。

2.2 存在的主要问题

通过对 CCN/NDN 技术特点的描述，内容网络主要是将内容和地址分离，其中内容可以是任何数据形式，这些内容都需要被标识后，才能够被用户使用。因此，如何对内容网络中的内容信息进行标识是首要解决的问题。从上文中可以看出，CCN/NDN 中主要采用的是分层结构化的内容标识方法。这种层次性标识方式虽然借用现有网络中的 URL 格式直接对其命名，大大降低了规范命名的工作量，但是仍然存在一些问题。具体问题如下：由于采用分层结构化的方式，在访问时仍与 DNS 一样呈树状的模式，这种中心化的结构和分布式管理模式有一定冲突；内容名字不是固定长度的，而且字符串长度也没有上限，导致查找时比 IP 地址复杂得多；由于数据量的不断增长，当发布或删除内容时将引起路由更新，快速实时路由更新是 CCN/NDN 路由查找所要解决的难题；每种内容资源都有一定的内容信息特征，分层结构化的标识方式没有考虑到不同类型内容的差异性；视频内容资源作为一种数据量较大、种类多样和信息冗余度高的内容资源，这种特殊性并没有被利用到标识实现过程中。综上，需要一种分布式的标识方法来对内容资源进行命名，从而改进内容分发的灵活性。因此，视频内容标识是内容网络中需要解决的问题。

3 视频内容标识

3.1 设计原则

与 CCN/NDN 的设计原则相似，在视频内容标识设计时，需要根据视频内容的特点来提出一套

基本的设计原则。具体的原则如下^[12]。

(1) 感知性

根据人类视觉系统对视频数据的认知，内容相同的视频资源所表达和传递的信息是一致的，即使视频像素有大有小、色彩有亮有暗、旋转角度各不一致等情况发生，由于它们所表达的内容是一样的，因此人类的感知信息也是一样的，从而最终生成的视频内容标识也是一样的。

(2) 简洁性

即产生的指纹数据量应尽量较少，这样可以利用视频内容的标识进行视频分类以及管理，减少处理时所产生的计算量。

(3) 唯一性

视频指纹应能唯一标识视频内容，即两个感知不同的视频，它们的指纹序列应该是不同的，感知相同的视频内容应该具有相同的视频指纹。

(4) 顽健性

视频指纹能够抵抗使得内容信息不变的攻击操作，如噪声、缩放、帧率变换和转码等，这种攻击操作，使得底层的字节数据发生一定的变换，但其表现的视觉内容仍是一样的。因此，需要抵抗这类攻击导致的视频指纹变化的情况，保证系统的顽健性。

(5) 高效性

生成指纹的算法复杂度要较低，降低视频指纹生成的成本，从而加快视频内容资源的分发速率。

(6) 便于匹配和查找

为后续的指纹匹配工作做准备，生成的视频指纹应具有较好的数据结构模式，这样便于使用资源时的匹配和查找工作。

3.2 主要方案

在视频内容标识的研究中，已有许多典型的方案，这些方案大体上可以分为两大类：一类是将视频看作一个整体的数据流，通过 hash 算法将数据流映射成相应的 hash 值，hash 值即

视频内容的标识；另一类是考虑了视频内容信息特征，通过一系列方法提取出视频的内容信息特征，基于内容信息特征得到相应的视频内容标识。下面将描述两种不同类型中的典型方案之一。

Walfish 等人^[13]阐述了虽然 DNS (domain name system) 是互联网上最成功的技术之一，但是由于 DNS 结构僵化，浏览器通过 URL (uniform resource locator) 来获取相关网页视频信息时，视频资源将会被特定的协议、主机号以及相关文件名所绑定，这样会导致资源不能够被灵活地分享和调度。因此，Walfish 等人^[3]以分布式散列表 (distributed hash table, DHT) 为基础，提出了一种无语义的内容标识技术 SFR (semantic free referencing)。这种方案用 DHT 将不同内容资源映射成一个 160 bit 的字符串，即 SFR 标签，并嵌入该内容标签记录 o-records (object records) 中，o-records 中也包含了该内容数据的地址信息以及其他元信息，如图 2 所示。该方案主要是利用 hash 表的特性，将内容数据通过 hash 算法生成相应的 hash 值，从而形成精简的内容标识。但该方案存在的缺点是：没有考虑视频内容数据的特征性，如同一内容不同格式的视频数据，在底层的字节信息可能会不大一样，这样由 hash 算法所产生的内容标识将会不同，从而同一视频内容将会拥有不同内容标识值，影响视频的分类与分发；由于视频数据量巨大，通过 hash 算法产生的 hash 值碰撞的概率也将提高。

SFR标签: 0xf01212099abcb679ac256ba4d...
地址: (ip, port),(DNS name, port),SFRtag
元数据信息: App-specific meta-data
生存时间: time-to-live: a caching hint

图2 内容标签记录 o-records

通过对上述方案的延伸，Anand 等人^[14]从内容信息的角度出发提出一种基于内容信息特征的视频内容标识方案 IBR (information-bound reference)。由于视频可以看成是一帧帧图像所组成的连续有序的集合，因此该方案受到图像特征提取的启发，将图像上的特征提取研究方案应用于视频标识上，为视频内容资源的标识提供了一种有效的方案。该方案的具体过程如图 3 所示。

该方案的具体过程如下。

步骤 1 通过视频镜头分割技术寻找出镜头切换点，该切换点即关键帧，以关键帧为边界点，将原视频分成一段段的视频片段。

步骤 2 在视频片段中选取该段的首帧和尾帧图像，利用图像特征提取技术生成其 IBR 值 (如图 3 中的 $IBR(F1)$ 和 $IBR(F3)$)，将首帧和尾帧图像的 IBR 值再加上该段音频信息的 hash 值合成一个 IBR 文件块作为该段的 IBR 标识。

步骤 3 将各个段的 IBR 文件块合在一起即形成原视频文件的最终 IBR 标识。其中图像的 IBR 值生成过程如图 4 所示。

图像 IBR 大致的生成思路就是利用离散余弦

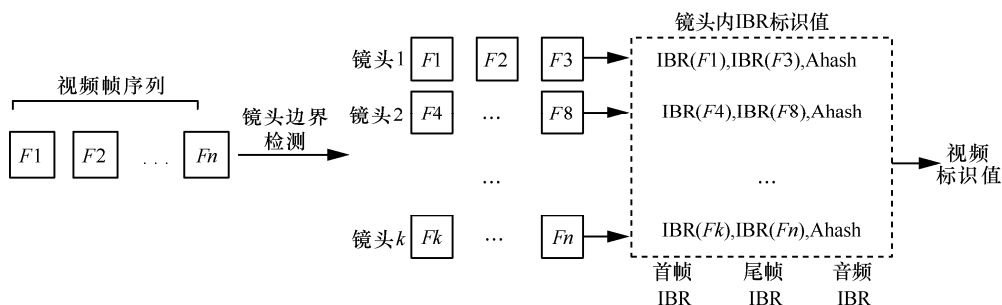


图3 基于关键帧 IBR 的视频标识生成过程

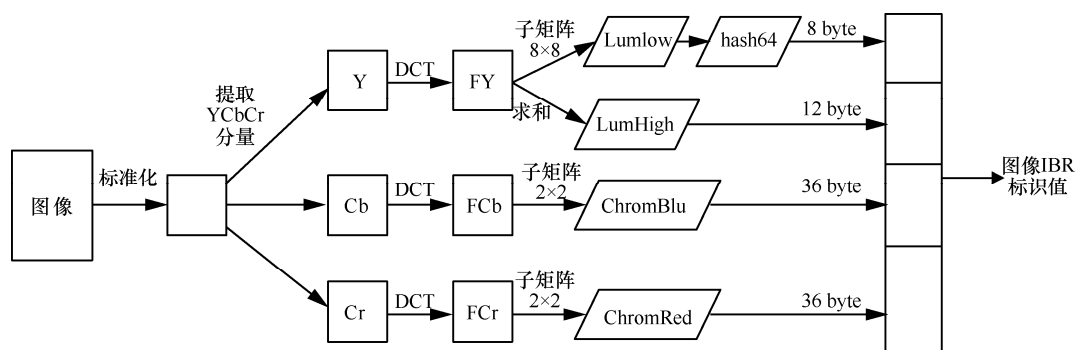


图4 关键帧图像 IBR 生成过程

变换 (discrete cosine transform, DCT) 提取出图像中 Y、Cb 和 Cr 成分的频率分量, 通过一系列整合得到最终的 IBR 值。因此, 可以看出 IBR 方案存在的问题是: 在提取特征时加入了色彩分量, 虽然提高了 IBR 值的有效性, 但也随之提高了运算成本, 使得生成时间变长; 由于在视频关键帧提取的过程中, 不同的视频镜头分割技术将会导致不同的结果, 而且也没对其所使用的技术进行详细阐述, 因此在该过程的处理上存在着一定的争议。

根据 IBR 的方案, 主要思路就是通过现有的图像或者视频处理方法提取视频的特征点来实现相应的内容标识。在图像和视频的研究领域中, 存在许多基于视频特征点的处理技术, 如视频指纹技术。在第 1 节中, 提到了对于同一内容视频的不同版本数据, 需要唯一的特征点为其标识, 生成同一个标识为视频的分类与管理提供方便。因此, 在提取视频特征点的过程中, 需要主要考虑的是视频的内容特征而不是结构特征^[15-16]。而视频指纹技术就考虑了这点, 不仅能解决 IBR 中存在的问题, 而且能为视频提供精简的标识值。因此, 本文将从视频指纹出发, 解析视频指纹的原理以及一种视频指纹的实现过程, 从而展现出视频指纹技术在内容网络中视频内容标识中的应用价值。

4 基于视频指纹技术的视频内容标识方案

视频指纹技术是通过提取视频内容特征信

息作为视频标识, 基于此标识建立视频指纹数据库, 然后通过相似性匹配算法进行查询、分类与管理^[17]。因此, 视频指纹技术主要由两个方面构成: 视频指纹的提取和视频指纹的匹配。下文描述了视频指纹技术的概念以及技术原理, 然后设计了一种基于时-频域视频指纹的视频内容标识方案。

4.1 视频指纹技术概述

早在 1999 年, 斯坦福大学的 Shivakumar N 和 Indyk P^[18]在博士论文中就如何在互联网中找出盗版视频数据为目标, 提出一套盗版视频数据检测技术。随着互联网的发展, 相关研究者希望用较少的数据成本来对视频元数据进行标识, 利用视频的标识进行基于内容的分类, 从而便于视频数据库的管理。由此, 以 Shivakumar N 和 Indyk P 的研究成果为基础, 不断发展形成了现在的视频指纹技术, 并且以关键技术写入了 MPEG-21 标准中。视频指纹是一个视频对象感知特征的简短摘要, 因此它有时又被称为视频摘要、视频散列、视频身份。它是通过一定的科学算法来提取视频内容本身所具有的某些特征信息 (如亮度、颜色、频谱等), 对这些特征信息进行统计、组合等方法后形成唯一的指纹序列, 通过与数据库中指纹间的相似性匹配计算, 能将视频本身与其他不同视频区分开来的数据。视频指纹提取主要分为视频预处理、特征提取、指纹建模和数据库构建^[19]4 个步骤。在这 4 个步骤中, 细节点较多但在大多数研

究中并没有详细解释的是视频预处理过程。因此,接下来主要对此过程进行阐述。

在视频预处理过程中,首先进行的就是视频解码处理,将数字视频信号流转换成一帧帧视频图像,然后再对帧图像进行后续的处理。视频解码的研究已经十分成熟,所以这里将不再重述。这里重点讲述下后续的处理过程,为了更好地解释该过程,先将视频组织结构进行分析。一段连续的视频,可划分为以下结构:视频镜头、视频关键帧、视频组以及视频场景^[20],如图 5 所示。可以看出,视频数据有很大的冗余量存在。因此,视频预处理一方面是为了消除这些冗余量,保留关键信息;另一方面是为后面的特征提取做格式上的准备,从而能生成标准化指纹格式。图 5 中,可见一个常用的预处理方法:视频镜头边缘检测。这种方法的原理是通过统计计算(帧差法或光流法等)来寻找镜头突变或者渐变的时间点,然后将整段视频分成一个个视频镜头片段,最后在一个镜头中选取或合成一个关键帧对该镜头进行表示。

关于特征的提取部分,基于不同的提取方法的特性信息,其相似性计算方法也不尽相同。目前已提出的视频指纹提取技术可分为三大类:基于时域、基于空域以及基于变换频域的视频指纹^[21]。指纹建模,是将提取的特征进行分析,用适合的数学工具建立起模型,以便进行指纹匹配。

视频指纹匹配,即在上述基于视频指纹特征构

建的数据库中,查询出所需视频的指纹数据,并获取该视频数据资源。该问题属于典型的查找型问题,最简单的算法即穷举法,但是由于现在视频数据量繁多,因此运用穷举法来解决该问题并不可取。虽然对于视频指纹的匹配算法研究甚少,但是对于海量文本数据或者其他类型数据的快速查找算法研究较多。因此,对于视频指纹的匹配算法的研究,一般把视频指纹数据看成比特串或者其他数据格式,根据已有的快速查找算法来解决问题。

最常见的快速最近邻查找算法是局部敏感散列(locality sensitive hashing, LSH)算法,该算法最早由 Indyk 等人^[22]提出,后续 Gionis 又进行了总结。Hu^[23]根据 LSH 算法实现了一个高效的视频检索系统,并证明了 LSH 算法在视频指纹应用的实效性。LSH 算法初衷是为了解决海量高维度数据的快速最近邻查找问题,其主要原理是:将高维度数据通过特殊的 hash 函数降维到低维数据,同时在一定程度上,保证原始数据的相似性不变。这样高维数据映射到低维不同的桶位中,在桶内进行线性查找,从而加速了查找的效率。

对这个特殊的 hash 函数有以下两个要求^[23]:

- 如果 $d(x,y) \leq d_1$, 则 $h(x) = h(y)$ 的概率至少为 p_1 ;
- 如果 $d(x,y) \geq d_2$, 则 $h(x) = h(y)$ 的概率至多为 p_2 。

其中 $d(x,y)$ 表示 x 和 y 之间的距离, $d_1 < d_2$,

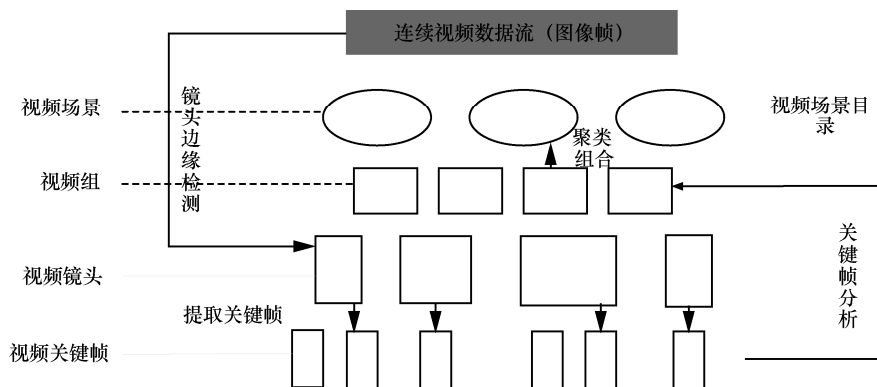


图 5 视频结构特征



$h(x)$ 和 $h(y)$ 分别表示对 x 和 y 进行 hash 变换。

虽然 LSH 提高了查找的速率,但是也降低了查找的准确率。为了提高查找的准确率,部分研究者进行了研究,提出了一系列算法。最具代表的是以聚类为基础的匹配算法。

4.2 视频内容标识方案设计与实现

根据前文视频指纹技术原理介绍,本节对时域和频域的视频指纹技术进行了融合,并提出了一种基于时-频域视频指纹的视频内容标识方案。该方案的基本原理是:先在视频的时间域中进行处理,利用视频镜头分割和图像合成技术得到一段视频的关键帧,然后对于每一关键帧进行 DCT 得到频域图像,从频域图像中提取出频域信息经过统计处理得到相应的视频指纹,对视频指纹进行整合后得到最终的视频标识。因此该方案分为 3 个过程:时域处理过程,又称视频预处理过程;频域处理过程;标识生成过程。具体过程如图 6 所示。

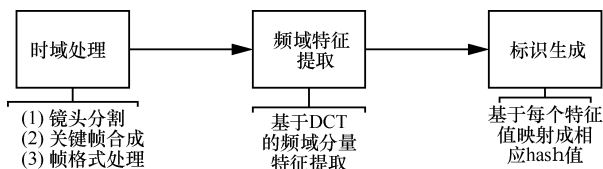


图 6 基于时-频域视频指纹的标识生成总过程

针对时域处理过程,下面将对每一个步骤进行详尽说明,具体过程如图 7 所示。

步骤 1 视频解码。目前存在大量的编解码

标准,为了适配不同的应用环境和终端设备,视频经常需要被转码成相应能处理的数据格式。造成同一内容的视频会以多种不同的编码形式存在。不同编码的数据具有不同的字节特征,这样想要直接从编码数据中提取唯一特性的指纹将变得很困难。因此,数字视频首先需要被解码,得到压缩前的视频帧序列。

步骤 2 帧序列分组和帧合成。这部分是大多数参考文献中提到的视频预处理过程,其目的是:剔除视频数据中的冗余量;为后续的处理做准备而进行帧图像标准化。本文的方案中,将步骤 1 得到的视频帧序列进行均等分组,先均等分成 63 组帧序列,最后的剩余作为第 64 组。这样得到 64 组帧序列后,需要利用图像合成技术,将组内帧序列合成一帧图像作为关键帧为后续处理做准备。这里所用的图像合成技术是 TIRI (temporally informative representative image) [24] 技术。这种方法的原理是:将一组图像进行权重相加,得到一个新的图像作为表示图像,计算式如下:

$$L_{m,n} = \sum_{k=1}^J w_k l_{m,n,k} \quad (1)$$

其中, $l_{m,n,k}$ 是一组总数为 J 帧序列中第 k 帧在 (m,n) 处的像素值, w_k 是第 k 帧的权重值。对权重值的设定,一部分参考文献认为将权重值设为指数值 (即 $w_k = \gamma^k$) 能很好地对一组帧序列进行表示,

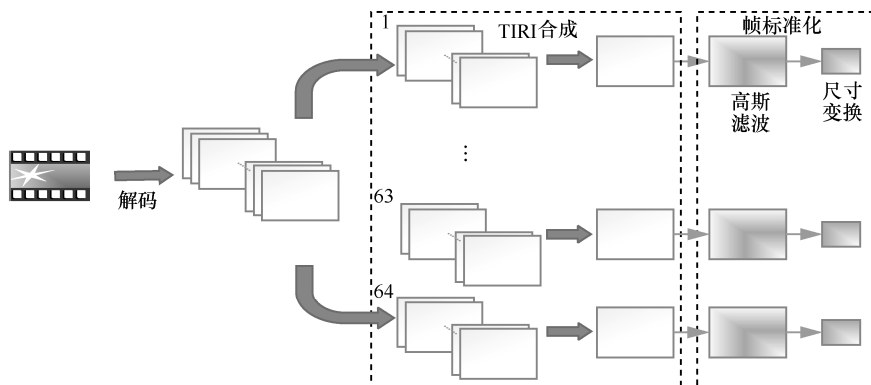


图 7 视频标识方案时域处理过程

但本文认为当组内帧数量足够多时,越往后帧图像的权重将越大,从而可能忽略前面的帧图像,导致信息丢失,没有对视频内容信息进行充分提取。因此,本文设置的权重值为 $w_k=1$,即对组内每帧图像进行均等对待,提取组内帧图像的像素平均值,保证最终提取的信息量的完整性。

步骤 3 关键帧的标准化。对关键帧标准化是为了最终生成标准形式的视频指纹格式。在这个标准化过程中包含两个方面:图像滤波和图像尺寸缩减。图像滤波,即在尽量保留图像细节特征的前提下对目标图像的噪声进行抑制,是图像预处理中不可缺少的操作,其处理效果的好坏将直接影响到后续图像处理和析的有效性和可靠性。本文所用的滤波器是高斯滤波器,是一种线性平滑滤波。因为其能在不破坏图像原始信息的情况下抑制噪声,所以广泛应用于图像处理的降噪过程。将关键帧图像降噪后,再进行图像的尺寸变化,生成标准化的图像大小,本文选择的是 $144 \text{ dpi} \times 176 \text{ dpi}$,这是 H.323 协议簇中规定的最小能表征图像信息的图像格式 (QCIF)。

通过上述时域处理,任何视频都将得到一个 $(144,176,64)$ 的关键帧集合,其中 $144 \text{ dpi} \times 176 \text{ dpi}$ 是关键帧图像尺寸大小,64 是关键帧的总数。在这个关键帧集合的基础上,这里将对其进行频域处理,提取出关键帧的频域信息。在进行频域处理过程中,有 3 个问题需要考虑:用何种变换工具来进行频域转换;提取多少频域分量信息;如何对提取出的分量信息进行量化。针对这 3 个问题,下文将一一阐述解决办法。

目前,在图像和视频的频域处理的研究中,应用最广泛的是 DCT^[25-26]。DCT 是对实信号定义的一种变换,它是从离散傅里叶变换 (discrete fourier transform, DFT) 推导出来的一种变换。相比 DFT 而言,DCT 可以减少一半以上的计算。因为 DCT 具有能量集中特性,即声音或图像信

号通过离散余弦变换后大多数信号能量集中在其低频区域,所以 DCT 在声音和图像数据处理中得到了广泛的使用。因此,本文也是利用 DCT 对关键帧集合进行频域转换。在 DCT 过后的频域矩阵中,需要选取合适大小的子矩阵作为该关键帧标识信息。为了确保最终视频指纹的简洁性和高效性,选取的子矩阵大小不能过大或过小,过大将导致视频指纹信息量过多,从而影响简洁性,过小会导致提取的信息量较少,从而使得部分重要信息丢失,影响其高效性。在本文中所选取的子矩阵大小是 16 行 $\times$ 16 列,每一关键帧将由 256 位频域分量标识。为了得到最终的视频指纹,最后就是将每关键帧的 256 位分量进行量化处理得到相应的比特串。与其他研究中的量化过程不同,本文将每一频率分量通道看成一组序列,在每一通道中计算出每组序列的中位数,根据以下计算式进行量化:

$$h_i = \begin{cases} 1, & C_i \geq m \\ 0, & C_i < m \end{cases} \quad (2)$$

其中, m 为通道中的中位值, C_i 为第 i 处的频率分量, h_i 为量化后的值。这样每一个视频数据将得到一个 64 行 $\times$ 256 位的比特指纹序列,图 8 形象地描述了上述过程。

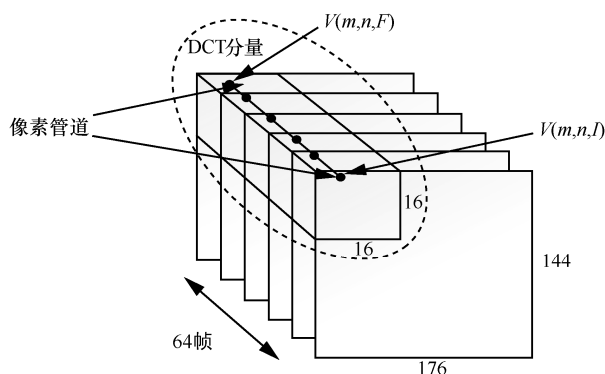


图 8 视频标识方案频域特征提取点示意

最后,通过 hash 函数对得到的视频指纹数据进行处理,生成最终的视频标识值。这种视频标



识不仅满足视频内容标识设计原则,也解决了内容网络中视频内容标识的问题。

5 结束语

内容标识一直都是内容网络中需要解决的问题,其中对于视频这一类型内容来说,它的内容标识与其他内容相比更为复杂。视频指纹技术是一种新型的视频内容标识技术。它通过一定的科学算法来提取视频内容本身所具有的某些特征信息(如亮度、颜色、频谱等),对这些特征量进行统计、组合等方法处理后形成唯一的指纹序列,通过与数据库中指纹间的相似性匹配计算,能将视频本身与其他不同视频区分开来的数据。本文研究了将视频指纹技术应用于解决内容网络中视频内容资源的标识问题,提出了一种基于时-频域视频指纹的标识方案,有效解决了视频内容在内容网络中的分类与管理问题。但随着科技的发展,现有的视频标识方案将会在日益丰富的视频数据中显现不足。因此,在后续的工作中,还需要进一步分析与研究。

参考文献:

- [1] 尹浩,袁小群,林闯,等.内容网络服务节点部署理论综述[J].计算机学报,2010,33(9):1611-1620.
YIN H, YUAN X Q, LIN C, et al. The survey of service nodes placement theories for content networks[J]. Chinese Journal of Computers, 2010, 33(9): 1611-1620.
- [2] 唐霏.内容网络中的缓存设计[D].成都:电子科技大学,2013.
TANG F. Caching design in content network[D]. Chengdu: University of Electronic Science and Technology, 2013.
- [3] 胡骞.以内容为中心的网络中缓存技术的若干问题研究[D].北京:北京邮电大学,2015.
HU Q. Research on caching technology in content centric network[D]. Beijing: Beijing University of Posts and Telecommunications, 2015.
- [4] 崔现东.内容中心网络网内缓存策略研究[D].北京:北京邮电大学,2014.
CUI X D. Research on caching strategy in content center network[D]. Beijing: Beijing University of Posts and Telecommunications, 2014.
- [5] 黄韬,刘江,霍如,等.未来网络体系架构研究综述[J].通信学报,2017,35(8):184-197.
HUANG T, LIU J, HUO R, et al. Survey of research on future network architectures[J]. Journal on Communications, 2017, 35(8): 184-197.
- [6] 刘斌,汪漪.内容中心网络中名字查找技术的研究[J].电信科学,2017,30(9):10-17.
LIU B, WANG Y. Research on name lookup in named data networking[J]. Telecommunications Science, 2017, 30(9): 10-17.
- [7] 张行功,牛童,郭宗明.未来网络之内容中心网络的挑战和应用[J].电信科学,2017,29(8):24-31.
ZHANG X G, NIU T, GUO Z M. Challenge and implementation of content-centric networking[J]. Telecommunications Science, 2017, 29(8): 24-31.
- [8] JACOBSON V, SMETTERS D K, THORNTON J D, et al. Networking named content[C]//The 5th International Conference on Emerging Networking Experiments and Technologies, December 1-4, 2009, Rome, Italy. New York: ACM Press, 2009: 1-12.
- [9] ZHANG G, LI Y, LIN T. Caching in information centric networking: a survey[J]. Computer Networks, 2013, 57(16): 3128-3141.
- [10] 胡骞,武穆清,郭嵩.以内容为中心的未来通信网络研究综述[J].电信科学,2012,28(9):74-80.
HU Q, WU M Q, GUO S. A survey of content-oriented future communication network[J]. Telecommunications Science, 2012, 28(9): 74-80.
- [11] 雷凯.信息中心网络与命名数据网络[M].北京:北京大学出版社,2015.
LEI K. Information-centric networking (ICN) and named-data networking (NDN)[M]. Beijing: Peking University Press, 2015.
- [12] 王大永.感知视频指纹算法研究[D].上海:上海交通大学,2012.
WANG D Y. Research on perceptual video fingerprint algorithm[D]. Shanghai: Shanghai Jiao Tong University, 2012.
- [13] WALFISH M, BALAKRISHNAN H, SHENKER S. Untangling the Web from DNS[C]//NSDI 2004, March 29-31, 2004, San Francisco, CA, USA. New York: ACM Press, 2004: 17.
- [14] ANAND A, BALACHANDRAN A, AKELLA A, et al. Enhancing video accessibility and availability using information-bound references[J]. IEEE/ACM Transactions on Networking (TON), 2016, 24(2): 1223-1236.
- [15] KIM M J, YOO C, KO Y W. Multimedia file forensics system exploiting file similarity search[J]. Multimedia Tools and Applications, 2017: 1-22.
- [16] NIE X, YIN Y, SUN J, et al. Comprehensive feature-based

- robust video fingerprinting using tensor model[J]. IEEE Transactions on Multimedia, 2017, 19(4): 785-796.
- [17] 许涛. 面向视频管理的指纹特征提取技术研究[D]. 成都: 电子科技大学, 2015.
- XU T. Research on fingerprint feature extraction technology for video management[D]. Chengdu: University of Electronic Science and Technology, 2015.
- [18] INDYK P, IYENGAR G, SHIVAKUMAR N. Finding pirated video sequences on the internet[R]. 1999.
- [19] OUALI C, DUMOUCHEL P, GUPTA V. Robust video fingerprints using positions of salient regions[C]//2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 5-9, 2017, New Orleans, USA. Piscataway: IEEE Press, 2017: 3041-3045.
- [20] 庄捷. 流媒体原理与应用[M]. 北京: 中国广播电视出版社, 2012.
- ZHUANG J. Principle and application of streaming media[M]. Beijing: China Radio and Television Publishing House, 2012.
- [21] LU J. Video fingerprinting for copy identification: from research to industry applications[C]// Media Forensics and Security. International Society for Optics and Photonics, January 19, 2009, San Jose, CA, USA. [S.l.:s.n.], 2009: 725402.
- [22] INDYK P, MOTWANI R. Approximate nearest neighbors: towards removing the curse of dimensionality[C]//The 30th Annual ACM Symposium on Theory of Computing, May 24-26, 1998, New York, USA. New York: ACM Press, 1998: 604-613.
- [23] HU S. Efficient video retrieval by locality sensitive hashing[C]// IEEE 2005 International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05), March 18-23, 2005, Philadelphia, Pennsylvania, USA. Piscataway: IEEE Press, 2005: 449-452.
- [24] 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.
- [25] OOSTVEEN J, KALKER T, HAITSMAN J. Feature extraction and a database strategy for video fingerprinting[M]. Berlin: Springer, 2002: 117-128.
- [26] COSKUN B, SANKUR B. Robust video hash extraction[C]// 2004 12th European Signal Processing Conference, Sept 6, 2004, Vienna, Austria. Piscataway: IEEE Press, 2004: 2295-2298.

[作者简介]



陈勋 (1994-), 男, 北京科技大学计算机与通信工程学院硕士生, 主要研究方向为内容中心网络、视频分发网络。

韩帅 (1988-), 男, 北京科技大学计算机与通信工程学院硕士生、中国洛阳电子装备试验中心助理工程师, 主要研究方向为视频编码与传输、视频分发网络。

刘姗姗 (1994-), 女, 北京科技大学计算机与通信工程学院硕士生, 主要研究方向为流媒体技术、视频分发网络。

阳小龙 (1970-), 男, 北京科技大学计算机与通信工程学院教授、博士生导师, 主要研究方向为网络安全、内容中心网络、视频分发网络和新一代互联网理论与技术。