基于场景分割的视频内容语义管理机制

邢 玲[1,2],马 强[2],胡金军[2]
(1. 河南科技大学信息工程学院,河南洛阳 471023;2. 西南科技大学信息工程学院,四川绵阳 621010)

摘 要: 针对视频内容管理在不同层面存在语义鸿沟的问题,提出基于 UCL(Uniform Content Locater)的视频语义描述框架,该框架包含了三个层次的语义:内容语义、控制语义以及物理属性信息. 而视频场景的分割则通过视频内容基于时空上的相似性实现. 对于每个视频场景,结合局部纹理复杂度、背景亮度和场景复杂度,选择最佳参考帧(I 帧)与非最佳参考帧(非 I 帧)以嵌入不同的语义信息:控制语义、物理属性信息嵌入 I 帧,内容语义嵌入非 I 帧. 利用数字语义水印技术来实现视频内容的语义管理,完成语义信息和载体信号的一体传输和存储. 实验中采用 JM 参考模型进行数字水印方法的验证,结果表明该方法鲁棒性强,且不会造成视频资源质量显著下降.

关键词: 视频描述; 语义管理; 语义水印; 场景分割; UCL
中图分类号: TN911.7 **文献标识码:** A **文章编号:** 0372-2112 (2016)10-2357-07
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2016.10.011

A Semantic Management Mechanism for Video Resources Based on Scene Segmentation

XING Ling[1,2], MA Qiang[2], HU Jin-jun[2]
(1. *School of Information Engineering, Henan University of Science and Technology, Luoyang, Henan* 471023;
2. *School of Information Engineering, Southwest University of Science and Technology, Mianyang, Sichuan* 621010, *China*)

Abstract: To tackle video management problem of semantics gap existing in different aspects, a video semantic description framework based on UCL (Uniform Content Locator) is proposed. The semantic description framework consists of three levels, i.e., content, control and physical. Video to be semantically managed is divided into different scenes based on spatial-temporal similarities of frames. For every scene, the most optimal reference frame (I-frames) and non-optimal reference frames (non I-frames) are selected based on local texture complexity, background luminance and scene complexity. Content semantic are imbedded into non I-frames while control and physical semantics are imbedded into I-frames. A semantic watermarking algorithm is incorporated into the management to realize the efficient storage and transmission of video content and its video semantics. JM reference model is adopted for experiments to verify the watermarking technique and results show that the method is robust and has little side effect on video quality.

Key words: video description; semantic management; semantic watermark; scene segmentation; UCL (uniform content locater)

1 引言

作为互联网络中最主要应用之一的视频业务,由于其内容具有易复制、易分发、难管理、难监控等特性,视频内容的有效管理成为了近年来的研究热点. 最早提出视频内容语义管理机制的是由 ETSI 等 300 多家工业组织制定使用的 EPG(Electronic Program Guide),它为数字视频内容创建了一组特有的表格,且使用单独的 TS 流进行传输[1];后续的研究如基于内容情感选项的视频建模与检索方法[2]、基于网络对于视频内容的分发与存储管理[3]、体育视频内容标志镜头分类与管理[4]. 与 EPG 有类似的特点,这些方法都是将视频数据和语义管理数据进行单独传输和存储,无法实现高效地信息一体化传输.

数字视频内容管理的困难主要集中在三个方面:(1)无内容语义描述集,导致视频内容重复冗余度高;

收稿日期:2014-12-17;修回日期:2015-01-30;责任编辑:郭游
基金项目:国家自然科学基金(No. 61171109);四川省科技厅应用基础项目(No. 2014JY0215);西南科技大学科研项目(No. 2014JY0215)

(2)无传输控制语义集,导致视频传播管控难度加大;(3)无安全语义集,造成了源端回溯不可信,缺乏认证安全性.视频数字水印技术的出现使得视频内容的版权信息得到了保证,且版权信息与载体同步传输,视频通信效率得以提高,如基于水印的开发式视频管理管理框架[5].但该方法输出端只能检测水印是否存在,以完成视频片段的认证,在无法获得水印原始信息的条件下,难以达到对视频内容的智能管理.由于视频资源仍然使用统一资源定位符(Uniform Resource Locator,URL)标识其引用,导致同一内容视频本体因无强制语义计算而得以重复冗余发布.因此,研究可靠的视频语义模型,结合数字水印技术,将提高视频内容的有效管理.

本文提出了一种基于语义水印的视频内容管理机制.在 UCL(Uniform Content Locater,UCL)的基础上[6],结合视频检索、内容管控等要求,提出 UCL 视频语义描述框架,框架中包括内容语义、控制语义以及物理属性信息;结合 H.264 视频具体编码算法,采用基于场景分割的视频语义水印算法,将控制语义、物理属性信息和内容语义信息分别嵌入所选视频场景中不同的视频帧中,以提高水印嵌入容量和水印的鲁棒性.利用数字语义水印技术实现视频内容的语义管理,完成语义信息和载体信号的一体传输和存储.最后基于 JM10.2 参考模型,对语义水印方法可见性、和鲁棒性进行了验证.

2 基于场景分割的视频语义水印方法

2.1 基于 UCL 的视频语义描述框架

视频数字水印信息针对不同的应用有不同的语义要求,如针对视频检索,有根据节目内容提出的语义要求,有根据节目名称提出的语义要求;针对网络可控,有对发布者、接收者、节目分级等方面的语义要求.这些多样的语义需求,要求提出相对普适的语义模型,以实现内容识别、选择、以及业务监管的功能.结合语义的物理特征(如摘要等纯文本信息量大,且对控制语义等信息鲁棒性较低),构建基于 UCL 的视频语义描述框架,如图 1 所示.其中包括:内容语义,控制语义和可选的物理属性信息.

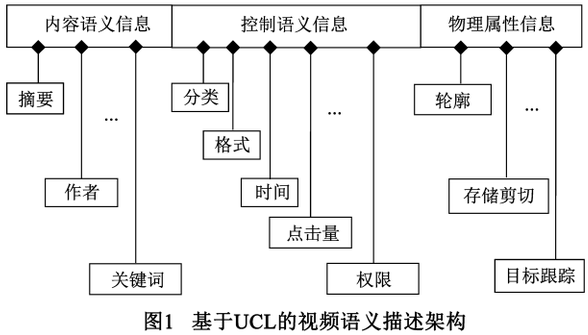

图1 基于UCL的视频语义描述架构

令视频的 UCL 语义模型为 $U\{U_{1x},U_{2y},U_{3z}\}$,其中 U_{1x} 属于内容语义,为纯文本信息,U_{2y} 为控制语义,为映射编码信息,U_{3z} 为可选的物理属性信息,为映射编码信息,x,y,z 分别为信息的元素个数.U_{1x} 包括的语义信息有:摘要、作者、标题、出版者、日期、关键词、扩充项;U_{2y} 包括的语义信息有:分类、格式、时间、点击量、语言、类型、文件大小、权限、扩充项;U_{3z} 包括的语义信息有:轮廓、存储剪切、拷贝许可、目标跟踪、扩充项.语义模型 U 中元素的大小或多少,与视频的具体应用背景有关,但并不影响视频内容的语义管理机制.

2.2 视频场景分割

场景指一个镜头所包含的视频帧序列.同一个场景,帧之间具有很强的相关性,因此可以利用这种时域和空域的相关性对一个场景进行压缩编码.另外,针对传输过程中的主动攻击,如帧删除、帧重组、帧平均,很难对整个场景进行完全删除或破坏的毁灭性的攻击.因此,论文通过利用场景分割技术来增强水印信号的鲁棒性,以提高针对时间同步攻击的自适应抵抗力.

目前,对于视频场景分割的研究,主要有像素比较、模板比较、直方图等方法,但他们有些共同的弊端,如算法复杂度较高,实时性不够强[7~11].考虑到视频数字水印实时性和视频解码同步的性能需求,论文提出了基于 DCT 系数变化量比较方法实现对视频场景的分割.由于图像的能量主要集中在变换域的 DC 系数上,相对离散的像素点具有更稳定的对应关系.结合视频编解码的子块结构,选择针对 16×16 宏块变换域 DC 系数做比较,如式(1):

$$Var(i)=\frac{1}{N}\sum_{n}(D(i,a,b)-D(i-1,a,b))^2 \quad (1)$$

其中 $D(i,a,b)$ 表示第 i 帧图像宏块 (a,b) 的 DC 系数,$Var(i)$ 表示则第 i 帧图像相对于前一帧图像的 DC 系数改变量,其中 $N=(a+b)*16$.由于 DC 系数表示子块图像像素点的均值,所以用宏块像素均值取代宏块的整数 DCT 变换,从而进一步降低算法的复杂度.

空间相似性 $Var(i)$ 越小,表示相邻两帧属于同一场景的可能性就越大,而 $Var(i)$ 值较大时,既可表示相邻两帧属于不同场景,也可表示同一场景中物体运动较为剧烈或背景变化较快,因此需要进一步计算他们的时间相似性.

$Var(i)$ 本身也可表示当前帧变化的剧烈程度,所以通过计算这种剧烈程度的放大或缩小的倍数来表示时间相似性,如式(2):

$$\alpha(i)=\frac{Var(i)-Var(i-1)}{\min(Var(i),Var(i-1))} \quad (2)$$

从式(2)可以看出该式为双极性式,$\alpha(i)$ 小于 0 表

示剧烈程度缩小的倍数,否则为放大倍数,$\alpha(i)$越接近0表示他们的时间相似性越高. 为了降低相对静止的两帧图像对场景分割产生误差,在计算 $\min(Var(i),Var(i-1))$时,若 $Var(i)<\overline{Var(x)}/3$,令 $Var(i)=\overline{Var(x)}/3$,其中,$\overline{Var(x)}$表示 x 场景中空间相似度的均值.

一个场景序列的第二帧相对于第一帧 DC 系数的改变量要小得多,$Var(x,2)<\beta_2$,变换的剧烈程度显著下降,$\alpha(x,2)<-\eta$;同理,下一个场景的第一帧相对于上个场景的最后一帧 DC 系数变化值很大,$Var(x-1,1)>\beta_1$,变换的剧烈程度显著增加,$\alpha(x-1,1)>\eta$,其中$(x,2)$为场景 x 第二帧的图像. 因此,综合考虑空间相似性和时间相似性,一个场景分割过程的首帧 F_f 和末帧 F_l 的判断式如式(3)、(4),其中 η 表示时间相似性的阈值,β_2 表示场景中第二帧图像的空间相似性阈值,β_1 为下一个场景中第一帧图像的空间相似性阈值.

$$F_f=\{i-1\mid\alpha(i)<-\eta\mid\mid Var(i)<\beta_2\} \tag{3}$$

$$F_l=\{i-1\mid\alpha(i)>\eta\mid\mid Var(i)>\beta_1\} \tag{4}$$

根据人眼的视觉特性,为了提高水印的不可见性,选择图像纹理复杂度高和帧间变化比较剧烈的场景嵌入视频数字水印信息. 将场景第二帧 DC 系数的梯度能量与第一帧 DC 系数改变量的乘积为场景复杂度 P,如式(5),

$$P=T(i)\times Var(i) \tag{5}$$

$$\begin{aligned}T(i)&=\frac{1}{N_a\times(N_b-1)}\sum_{N_a}\sum_{N_b}(D(i,a,b+1)-D(i,a,b))^2\\&+\frac{1}{(N_a-1)\times N_b}\sum_{N_a}\sum_{N_b}(D(i,a+1,b)-D(i,a,b))^2\end{aligned} \tag{6}$$

其中式(6)为 DC 系数的梯度能量,$D(i,a,b)$表示第 i 帧图像宏块(a,b)的 DC 系数,即像素均值,根据 P 的定义,其中 $i=2$,即为本场景序列中的第二帧,通过对 P 值与阈值的比较来选择适合嵌入水印的场景.

2.3　目标矩阵生成

由于人眼对嵌入水印变化域的敏感性较低,所以水印信息不仅和帧内纹理复杂度和背景亮度有关,帧间变化剧烈程度也同样影响着水印信息的不可见性. 为了使水印信息更接近于噪声信号,具有更好的不可见性,论文引入了场景复杂度,即综合考虑背景亮度、帧内空间复杂度、场景复杂度三要素来决定水印嵌入强度 S,形成一个目标矩阵 M.

针对所选中的分割场景,首先计算图像中每个 16×l6 宏块的背景亮度、帧内纹理复杂度,得出宏块的局部图像复杂度 H;然后,结合场景复杂度 P 得到水印嵌入强度 $S_{a,b}$,判断与阈值 S_{th}关系,当小于阈值时,水印的目标矩阵项 $M_{a,b}=0$,表示在此宏块不适合水印信息的嵌入;相反,$M_{a,b}=1$. 在视频解码端根据密钥再次生成目标矩阵,进行视频数字水印信息的检测与提取.

局部图像复杂度 H 的客观描述,来自于该宏块的灰度均值和纹理复杂度的加权组成的线性函数,如式(7):

$$H_{a,b}=\alpha_1\sigma^2_{a,b}+\alpha_2 e_{a,b} \tag{7}$$

其中,$1\leqslant a\leqslant N_1/16$,$1\leqslant b\leqslant N_2/16$,视频序列为 $N_1\times N_2$ 的图像,$e_{a,b}$为宏块的灰度均值,$\sigma^2_{a,b}$为宏块 Y 分量中 $Y_{a,b}$ 的纹理复杂度,$\alpha_1,\alpha_2\in[0,1]$为加权因子,其中宏块的纹理复杂度如式(8)所示:

$$\sigma^2_{a,b}=\frac{1}{8}\sum_{(i,j)\in Y_{a,b}}\vartheta(e_{a,b})\frac{\mid Y_{a,b}(i,j)-e_{a,b}\mid}{e_{a,b}} \tag{8}$$

$$\vartheta(e_{a,b})=(1/e_{a,b})^{\beta} \tag{9}$$

其中 $\vartheta(e_{a,b})$为加权系数,它作为修正因子来使宏块的纹理复杂度和灰度均值在同一个数量级成线性关系,论文中取值范围为 0.5~0.8.

为了减低过多修正因子给算法带来额外的计算复杂度,故将局部图像复杂度 $H_{a,b}$与场景复杂度 P 进行“×”操作得出水印嵌入强度,如式(10):

$$S_{a,b}=P\times H_{a,b} \tag{10}$$

其中,$S_{a,b}$的值随 α_1、α_2 和 β 取值而各异,从而生成不同的目标矩阵 M,因此可以将这三个参数作为水印算法中的密钥使用.

2.4　语义水印的嵌入与提取

H.264 中一个宏块包括一个 16×16 亮度分量 Y 和两个 8×8 的色差分量 Cb、Cr. 由于人眼对视频的色度较敏感,故算法仅考虑亮度分量 Y 信息. 首先,将视频图像的亮度分量 Y 分割成 16×16 块,则水印目标矩阵 M 的结构为 $N_1/16\times N_2/16$,其中 $M_{a,b}\in\{0,1\}$,$1\leqslant a\leqslant N_1/16$,$1\leqslant b\leqslant N_2/16$. 当 $M_{a,b}=1$ 表示 $Y_{a,b}$为水印信息的载体,然后,将 $Y_{a,b}$块划分为 16 个 4×4 子块,对每个子块进行整数 DCT 变换,如图 2 所示,左上角的 DCT_0为 DC 系数.

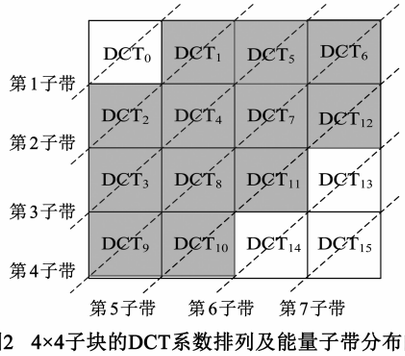

图2　4×4子块的DCT系数排列及能量子带分布的图

经过 DCT 变换后,4×4 子块能量如图 2 中的第 1 子带到第 7 子带逐渐递减. 其中 AC 高频系数(第 6、7

子带)多数为零不适合水印的嵌入,故选择第 2 子带到第 5 子带的 12 中频系数进行水印信息的嵌入,则嵌入规则如式(11)、(12)和式(13):

$$DCT_{mean} = \frac{1}{12}\sum_{i=1}^{12} DCT_i \tag{11}$$

$$DCT_{mean1} = \frac{1}{6}\sum_{i=0}^{2}(DCT_{i+3} + DCT_{i+10}) \tag{12}$$

$$DCT_{mean2} = \frac{1}{6}\left(\sum_{i=1}^{2} DCT_i + \sum_{i=6}^{9} DCT_i\right) \tag{13}$$

其中,DCT_{mean}为 12 个中频系数的均值,DCT_{mean1}为第 3 子带和第 5 子带 6 个中频系数的均值,DCT_{mean2}为第 2 子带和第 4 子带 6 个中频系数均值,通过调整 12 个中频系数来改变 DC_{Tmean}、DCT_{mean1}和 DCT_{mean2}三者之间的关系进行水印信息($w_{x,y}$)的嵌入,即为水印信息的编码,如式(14)和(15):

$$DCT_{mean1} > DCT_{mean} > DCT_{mean2},\ w_{x,y} = 1 \tag{14}$$

$$DCT_{mean2} > DCT_{mean} > DCT_{mean1},\ w_{x,y} = -1 \tag{15}$$

本文采用基于场景的语义水印算法,故将水印信息 U 中的 U_{1x}内容语义信息、U_{2y}控制语义信息和 U_{3z}可选的物理属性信息,采用相同的水印嵌入方案在不同的嵌入点进行水印嵌入操作,所以即使采用相同的水印嵌入方案,其生成目标矩阵的参数 P、α_1、α_2、β 及其阈值 S_{th}也不相同,这些都可作为密钥以提高水印的安全性.

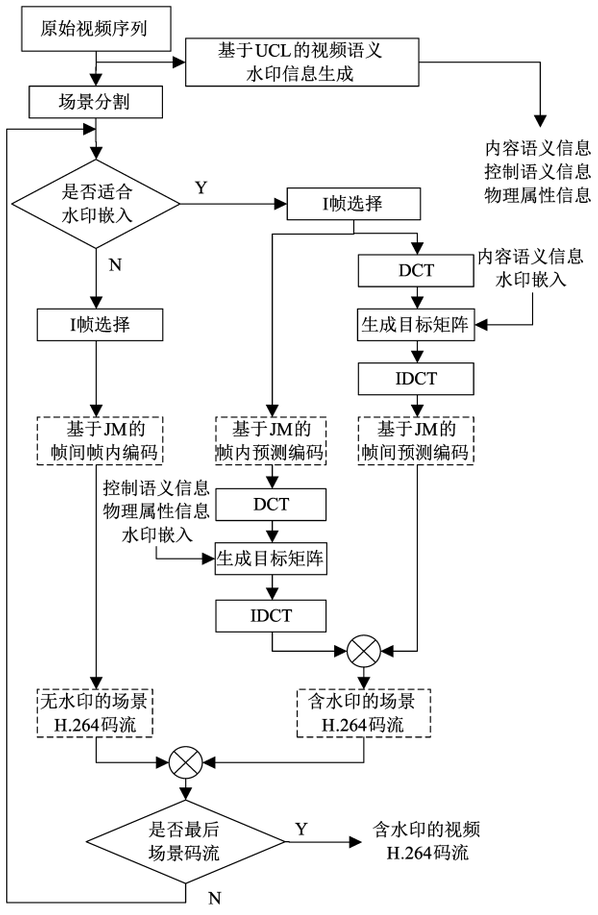

图3 场景分割的视频水印嵌入流程

结合压缩域水印嵌入量小和原始域水印鲁棒性较差的各自弊端,针对 H. 264 编码标准和 JM 实验的仿真平台,视频水印嵌入的流程为图 3 所示.

(1)对视频原始序列(YUV 格式文件)进行 UCL 标引并采用扩频技术生成视水印语义模型信息集 U,其中 U_{1x}属于内容语义,U_{2x}是控制语义,U_{3x}为可选的物理属性信息;

(2)对视频原始序列进行场景分割,形成基于场景的视频信息集 F,元素 $F(i,j,k)$中的 i 表示场景编号,j 表示帧图像相对场景的序号,k 表示在原始视频序列中帧编号;

(3)计算视频场景中场景复杂度,来选择适合进行水印嵌入的场景 $F'(i,j,k)$;

(4)将场景的第 y'帧作为最佳参考帧(I 帧),当一个场景的帧图像数大于 15,按照 GOP 标准生成 I 帧,其中满足式(16),其中 y 表示场景中视频帧的数量.

$$y' = \alpha * 15, \alpha \in 0,1,2\cdots, y''£\ y \tag{16}$$

(5)场景 $F'(i,j,k)$中的 I 帧 F'_I,借助 JM 编码器进行帧内预测编码,将压缩编码图像通过 DCT 变换,实现部分水印信息(U_{2y}和 U_{3z})的嵌入;

(6)对场景 $F'(i,j,k)$中的其余帧 F'_P 通过 DCT 变换完成部分水印信息(U_{1x})的嵌入,然后借助 *JM* 编码器对其进行帧间预测编码;

(7)将含有水印信息的 I 帧和 B、P 帧重新组合生成含水印的场景 H. 264 压缩码流;

(8)结合第(4)步的最佳参考帧选择算法,对第(3)步筛选出不适合水印嵌入的场景,借助 JM 编码器进行帧内和帧间编码,生成未含水印的压缩码流.

将第(7)步和第(8)步生成的压缩码流进行排序整合,生成基于 H. 264 的视频压缩码流.

水印的检测与提取在 H. 264 解码端完成,根据编码端对应的密钥生成目标矩阵 M 确定含水印的宏块位置,对含水印的宏块按照式(11)、(12)和(13)计算出 DCT_{mean}、DCT_{mean1}和 DCT_{mean2}的值,以重构水印信息,如式(17):

$$\hat{w}_{x,y} = \begin{cases} 1, & DCT_{mean1} > DCT_{mean} > DCT_{mean2} \\ -1, & DCT_{mean1} > DCT_{mean} > DCT_{mean2} \end{cases} \tag{17}$$

其中,$\hat{w}_{x,y}$表示宏块中的 4×4 子块(x,y)嵌入的水印信息,再经过重组就得到嵌入完整的的视频数字水印信息.

3 实验结果

实验中在 VS2008 开发环境中完成 JM10. 2 最佳参

考帧选择算法的移植和优化、原始视频序列的场景分割、水印的嵌入和含水印的 H.264 码流的解码工作,由 MatlabR2010b 对原始视频序列和含水印的视频序列进行数据统计,最后针对数字水印的性能指标得对算法进行性能评估. 视频采用标准视频序列 News、Foreman 和 Akiyo,所有视频序列都是 QCIF 格式(176 × 144),YUV(4: 2: 0),序列长度均为 300 帧,视频场景分割时参数选择为 $\eta=2,\beta_1=500,\beta_2=50$.

对于重构视频图像质量的判断,选择 PSNR(Peak Signal to Noise Ratio)峰值信噪比作为评判标准. PSNR 表示视频载体信号嵌入水印后的视频质量变化情况,其值越高表示其透明性越好,其计算过程如式(18):

$$PSNR = 10\log_{10}\left[\frac{\max_{\forall}(x,y)f^2(x,y)}{MSE}\right] \tag{18}$$

其中$\max_{\forall}(x,y)f^2(x,y)$为原始视频图像 f 上所有像素点中的最大像素值,针对 8bit 的灰度图像,其最大值为 255,则典型算法的 PSNR 值主要集中在 20 ~ 40dB 之间.

采用归一化互相关系数 NC(Normalized Correlation)用来度量重构的水印和原始水印之间的相似程度,如式(19):

$$NC = \frac{\sum_{k=0}^{N-1} W(k)W'(k)}{\sum_{k=0}^{N-1} W(k)^2} \tag{19}$$

其中,W 表示原始水印信息,W'表示提取出来的水印信息,N 为水印信号的长度,通常情况下,当 $NC>0.9$ 时,认为重构水印是可识别的.

3.1　水印的不可见性

视频数字水印的不可见性指确保人眼无法察觉,由于水印嵌入造成图像质量的下降. 实验中对 Akiyo 视频第 150 帧和 151 帧图像的原始序列图像、压缩后的图像、含水印的视频图像量的质量变化进行展示,其中第 150 帧为 H.264 编码中的最佳参考帧 I 帧,前者采用基于压缩域的视频水印嵌入方案,后者为基于原始域的视频数字水印嵌入方案. 从图 4 中很难察觉到由于压缩和水印的嵌入引起视频图像质量的变化.

(a)原始帧　(b)压缩后帧　(c)含水印帧

(d)原始帧　(e)压缩后帧　(f)含水印帧

图4　News视频序列第150帧(上)、151帧(下)

Akiyo 视频序列压缩后和嵌入水印后的视频图像前 90 帧的 PSNR 值如图 5 所示. 一般情况下,当 PSNR 值大于 30dB 以上,人眼就难以辨别两幅图像差别. 从图 5 可见,Akiyo 视频原始序列压缩后和嵌入水印后第 $y'(y=i*15)$帧的 PSNR 值较高,主要是由于第 y'帧在 H.264 视频编码中作为最佳参考帧,编码准确率最高. 总体上两曲线非常接近,且 PSNR 最小值为 35.91,说明本文提出的视频数字水印具有很好的不可见性.

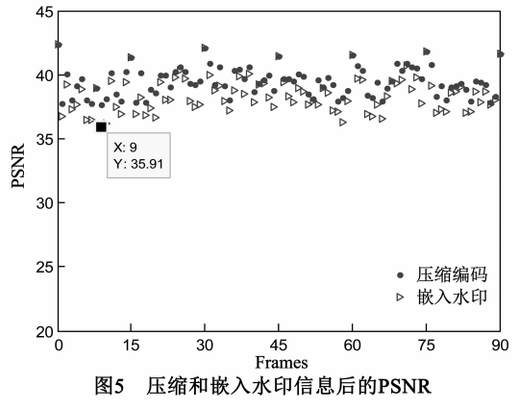

图5　压缩和嵌入水印信息后的PSNR

3.2　水印的鲁棒性

实验中若 $NC>0.9$,则认为该帧内含有水印信息,同一场景内有一幅图像含有水印信息,认为该场景为水印信号的载体. 实验对象为 Akiyo、News、Foreman、Sum 四个视频序列,其中 Sum 为前三者视频拼接序列. 对其分别统计视频序列的场景数(SC),含有水印信息的场景数(SC_w),检测到水印载体场景数(DSC_w),错误检测到的场景数(ESC_w),如表 1 所示.

表 1　嵌入水印的场景检查

视频序列	SC	SC_w	DSC_w	ESC_w
Akiyo	1	1	1	0
News	4	4	4	0
Foreman	5	3	3	0
Sum	12	8	8	0

从表 1 中可看出,在未受攻击的状态下,实验中嵌入信息的场景都能准确的检查出来. 由于在同一场景中嵌入相同内容,故实验中采用的水印场景检测标准($NC>0.9$)足以重构出原水印信息. 且由于采用原始域与压缩相结合的水印算法,所以为了进一步说明一个场景中关于内容语义水印信息的鲁棒性,以 News 视频序列为例,统计其前 90 帧(第 90 帧为第 2 个场景头帧)的 NC 值,如图 6 所示.

从图 6 中可知,第 $y'(y=i*15)$帧图像的 NC 值要明显高于其他图像,这是由于 y 帧为编码参考帧(I

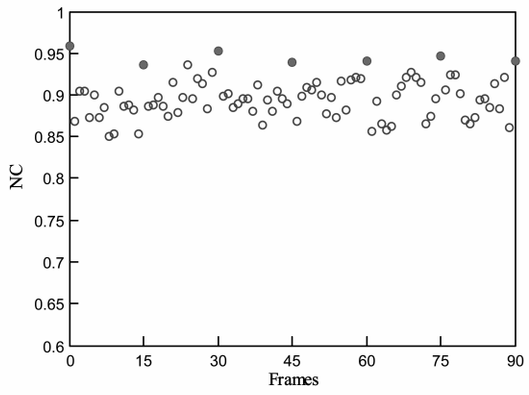

图6 News的NC值

帧),其量化后的非零 DCT 系数较多,且采用基于压缩域的水印方案,避免了由于视频信息的频繁解压缩,造成的水印信息丢失. 虽然非 I 帧域的 NC 值相对较低,但该域采用基于原始域的水印嵌入方案大大增加了水印信息的嵌入量,且该域的纯文本水印信息(摘要、关键词等)在 NC > 0.7 的情况下不会对语义理解造成歧义,一般情况下,NC > 0.6 就可以重构出水印信息.

3.3 抗噪声攻击能力

视频载体信号在传输和处理的过程中,最常见的攻击方式就是噪声攻击,因此水印算法抗噪能力是其性能评判的重要指标. 实验同样对 Foreman 视频序列的前 90 帧图像分别加载了密度为 0.005、0.01、0.03 的椒盐噪声,计算出重构视频图像的 PSNR 值和重构水印信息的 NC 值,其中 PSNR 值如图 7 所示. 可以看出,相对密度为 0.005、0.01、0.03 的椒盐噪声,水印信息对视频帧质量的影响反而更小,说明该算法对视频原始图像的影响几乎忽略不计. 在密度为 0.03 的椒盐噪声下 PSNR 最小值为 31.21,故重构的视频图像相对于原始图像的变化在人眼察觉范围之外.

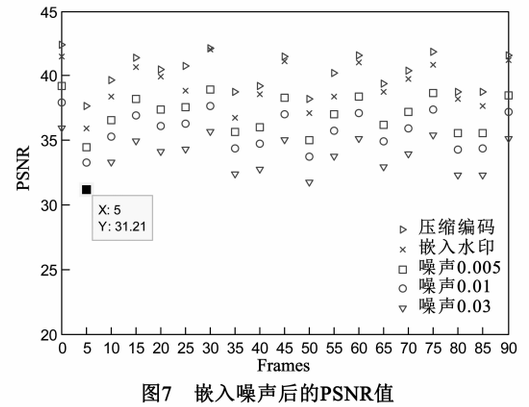

图7 嵌入噪声后的PSNR值

视频数字水印信息受到不同程度的噪声攻击后,NC 值变化情况如图 8 所示. NC 值出现很大程度的衰减,特别是非 I 帧($y \neq i * 15$)中的水印信息. 由于非 I 帧采用帧间预测编码,只保留部分残差信息,且该域水印信息经过 JM10.2 的重压缩编码,使该域水印信息的 NC 值衰减的相对比较厉害. 如在密度为 0.03 椒盐噪声下,最小 NC 值为第 5 帧(非 I 帧)的 0.4943,但经过对数据的统计发现,在相同强度噪声攻击下该场景中非 I 帧的最大 NC 值为 0.6357,由于同一场景内嵌入相同的水印信息,所以即使在较高密度的噪声攻击下,仍然可以重构出不影响人们观看的水印信息. 由此可见,对于噪声攻击,I 帧的鲁棒性表现的比较满意,故将 I 帧作为控制语义信息和物理属性语义信息的载体.

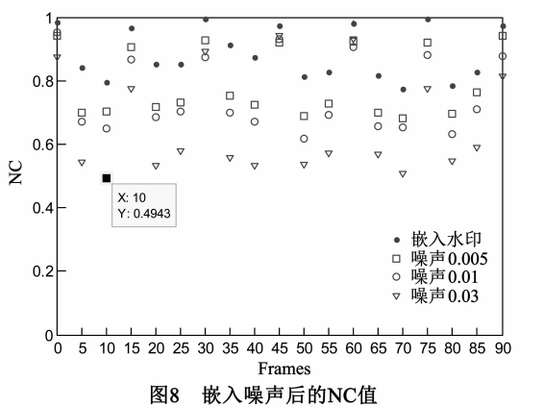

图8 嵌入噪声后的NC值

3.4 抗其它主动攻击能力

同时对 Akiyo、News、Foreman 三个视频分别进行重量化、中值滤波和帧删除攻击,视频数字水印受到攻击后的 NC 值如表 2 所示,结果为三段含水印的视频序列前 300 帧中,有效 NC 的均值. 由于同一场景中嵌入相同的水印信息,当 NC 值 < 0.5 时,视该帧水印信息为无效水印.

表 2 水印信息鲁棒性分析

攻击类型	Akiyo(NC)		News(NC)		Foreman(NC)	
	I 帧	P 帧	I 帧	P 帧	I 帧	P 帧
未受攻击	1.0	0.96	1.0	0.93	1.0	0.94
重量化(QP_{36})	0.85	0.65	0.87	0.72	0.81	0.70
中值滤波	0.89	0.72	0.91	0.79	0.89	0.69
帧删除	1.0	0.96	1.0	0.93	1.0	0.94

从表 2 中看出,I 帧中水印信息在遭受重量化、中值滤波和帧删除等攻击时,表现出较好的鲁棒性. 其中帧删除攻击对水印信息没有任何影响,主要是帧删除很难实现完全删除整个视频场景.

4 结论

数字视频内容的有效管理有助于互联网中视频业务高效、可靠的开展. 本文提出一种基于场景分割的视频内容语义管理机制,将语义模型从特性上分为三个

子集:内容语义、控制语义以及物理属性信息;对视频内容按照时间与空间相似性,构建基于 DCT 系数变化比较方法来实现视频的场景分割;并且按照背景亮度、帧内空间复杂度和场景复杂度三要素来决定语义水印嵌入位置即目标矩阵,通过修改 DCT 中 AC 系数实现水印的嵌入.将内容语义信息嵌入到场景中的非最佳参考帧,语义信息、物理属性信息则嵌入到最佳参考帧,利用数字语义水印技术实现了视频内容的语义管理,完成语义信息和载体信号的一体传输和存储.

参考文献

[1] Basic R,Mocinic M. User′s requirements for electronic program guide (EPG) in interactive television (iTV)[A]. Region 8 International symposium on video/image processing and multimedia communication[C]. Zadar :IEEE, 2002. 457 – 462.

[2] Alan H,Xu L Q. Affective video content representation and modeling[J]. IEEE transactions on multi-media, 2005, 7(1),143 – 154.

[3] 吴宣够,熊焰,印凤行. 树形网络中的一种有效视频内容分发算法[J]. 小型微型计算机系统,2013,34(8):1728 – 1731.
Wu Xuan-gou, Xiong Yan, Yin Feng-hang. An Efficient video content distribution algorithm for tree networks[J]. Journal of Chinese computer systems,2013,34(8),1728 – 1731. (in Chinese)

[4] 朱映映,朱艳艳,文振焜. 基于类型标志镜头与词袋模型的体育视频分类[J]. 计算机辅助设计与图形学学报,2013,25(9),1375 – 1383.
Zhu Ying-ying,Zhu Yan-yan,Wen Zhen-kun. Sports video classification based on marked genre shots and bag of words model[J]. Journal of computer-aided design and computer graphics, 2013, 25(9), 1375 – 1383. (in Chinese)

[5] 刘宇驰,等. 一种开放式视频管理框架[J]. 国防科技大学学报,2006(28),73 – 76.
Liu Yu Chi, et al. An open framework for video management[J]. Journal of national university of defence technology,2006(28),73 – 76. (in Chinese)

[6] XING Ling,MA Qiang,ZHU Min. Tensor semantic model for an audio classification system. SCIENCE CHINA Information Sciences,2013,56(6):1 – 9.

[7] Yun Z, Mubarak S. A General Framework for Temporal Video Scene Segmentation[A]. International Conference on Computer Vision[C]. Beijing: IEEE, 2005. 1111 – 1116.

[8] Panagiotis S,Vasileios M,Ioannis K,et al. Temporal video segmentation to scenes using high-level audiovisual features[J]. IEEE Transactions on Circuits and Systems for Video Technology,2011,21(8),1051 – 8215.

[9] Zhu S H,Liu Y C. Scene Segmentation and Semantic Representation for High-Level Retrieval[J]. IEEE Signal Processing Letters,2013,15,713 – 716.

[10] Mostafa T,Mahmood K,Shohreh K. Event Detection and Summarization in Soccer Videos Using Bayesian Network and Copula[J], IEEE Transactions on circuits and Systems for Video Technology,2014,24(2),291 – 304.

[11] He Hu,Ben U. Automatic object segmentation of unstructured scenes using colour and depth maps[J], IET computer vision,2014,8(1),45 – 53.

作者简介

邢　玲　女,1978 年 11 月生,四川攀枝花人,河南科技大学信息工程学院教授,硕士生导师,主要研究方向为网络信息智能处理与主动服务技术.
E-mail:xingling_my@163.com

马　强　男,1982 年9 月生,四川绵阳人,西南科技大学信息工程学院讲师,主要研究方向为多媒体安全认证、语义计算.
E-mail:maqiang_my@163.com

胡金军　男,1986 年6 月生,河南信阳人,西南科技大学信息工程学院硕士,主要研究方向为视频编解码、视频质量评估.
E-mail:hujingjun_my@163.com