

Video Deduplication Based on Perceptual Hash

Xueqing Hu

Hubei Intelligent Wireless Communication Laborator, College of Electrical and Information Engineering, South-Central University for Nationalities, Wuhan Hubei
Email: 554653201@qq.com

Received: Apr. 11th, 2018; accepted: Apr. 23rd, 2018; published: Apr. 30th, 2018

Abstract

With the continuous development of human information technology, the total amount of data accumulated in all walks of life is becoming larger and larger in the whole Internet. There is a lot of redundancy in the multimedia data at the cloud end. At the same time, a lot of new duplicates are uploaded through massive clients, and a lot of bandwidth is wasted. Therefore, how to effectively re-emphasize it has become an urgent problem. This paper focuses on video deduplication in multimedia data deduplication. An adaptive threshold key frame extraction method based on perceptual Hash is proposed. Based on this, a video de duplication scheme is proposed, which includes key frame extraction, matching sorting and video quality comparison. Experiments confirm that the scheme has high accuracy.

Keywords

Perceptual Hash, Video Deduplication, Key Frame Extraction, Matching Sorting, Quality Evaluation

基于感知哈希的视频去重

胡雪晴

中南民族大学电子信息工程学院, 湖北省智能无线通信实验室, 湖北 武汉
Email: 554653201@qq.com

收稿日期: 2018年4月11日; 录用日期: 2018年4月23日; 发布日期: 2018年4月30日

摘要

随着人类信息技术的不断发展, 整个互联网中, 各行各业累积的数据量越来越大, 在云端的多媒体数据

中存在着大量的冗余,同时又不断的通过海量客户端,上传新的重复数据,浪费了大量的带宽。因此如何高效去重成为亟待解决的问题。本文主要研究多媒体数据去重领域中的视频去重。提出了一种基于感知哈希的自适应阈值关键帧提取方法。并在此基础上提出了一个视频去重的方案,该方案包括关键帧提取,匹配排序和视频质量比较。并通过实验证实该方案具有较高的准确率。

关键词

感知哈希, 视频去重, 关键帧提取, 匹配排序, 质量评价

Copyright © 2018 by author and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

目前媒体和网络技术的发展,呈现多平台,多渠道,多用户的发展趋势,各种云平台,门户网站,手机移动端直播平台每时每刻都在吞吐数据。互联网在给人们传输数据带来了巨大的便利,使得数据更容易传输,压缩,存储,甚至很简单的进行复制和相关处理。但同时也给信息检索带来了巨大的挑战。在各大平台中,存在着大量的冗余的多媒体数据。浪费了存储带宽,降低了用户体验。特别是视频数据,由于其传播信息量大,所占数据量大,因此高效的进行重复视频识别并删除可以减少网络带宽,可以使得网络环境更加简洁,增强用户体验。重复视频是指由原始视频通过一系列的内容保持操作,如滤波,剪裁,增加字幕等手段得到的视频副本。在互联网中,大多数的重复视频主要有三种,一是简单的格式转换或者分辨率转换,二是通过添加字幕和徽标的转换,三是视频的剪裁[1] [2]。

视频不同于文本和图像数据,占有较大的存储空间,且结构复杂。不可以通过简单的二进制流来检测重复视频。目前现有的视频去重方式有以下几种:通过检测视频的描述性文本信息的相似性对视频进行去重,描述性文本信息包括视频的标题,视频传播时的描述标签等,但是通过这种方式的错判率较高,视频的标题很容易进行人为修改[3]。有根据视频的内嵌字幕进行去重的方法,通过视频时长,对其进行分段,并提取相应时长内的文本信息,通过对比文本信息的相似程度来判断视频的相似性,但对于内嵌字幕文本信息较少的视频,缺少准确性[4]。相对有效方法的则是基于视频内容进行去重,视频是由一帧帧图像组成的,基于视频内容去重的思想是把视频检测转化为相应特征帧的图像相似性检测问题,主要过程是先进行特征帧选择,特征帧图像特征提取,特征匹配,其最关键的过程为特征图像特征的选取。在现有的特征提取方式中,基于颜色直方图的特征提取算法,不具有时空性,无法描述其特征在连续空间上的分布[5];而基于视频运动矢量的算法,不仅运动矢量的计算复杂,而且当视频发生改变时,由于快效应等的存在,很难捕捉到视频的边缘信息;更有一些基于 SIFT 的尺度不变算法,但其运算量过大,较为复杂[6]。

因此针对海量重复视频数据,如何准确高效的进行重复检测,关键是要找到视频中感知不变的特征量。在这里,我们进行了大胆的尝试,利用感知哈希算法提取视频特征帧的哈希序列,进一步进行相似性匹配,实现视频去重过程。

本文提出了一种基于感知哈希的视频关键帧提取算法,并提出了一种基于感知哈希的视频去重方案,该方案利用所提取关键帧的感知哈希特征构建映射,通过匹配排序进行重复视频识别,主要包括关键帧提取,匹配排序和视频质量等过程。

2. 感知哈希算法

感知哈希提取多媒体的主要感知信息，并将其量化压缩转化为一些列的二值序列做为指纹摘要，并用该指纹摘要唯一标识多媒体文件。在实际应用过程中如海量图像检索与图像拷贝检测中，通过图像指纹摘要的存储，极大的减少的数据的存储量以及传输带宽，降低了存储成本，极大的方便了数据的管理与维护，同时极大提高了工作效率。对于感知哈希的定义，参考文献[7] [8] [9]中提到的概念，归纳来说，感知哈希是多媒体数据集到感知摘要集的一种单向映射，即将具有相同感知内容的多媒体数据唯一地映射为一段指纹摘要，并满足感知鲁棒性和安全性要求。

本文所用的感知哈希算法通过把图像进行分块 DCT 变换，提取变换后的重要系数，并对系数进一步处理生成图像的感知哈希序列 BDCT-hash，用来代表每一幅图像。并用两幅图像 BDCT-hash 的汉明距离判断任意两幅图像间的距离。

2.1. 生成感知哈希序列

生成感知哈希序列包括两个步骤，首先进行图像预处理，然后提取图像特征向量。

第一步：图像预处理

预处理包括：

1) 彩色图像灰度化，本文仅考虑灰度情况下的图片，降低了亮度对图像分析的影响，彩色图像首先通过 RGB 转化为灰度图像。使用下列公式将图像转换为灰度形式：

$$Gray = R \times 0.299 + G \times 0.587 + B \times 0.114$$

2) 统一图像大小，对图像进行双线性插值，使分辨率统一为 64×64 ，以保证生成的感知哈希序列长度一致。

第二步：提取图像特征向量

1) 将预处理后的图像分成 64 个 8×8 的块，通过分块 DCT 变换，得到每个分块的 DCT 系数矩阵。

2) 分别取每个块的 DC 分量以及按照 zigbig 扫描的 AC1 和 AC5 分量，将相同位置的分量按照从左至右从上至下的顺序连接起来组成三个长度为 64 的一维向量 A_i 。

3) 将三组向量分别进行二值化处理得到 F_i 。

$$F(i, j) = \begin{cases} 0, & Gray(i, j) < Avg \\ 1, & Gray(i, j) \geq Avg \end{cases}$$

4) 依次连接 F_i ，生成长度为 192 的一维特征向量 F 。

特征向量 F 即为图像经过 BDCT-hash 得到的感知哈希序列值。

2.2. 相似性度量

我们采用汉明距离作为比较两个感知哈希序列值之间相似性的一种度量。

设 $x = \{x_1, x_2, \dots, x_N\}$ ， $y = \{y_1, y_2, \dots, y_N\}$ 表示两个有限长字符串。 x, y 之间的汉明距离 $D(x, y)$ 被定义为：

$$D(x, y) = \sum_{i=1}^N Q(x_i, y_i)$$

这里 $Q(x_i, y_i)$ 是一个函数，当 x 和 y 不同时值为 1，相同时等于 0。为了便于比较，

汉明距离可以相对于字符串的长度 n 进行归一化。归一化汉明距离定义为：

$$D(x, y) = \frac{\sum_{i=1}^N Q(x_i, y_i)}{N}$$

特别的，二进制编码的数字之间的汉明距离可以使用 XOR 运算来计算。

$$\sum_{i=1}^N (x_i \oplus y_i)$$

两个字符串之间的汉明距离可以被用来确定他们的相似性。如果汉明距离越大，字符串就越不一样。为了识别类似的字符串，我们通常为汉明距离设置一个阈值 T 。

$D(x, y) \leq T$ 时， xy 被识别为相似的字符串，此时认为两个感知哈希序列字符串代表的图像为相似图像。

$D(x, y) > T$ 时， xy 被识别为不同的字符串，此时认为两个感知哈希序列字符串代表的图像为不同图像。

3. 视频的结构和特点

视频是由一张张连续的静态图片构成的，利用人类视觉暂留效果，连续变化的静态图像构成动态视频。如果说图片是二维的，图片数据记录了像素点的信息以及相关位置信息，视频是三维的，增加了时间信息，使得视频更加复杂。因此分析时不能单独来看每一张静态的画面，时序上的关系也是视频的特征之一[10]。一般来说按照视频序列的结构层次可以如下划分：视频序列有单个或者多个场景构成，场景又划分为不同的镜头，镜头由不同的视频帧组成。如图 1 所示为视频的基本组成结构。

帧是视频序列的最小组成部分，最基本的单元，一个静态图像即为一帧，视频是有连续帧构成的图像序列。帧中包含了全部的视频信息。

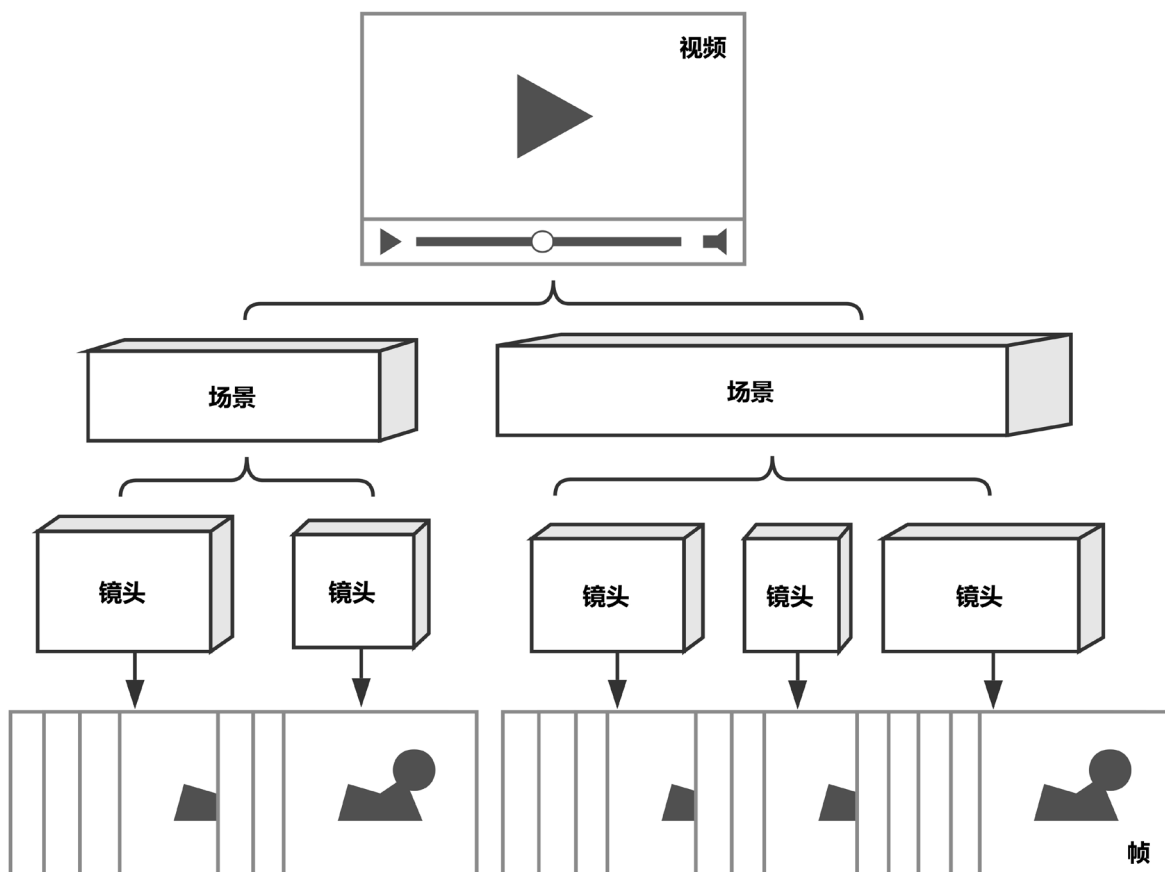


Figure 1. The basic structure of video

图 1. 视频的基本组成结构

镜头有一系列连续帧构成，他通常描述的是同一场景下，一个事物主体连续的动作。在同一个镜头内，相邻帧之间的差异不大，且往往描述的是同一个事物主体在镜头内的变化，事物的变化包括平移，缩放等连续的动作。

场景有多个镜头构成，同一场景内的背景往往不会有太大变化，从不同角度描述事物主体的变化，不同场景间的背景不同，这是用来划分场景的常用手段。

视频序列又由一个或者多个场景构成，描述事件的开始到结束，整个视频文件是一个视频序列。

4. 基于感知哈希的视频去重

基于感知哈希的视频去重方案流程图如图 2 所示。该方案主要由三部分构成：特征帧提取，匹配排序，视频质量评价。视频是由一帧帧图像构成的，同一镜头内连续帧的内容相似性极强，而不同镜头帧间差别很大。在一定程度上可以认为同一镜头内的连续帧为同源图像。我们可以借鉴感知哈希对同源图像检索的方法，利用对每一特征帧提取的感知哈希序列识别同一镜头内的图像帧，当相邻帧的感知特征序列距离小于设定的阈值时，将其归为同一镜头，而当相邻帧间的感知特征序列大于设定的阈值时，此前帧为临界帧。将视频划分为不同的镜头后，再选取特定数量的特征帧，组成特征帧序列。得到特征帧序列后，每一个特征帧图像进行感知哈希计算生成特征帧感知哈希序列。所有的特征序列存放在特征帧数据数据库中，当进行视频去重时，在特征值数据库中对检索视频的特征帧序列中的每一个特征帧进行检索，记录检索结果得到特征帧图像候选集，最后将特征帧序列的检索结果整合得到重复视频序列的候选集。得到候选集后再通过质量评价阶段，对候选集内的重复视频进行评价，得到质量最好的视频序列，进行标记。同时可选择删除其余重复视频。下面将详细介绍该方案的主要过程。

4.1. 基于感知哈希的自适应阈值关键帧提取方法

视频是由多个不同的镜头构成的，同一个镜头内相邻帧间的差异不大，具有一定程度上的相似性。而不同镜头间的相邻帧则差异较大，图像内容会发生突变。基于镜头分割的关键帧提取算法的思路正是

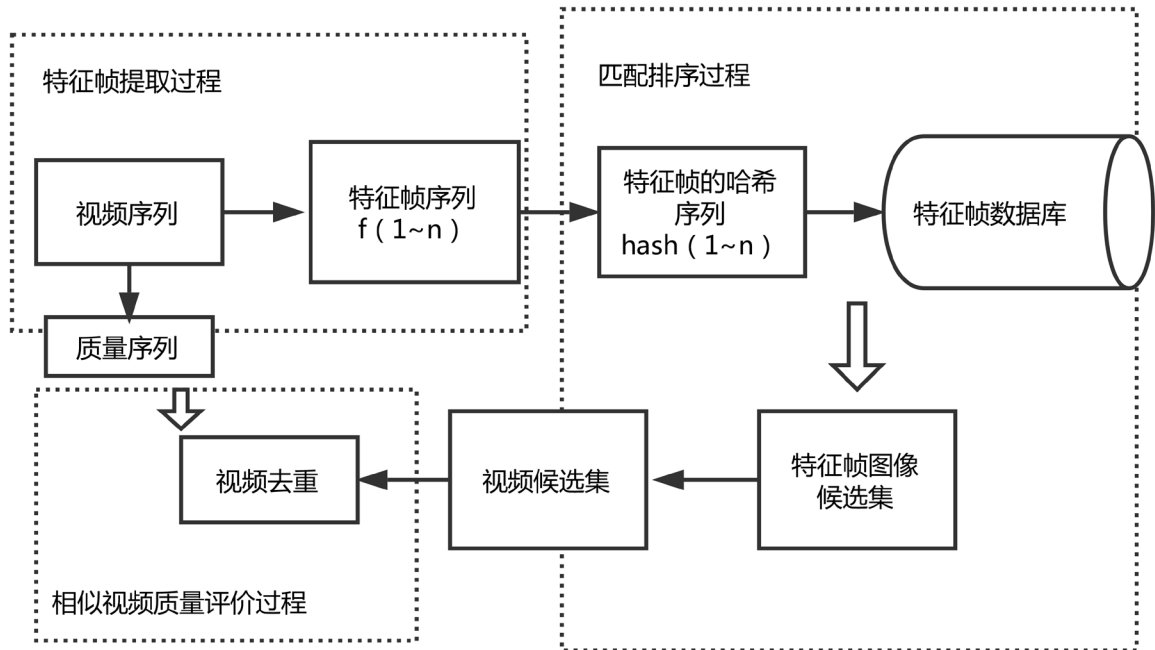


Figure 2. Block diagram of video deduplication scheme based on perceptual Hash
图 2. 基于感知哈希的视频去重方案框图

利用这一点, 首先计算每一帧图像的感知哈希序列值, 然后计算视频相邻帧间的汉明距离差异值, 当该差异值超过所设定的阈值时, 便判断该相邻帧为镜头的分界帧, 找到分界帧后便可以将视频序列划分成多个镜头, 然后通过划分的不同镜头, 按照一定的数量关系提取关键帧。但基于镜头分割的算法只能针对特定的场景, 因为当场景发生变化时, 不同镜头间的差异值发生变化, 此时需要按照不同的场景调整算法中的阈值至合适的大小, 过程麻烦。不同视频场景的镜头分界值会随着场景的不同发生变化, 远景轮廓背景场景内不同镜头间差异较小, 而近景细节背景场景内不同镜头间的差异会较大。根据这个原理, 本文提出了一种基于感知哈希的镜头自适应阈值分割算法。

算法具体步骤:

视频帧感知哈希特征提取。视频是由图像帧构成的, 提取视频图像帧, 分别对提取出的视频帧进行感知哈希计算, 利用上述的 BDCT-hash 算法, 计算每一帧图像的感知哈希序列。视频每一帧 t 的感知哈希特征用 F_t 表示:

$$F_t = \text{BDCT-hash}(I)$$

相邻帧差值计算。相邻帧间差异值用汉明距离计算, 公式为:

$$\text{diff} = \left\| \sum_{t=1}^N Q(F_t, F_{t-1}) \right\|$$

镜头边界帧判断。

设定滑动窗口为 Seq , 设定两个阈值 m, n (且 $n > m$), $\overline{\text{diff}_{seq}}$ 表示为滑动窗口内所有帧差的平均值, diff_{curr} 表示当前帧与上一帧的差异值。默认把第一帧加入滑动窗口, 每计算一帧的差异值后进行如下操作:

若 $\text{diff}_{curr} \geq m \times \overline{\text{diff}_{seq}}$ 且 $\text{diff}_{curr} \leq n \times \overline{\text{diff}_{seq}}$, 则将当前帧加入到滑动窗口中。

若 $\text{diff}_{curr} \geq n \times \overline{\text{diff}_{seq}}$, 则当前帧即为镜头的分界帧。

找到第一个分界帧后清空滑动窗口, 并将接下来的一帧做为初始帧放入滑动窗口中, 重复上述步骤, 直到视频最后一帧。

镜头合并。在上一步检测出所有关键帧后, 根据关键帧的时间定位, 将时间相差较小, 小于 1 s 的关键帧的相应镜头合并。

5) 镜头关键帧选取。每个镜头中找到与该镜头的直方图特征平均值最接近的一帧, 将该帧做为该镜头的关键帧。

6) 关键帧排序。根据每个镜头不同的时长, 赋予镜头关键帧不同的权重进行归一化处理并把关键帧按照权重排序。需要时选取前 n 位关键帧作为视频关键帧。

对于视频任意视频, 该过程可以表示为

$$\text{video}_{query} \rightarrow \{f_1, f_2, \dots, f_n\}$$

4.2. 匹配排序

匹配排序过程分为两大部分。一是视频到图像特征的映射。在关键帧提取过程中, 得到了视频到图像的映射, 对于任一个视频得到了其相应的关键帧序列。匹配排序, 首先要对关键帧序列建立索引。建立视频到图像的联系。关键帧序列实际上是有一系列图像构成的集合, 对于这些图像数据, 从前几章的内容中可知, 我们很容易获得每一张图像的感知哈希序列, 将该哈希序列做为关键帧序列中图像的特征值。提取所有视频的全部的关键帧特征序列建立关键帧的特征值索引库, 用进行后续的匹配。

特征值索引库中的每一个关键帧的序列包括以下信息: $f_i \rightarrow (\text{hash}, \text{Video}, \text{weight}_{mi})$ 。

包括哈希序列值，所属的视频，以及该特征帧所在的镜头在视频序列中所占的权重。

二是图像到视频的映射。从上一步中，我们可以得到关键帧序列中全部图像的感知哈希序列，存放在特征值索引库中。这样相似视频检索就转化成了相似图像检索的过程。我们在特征值数据库中对待检测视频的关键帧图像一一进行检索，与上文所述的相似图像检索过程相同，通过对 hash 值汉明距离的比较，得到关键帧序列中每一张图像在特征值数据库中的相似图像。同样，得到的匹配后的索引图像也包含以上三种信息，哈希序列值，所属的视频，以及在视频中所占的权重。

该算法的具体步骤如下：

1) 关键帧提取。按照 4.1 中的算法提取待查询视频的关键帧。其中 f_n 代表第 n 个图像关键帧。

$$video_{query} \rightarrow \{f_1, f_2, \dots, f_n\}$$

2) 关键帧的特征提取

按照 2.1 中所提到的 BDCT-hash 算法提取每一个关键帧的感知特征序列。其中第 n 个图像的感知哈希特征序列用 $hash_n$ 表示。

$$video_{query} \rightarrow \{f_1, f_2, \dots, f_n\} \rightarrow \{hash_1, hash_2, \dots, hash_n\}$$

3) 图像检索

在特征值索引库用待检测视频的关键帧特征序列搜索其相似图像，每个关键帧可以得到 k 个相似图像候选集， $score$ 表示图像的相似度评分， $score$ 的计算有汉明距离的不同位数占总序列长度的比例得到。

$$f_i \rightarrow \{(f_1, score_1), (f_2, score_2), \dots, (f_k, score_k)\}$$

4) 结果整合

在关键帧的提取过程中，可以得到每个关键帧的所属视频以及其所属镜头占视频的权重 $weight$ 。

$$f_i \rightarrow (Video, weight_{mi})$$

表示图像序列 f_i 属于视频 $Video_m$ 且其所占权重为 $weight_{mi}$ 。

整合后的检索图像序列为：

$$f_{query} \rightarrow \{(Vieao_1, score_{1i} \times weight_{1i}), (Vieao_1, score_{1k} \times weight_{1k}), \dots, (Vieao_m, score_m \times weight_m)\}$$

最后合并所有的检索图像得到最终的视频对应序列 $video_{query}$ ，得到重复视频结果的候选集。

4.3. 视频质量评价

在得到重复视频候选集后，要通过质量比较后再进行进一步的去重工作。以保留质量最好的视频文件。关于视频的质量评价算法，目前已有大量的研究。对于视频的质量评价是以人眼观测为标准的，因此主观的人眼观测方法也是最精准的方法，但需要耗费大量的人力，不能用于大规模的应用。客观的质量评价方法是目前的主要研究方向。与图像质量评价算法相似，根据算法对原始数据的依赖程度，客观的视频质量评价算法主要分为，全参考评价算法，半参考评价算法和无参考评价算法。其中全参考评价算法需要利用整个原始视频的全部信息，用完整的原始视频与待评价视频做比较，得到与其相似度，经典的全参考算法如 MSE 和 PSNR 被广泛应用，还有研究人员对 MSE 和 PSNR 进行改进，提出基于结构相似度的视频质量评价方法[11]。但这类算法需要客户端上传整个原始视频至服务器，由于带宽限制，很难实现。半参考评价算法从原始视频中提取特征值，传至服务器用以比较待评价视频与原始视频的相似度，虽然传输所占带宽较全参考视频小，但其准确度有一定的削减。半参考的质量评价算法常适用于在线视频质量实时监控中。

在这里更适合使用无参考的视频质量评价模型。无参考的视频质量评价模型中有很多分类，主要分为像素域的无参考视频质量评价还有压缩域的视频质量评价。像素域的无参考评价又有以下几种形式，基于失真种类的评价算法，通过对视频失真种类进行分类，分析在该分类中失真效应；基于像素差异的估计评价算法还有基于神经网络的算法等。林翔宇在文献[12]一文中提出了一种基于失真度估计的无参考视频质量评价算法。该算法通过计算视频的局部失真度与全局失真度结合得到整体视频的失真度，并通过帧间预测和帧内预测得到视频复杂度，最后结合失真度和计算法复杂度得到最终的视频质量。

在我们的视频去重方案中，视频质量评价过程将直接采用文献[12]中提出的质量评价算法。在得到视频候选集后计算候选集内视频的质量。通过质量比较标记质量最好的视频，同时对其余重复视频进行选择删除。至此完成了整个框架中视频去重的任务。

5. 实验分析

实验采用 CC WEB VIDEO 数据集与 TRECVID2010 数据集。

其中 CC WEB VIDEO 是视频检索领域常用的视频数据集，由香港城市大学和卡耐基大学共同完成。它包含 24 个不同的关键词，统计了不同的关键词分别在谷歌视频，雅虎视频和 youtube 上的检索的冗余结果。其中冗余主要存在编码格式的不同，帧率、比特率、分辨率等格式差异，以及颜色、亮度等光学变化和添加字幕等内容变化。图 3 显示为 CC WEB VIDEO 视频数据集的相关信息。

我们选取 TRECVID2010 中 400 个视频序列，并将其分类为动画，风景，体育，新闻四大类。分别统计每一类视频中的镜头数。用来判断我们镜头分割算法的准确率。图 4 为 TRECVID2010 视频数据集的部分视频内容。

5.1. 关键帧提取算法

关键帧提取中关键在于可变阈值镜头分割算法对镜头正确划分的准确率。在这里采用多媒体信息检测中常用的两个公式来计算该基于镜头的可变阈值分割算法的性能。分别用查全率和查准率表示。查全率指的是所检测出的镜头占整体视频镜头数的百分比，从而反映出检索的全面性。查准率指的是，在所有检测出的镜头中，所能正确检测出视频镜头所占的百分比，从而反应出检索的准确性。计算公式如下：

$$\text{查全率} = \frac{\text{正确检出数}(P)}{\text{正确检出数}(P) + \text{漏检数}(S)} \times 100\%$$

$$\text{查准率} = \frac{\text{正确检出数}(P)}{\text{正确检出数}(P) + \text{错误检出数}(N)} \times 100\%$$

对 TRECVID2010 数据库中四类视频进行镜头划分，经过统计得到表 1 中数据。

从数据中可以看出，该算法有较高的查全率和查准率。

5.2. 匹配排序过程

对 CC WEB VIDEO 实验库中的视频序列按照方案的流程从进行匹配检索。这里我们只分析最终结果，准确率定义为可以检测出相似视频占相似视频总数的百分比。分别分析带权重与不带权重情况下的检测准确率。表 2 中显示为，在 CC WEB VIDEO 数据库视频中分别进行带权重的匹配和不带权重匹配后的结果的准确率。

经过分析可得，匹配过程中选择考虑关键帧权重的算法可以使得检索结果更接近真实的情况。特别是当镜头数量较多时，关键帧较多，且差异更小。使得其对查重检索的结果影响更小，使得结果更加准确。

| Queries | | | Near-Duplicate | |
|---------|---|-------|----------------|-----|
| ID | Query | # | # | % |
| 1 | The lion sleeps tonight | 792 | 334 | 42% |
| 2 | Evolution of dance | 483 | 122 | 25% |
| 3 | Fold shirt | 436 | 183 | 42% |
| 4 | Cat massage | 344 | 161 | 47% |
| 5 | Ok go here it goes again | 396 | 89 | 22% |
| 6 | Urban ninja | 771 | 45 | 6% |
| 7 | Real life Simpsons | 365 | 154 | 42% |
| 8 | Free hugs | 539 | 37 | 7% |
| 9 | Where the heil is Matt | 235 | 23 | 10% |
| 10 | U2 and green day | 297 | 52 | 18% |
| 11 | Little superstar | 377 | 59 | 16% |
| 12 | Napoleon dynamite dance | 881 | 146 | 17% |
| 13 | I will survive Jesus | 416 | 387 | 93% |
| 14 | Ronaldinho ping pong | 107 | 72 | 67% |
| 15 | While and Nerdy | 1771 | 696 | 39% |
| 16 | Korean karaoke | 205 | 20 | 10% |
| 17 | Panic at the disco I write sins not tragedies | 647 | 201 | 31% |
| 18 | Bus uncle | 488 | 80 | 16% |
| 19 | Sony Bravia | 566 | 202 | 36% |
| 20 | Changes Tupac | 194 | 72 | 37% |
| 21 | Afternoon delight | 449 | 54 | 12% |
| 22 | Numa Gary | 422 | 32 | 8% |
| 23 | Shakira hips don't lie | 1322 | 234 | 18% |
| 24 | India driving | 287 | 26 | 9% |
| Total | | 12790 | 3481 | 27% |

Figure 3. The related information of the CC WEB VIDEO video database

图 3. CC WEB VIDEO 视频数据集的相关信息



Figure 4. CC TRECVID2010 video data set part content

图 4. CC TRECVID2010 视频数据集部分内容

Table 1. Recall and precision of key frame extraction algorithm

表 1. 关键帧提取算法的查全率和查准率

| | 镜头数 | P | N | 查全率 | 查准率 |
|----|-----|-----|---|-------|-------|
| 动画 | 207 | 207 | 2 | 100% | 99.1% |
| 风景 | 126 | 125 | 4 | 99.2% | 96.8% |
| 体育 | 262 | 258 | 8 | 98.4% | 96.9% |
| 新闻 | 180 | 179 | 1 | 99.4% | 99.4% |

Table 2. The accuracy of the video deduplication program is weighted
表 2. 该视频去重方案在是否带权重情况下的准确率

| | 带权重 | 不带权重 |
|-----|-------|-------|
| 准确率 | 98.3% | 92.4% |

5.3. 实验结果总结

我们在 CC WEB VIDEO 和 TRECVID2010 两大视频数据集上, 对视频去重的方案做了模拟实验。分别从关键帧提取与匹配检测两大过程进行了实验分析。实验结果显示, 该方案的关键帧提取在查准率和查重率上达到了理想的水平且采用加权重的匹配方式可以在一定程度上提高重复视频检索的准确率。

6. 结束语

视频数据是当今互联网数据中的重要组成部分, 在存储和传播中产生了大量冗余, 本文提出了一种基于感知哈希的自适应阈值关键帧提取方法。并在此基础上提出了一个视频去重的方案, 该方案包括关键帧提取, 匹配排序和视频质量比较。并通过现有的视频质量评价算法, 评估视频候选集的视频质量, 最终进行视频去重操作。实验证实该方案具有较高的准确率。未来计划完善视频去重方案模型功能, 结合视觉模型特征, 按照人类视觉对视频内容的不同的关注度, 对视频内容进行分块, 并赋予不同的权重。从而使得视频去重获得更高的效率。

基金项目

国家自然科学基金(No.61702563), 中民族大学中央高校基本科研基金(No.CZP17075)。

参考文献

- [1] Stanek, J., Sorniotti, A., Androulaki, E., *et al.* (2014) A Secure Data Deduplication Scheme for Cloud Storage. *International Conference on Financial Cryptography and Data Security*, Springer Berlin Heidelberg, 99-118. https://doi.org/10.1007/978-3-662-45472-5_8
- [2] Katiyar, A. and Weissman, J. (2011) ViDeDup: An Application-Aware Framework for Video De-Duplication. *USENIX Conference on Hot Topics in Storage and File Systems*, USENIX Association, 7.
- [3] Lee, S. and Yoo, C.D. (2008) Robust Video Fingerprinting for Content-Based Video Identification. *IEEE Transaction on Circuits and Systems for Video Technology*, **18**, 983-988. <https://doi.org/10.1109/TCSVT.2008.920739>
- [4] Lee, S. and Yoo, C.D. (2008) Robust Video Fingerprinting Based on 2D-OPCA of Affine Covariant Regions. *IEEE International Conference on Image Processing (ICIP)*, San Diego, CA, 12-15 October 2008, 2156-2159.
- [5] Mohan, R. (1998) Video Sequence Matching. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, **6**, 3697-3700. <https://doi.org/10.1109/ICASSP.1998.679686>
- [6] Kim, C. and Vasudev, B. (2005) Spatiotemporal Sequence Matching for Efficient Video Copy Detection. *IEEE Transaction on Circuits and Systems for Video Technology*, **15**, 127-132. <https://doi.org/10.1109/TCSVT.2004.836751>
- [7] 牛夏牧, 焦玉华. 感知哈希综述[J]. 电子学报, 2008, 36(7): 1405-1411.
- [8] 张慧, 牛夏牧. 基于人类视觉系统的图像感知哈希算法[J]. 电子学报, 2008, 36(s1): 30-34.
- [9] Rashid, F., Miri, A. and Woungang, I. (2016) Secure Image Deduplication through Image Compression. *Journal of Information Security & Applications*, **27**, 54-64. <https://doi.org/10.1016/j.jisa.2015.11.003>
- [10] Fridrich, J. and Goljan. M. (2000) Robust Hash Functions for Digital Watermarking. *International Conference on Information Technology: Coding & Computing*, Las Vegas, NV, 29-29 March 2000, 178-183. <https://doi.org/10.1109/ITCC.2000.844203>
- [11] Wang, Z., Lu, L. and Bovik, A.C. (2002) Video Quality Assessment Using Structural Distortion Measurement. *2002 International Conference on Image Processing*, Rochester, NY, 22-25 September 2002, 65-68.
- [12] 林翔宇. 无参考视频质量评价方法研究[D]: [博士学位论文]. 杭州: 浙江大学, 2012.

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2325-2286，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：sea@hanspub.org