

# 基于 AVS 的嵌入式音视频同步方法

陈 健, 赵 岩, 陈贺新

(吉林大学通信工程学院通信新技术重点实验室, 长春 130012)

**摘要:** 音视频同步是数字电视广播和多媒体通信等应用的关键技术。该文提出一种基于 AVS 并结合嵌入技术的音视频同步方法。将压缩音频数据嵌入 AVS 视频编码系统, 保证传输或存储、接收端解码与播放过程中的音视频始终同步。实验结果表明, 该方法实现了音视频完全同步, 能减小用于同步的开销。

**关键词:** 视频; 音频; 嵌入式; 同步

## Embedded Audio-video Synchronization Method Based on AVS

CHEN Jian, ZHAO Yan, CHEN He-xin

(New Communication Technique Key Lab, College of Communication Engineering, Jilin University, Changchun 130012)

**【Abstract】** Audio-video synchronization is the key technology in applications such as digital TV broadcasting and multimedia communication. This paper proposes an audio-video synchronization method based on AVS and embedding technique. It embeds compressed audio data into AVS video coding system to keep synchronization all the time between audio and video during transmission or storage, decoding at the receiver and playback. Experimental results show that this method can achieve synchronization between audio and video, and reduce overhead for synchronization.

**【Key words】** video; audio; embedded; synchronization

### 1 概述

随着 Internet 的快速普及和发展, 通信方式已从传统语音方式发展到语音、文字、图像、视频相融合的多媒体通信方式。在 MPEG-2 标准中, 音频编码和视频编码是独立进行的, 音频流与视频流之间的同步通过时间标签实现。针对 MPEG-2 同步方案的缺点, 国内外学者提出了很多针对 MPEG-2 进行改进的同步方案<sup>[1-3]</sup>。国外学者针对音视频同步问题提出利用语音(音频)与其唇形之间的对应关系解决视频会议、可视电话等应用中的音视频同步(唇同步)问题<sup>[4-5]</sup>。

AVS 标准是我国数字音视频编解码技术标准工作组制定的数字音视频编解码标准, 基于我国自主创新技术和国际公开技术构建<sup>[6]</sup>。本文提出一种基于 AVS 的嵌入式音视频同步方法, 给出系统整体同步方案、将音频数据嵌入视频编码器的嵌入方法和相应的解码器提取方法。

### 2 基于 AVS 的嵌入式音视频同步系统

基于 AVS 的嵌入式音视频同步系统编码器和解码器结构分别如图 1、图 2 所示。

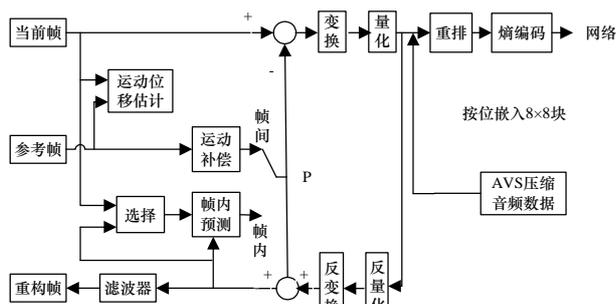


图 1 编码器结构

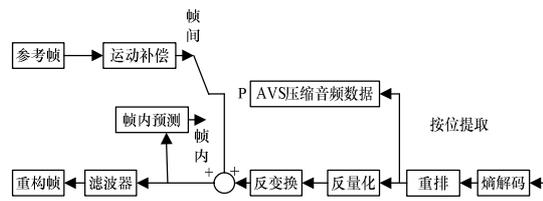


图 2 解码器结构

### 3 音频数据嵌入视频及其提取

#### 3.1 算法描述

编码器中音频数据的嵌入算法如下:

$$\begin{cases} C(u_B, v_B) = C(u_A, v_A) & \text{if } a_i = 0, C(u_A, v_A) \neq C(u_B, v_B) \\ C(u_B, v_B) = C(u_A, v_A) + 1 & \text{if } a_i = 1, C(u_A, v_A) = C(u_B, v_B), \\ & C(u_A, v_A) < -1 \\ C(u_B, v_B) = C(u_A, v_A) - 1 & \text{if } a_i = 1, C(u_A, v_A) = C(u_B, v_B), (1) \\ & C(u_A, v_A) > 1 \\ C(u_A, v_A) = C(u_B, v_B) + 1 & \text{if } a_i = 1, C(u_A, v_A) = C(u_B, v_B), \\ & |C(u_A, v_A)| \leq 1 \end{cases}$$

解码器中音频数据的提取算法如下:

$$\begin{cases} a_i = 0 & \text{if } C(u_A, v_A) = C(u_B, v_B) \\ a_i = 1 & \text{if } C(u_A, v_A) \neq C(u_B, v_B) \end{cases} \quad (2)$$

**基金项目:** 国家自然科学基金资助项目(60672100); 科技部国际合作基金资助项目(2005DFA10300); 吉林省科技发展计划基金资助项目(20070528)

**作者简介:** 陈 健(1981-), 男, 硕士, 主研方向: 数字图像处理, 音视频系统同步; 赵 岩, 副教授、博士; 陈贺新, 教授、博士、博士生导师

**收稿日期:** 2008-07-22 **E-mail:** zhao@email.jlu.edu.cn

其中,  $C(u_A, v_A), C(u_B, v_B)$  分别代表  $A, B$  两点量化后的系数值;  $a_i$  代表需要嵌入的音频比特值。

### 3.2 音频数据嵌入与提取的具体方法

#### 3.2.1 音频数据嵌入方法

设视频编码系统中, 量化后系数块内具有最小距离的 2 个中频系数分别为  $A$  和  $B$ , 如果待嵌入的音频编码比特为 0, 且  $A=B$ , 则  $A$  和  $B$  的值保持不变。如果  $A \neq B$ , 则改变系数  $B$  的值, 使  $B=A$ 。如果待嵌入的音频编码比特为 1, 且  $A \neq B$ , 则  $A$  和  $B$  的值保持不变。如果  $A=B$ , 则分为如下 3 种情况:

- (1) 如果  $A$  的值小于 -1, 则改变系数  $B$  的值, 使  $B=A+1$ ;
- (2) 如果  $A$  的值大于 1, 则改变系数  $B$  的值, 使  $B=A-1$ ;
- (3) 对于  $A$  取值的其他情况, 改变系数  $A$  的值, 使  $A=B+1$ 。

#### 3.2.2 音频数据提取方法

如果  $A=B$ , 则提取的音频比特为 0。如果  $A \neq B$ , 则提取的音频比特为 1。

在解码过程中恢复原视频系数的方式如下: 若提取的压缩音频比特为 0, 则使上述 2 个中频系数  $A$  和  $B$  的值保持不变。若提取的压缩音频比特为 1, 则改变调节系数  $B$  的值, 使其与参考系数  $A$  相等。

### 3.3 嵌入点的确定

在嵌入算法的应用中, 涉及固定点  $A, B$  的位置选取问题。

本文分析 AVS 视频编码中的  $8 \times 8$  块变换系数, 找到距离小于一定门限的一些点的集合, 并分别对这些点进行测试, 得到 PSNR, 比较并确定所需距离最小的  $A, B$  点。测试并分析  $8 \times 8$  块中的 64 个点, 选取的最佳嵌入点为  $A: (1,3), B: (2,2)$ 。

## 4 实验结果与分析

本文应用 AVS 编码器测试 5 种不同标准格式的序列, 且每种格式使用 2 个不同序列(即 2 种运动变化程度的序列)进行测试, 以全面考察本文算法的通用性。除了分辨率选项, 对各个不同序列的其他参数设置相同, 分别如下: 编码帧数为 10; 量化参数 QP 为 36; 帧间预测模式为  $16 \times 16, 16 \times 8, 8 \times 16, 8 \times 8$ ; 打开环路滤波器; 帧间交织编码为帧编码。

对算法各个序列的实验结果如表 1 所示。

表 1 各种序列的实验结果

格式	序列名	PSNR Y 下降/dB	嵌入开销/(%)
1080i	parkrun_ter_1080i.yuv	0.10	0.84
	Kayak_1080i.yuv	0.13	0.54
576i	parkrun_ter_576i.yuv	0.10	0.62
	husky_576i.yuv	0.13	0.36
480i	football_480i.yuv	0.11	0.72
	teacher_480i.yuv	0.06	0.66
CIF	bus_cif.yuv	0.13	1.51
	mobile_cif.yuv	0.18	1.15
QCIF	football_qcif.yuv	0.11	0.71
	tempe_qcif.yuv	0.04	1.05

PSNR Y 下降指未嵌入音频数据和嵌入音频数据后, AVS 视频解码器重构相应序列图像的平均  $Y$ (亮度) 帧的峰值信噪比(PSNR)之差。PSNR 是解码恢复图像质量的客观评价, 即

$$PSNR = 10 \lg \frac{255^2}{\frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [f(x, y) - F(x, y)]^2} \quad (3)$$

其中,  $M$  是图像的宽度;  $N$  是图像的高度;  $f(x, y)$  是原图像的像素值;  $F(x, y)$  是解码恢复图像的像素值。

嵌入开销指将音频嵌入视频带来的开销, 本文用百分数表示。嵌入开销的计算如式(4)所示。

$$O_e = \frac{H_e - O_v - A_e}{O_v + A_e} \times 100\% \quad (4)$$

其中,  $O_e$  代表嵌入开销(embedding overhead), 用百分数表示;  $O_v$  代表未嵌入音频的视频编码后生成的数据量(original video);  $H_e$  代表嵌入音频后的音视频混合编码后生成的数据量(embedded hybrid);  $A_e$  代表嵌入的音频数据量(embedded audio)。

分析实验结果, 得到如下结论:

(1) 音频数据嵌入对视频的影响。将音频嵌入视频必定对视频产生影响, 该影响可以通过客观和主观 2 个方面来衡量。客观衡量即信噪比 PSNR  $Y$  的下降程度。由表 1 可见, PSNR  $Y$  下降的最大值是 0.18 dB, 最小值只有 0.04 dB, 在视频处理中可以忽略不计。为了说明音频嵌入视频对重构图像主观质量的影响, 图 3 给出了运动最剧烈的 football\_qcif.yuv 第 9 帧原图像和未嵌入音频时以及嵌入音频后 AVS 视频解码器恢复图像。可以发现, 很难区分嵌入音频的重构视频与未嵌入音频的重构视频, 其差别达到了人眼无法察觉的程度。

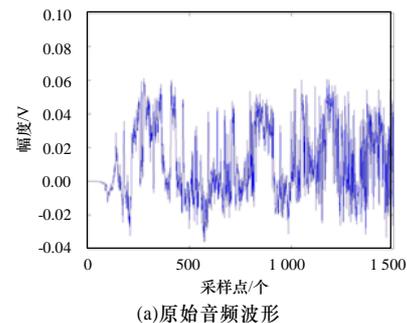


图 3 football\_qcif.yuv 序列第 9 帧原图像及其恢复图像

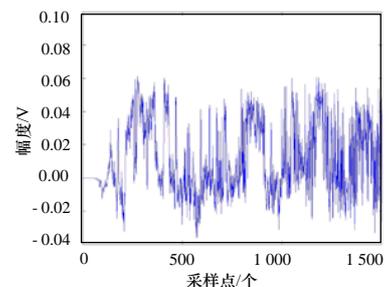
(2) 嵌入开销。由表 1 可见, 在嵌入开销最大值为 1.51%、最小值为 0.36% 时, 本文方法可以做到音视频同步应用。与 MPEG-2 中使用时间戳技术带来的 3% 的系统信道开销相比有了很大下降。

(3) 嵌入和提取处理对音频的影响。本文采用的嵌入与提取方法保证了所提取音频比特与嵌入比特完全相同, 因此, 本文嵌入式音视频同步方法对音频信息没有影响, 即解码器中提取出的压缩音频数据可以通过 AVS 音频解码器正确解码, 不存在因嵌入和提取操作而引入的失真。

图 4 给出了嵌入的原始音频信号波形和恢复后的音频信号波形, 可以看出, 音频信号中不存在由嵌入引起的失真。



(a) 原始音频波形



(b) 恢复后的音频波形

图 4 原始音频信号和恢复后的音频信号波形

(下转第 256 页)