

基于 DCT 频域条件帧间补偿法的 视频编码方案及其实现*

郑伟国 肖自美

(中山大学电子系, 广州 510275)

摘要 运动图象的压缩是实现多媒体通信的关键技术之一。本文提出一种基于离散余弦变换(DCT)系数的条件帧间补偿法及其视频编码方案。该算法利用 DCT 变换的特点和人眼视觉特性,快速高效地对运动图象进行编码。由于运算量低,该算法可以在 PC 机上用软件实时实现,适用于基于 PC 平台的可视电话、会议电视和可视多媒体通信系统。

关键词 运动图象压缩编码,频域条件帧间补偿,视频编码

1 引言

运动图象信息巨大数据量,给存储和通信带来了极大的困难。但是,我们可以利用运动图象的空间和时间冗余对其进行压缩。目前的国际标准 H. 261 和 MPEG 正是这类算法的代表。对于活动图象或准活动图象等序列图象,利用帧间的相关性,有效减少序列图象的时间冗余,是实现运动图象高效压缩的重要环节。因此,帧间编码是对运动图象进行压缩编码的关键技术。

运动补偿法(MC)和条件帧间补偿法(CFR)是帧间编码的两个基本方法。

H. 261 和 MPEG 标准都采用运动补偿法来作为其帧间编码方法^[1,2]。基于 DCT/MC 的视频编码结构(图 1),运动补偿是通过运动估值来进行预测或内插补偿的,象素块的运动由一个二维的运动矢量来描述。运动补偿帧间编码的核心问题是运动估值(ME)。运动估值假定图象在相邻帧之间的变化可以视为构成物体的那些象素在相邻帧之间发生了位移。即使采用快速算法,运动补偿 MC 也具有庞大的运算量,另外,在甚低码率的情况下,由于带宽

的限制,每秒传输的帧数较少,在出现跳帧的情况下,为了保证压缩效率,必然要求加大运动估值的搜索窗口的尺寸,导致 ME 运算量的进一步增大。这给 MPEG 和 H. 261 编码器的实时实现与降低成本带来很大困难。

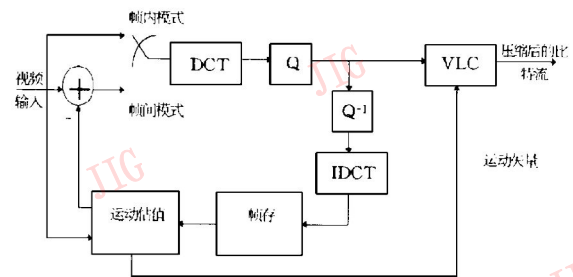


图 1 DCT/MC 视频编码方案

Fig. 1 A DCT/MC video coding scheme.

作为消除时间冗余度的帧间编码基本方法之一的条件帧间补偿法(CFR)的基本原理是:计算前后帧象素的帧间差值 FD ,并规定一个阈值 T ,如果 $|FD| \leq T$,则认为当前帧的象素位于序列图象的静止部分(如背景区)或相对静止部分,对于这类象素不进行传送。反之,如果 $|FD| > T$,则认为当前帧的象素位于运动物体区,需要进行传送^[3]。图 2 是一种 DCT/CFR 的视频编码方案。

* 广东省重点科技攻关项目资助(962200202)

收稿日期:1996 年 10 月;收到修改稿日期:1997 年 1 月。

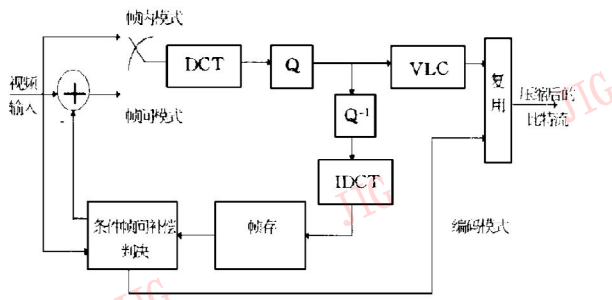


图 2 DCT/CFR 视频编码方案

Fig. 2 A DCT/CFR video coding scheme.

常规的条件补偿法通常是基于空间域的,具有算法简单,易于实现的优点,但压缩比较低。而且,其编码器比解码器除了要多做帧间检测外,还要多做一次 IDCT 变换和逆量化运算,因此,编解码方的运算量是不平衡的。另外,空域帧间检测对噪声比较敏感,阈值的选取比较困难。当光照条件好时阈值 T 较小,当光照条件较差时,图象的信噪比较低,此时,阈值 T 的取值要大些,才能减轻噪声的影响。由于空间域条件帧间补偿对摄像机、环境等的噪声很敏感。帧间变化检测阈值的优化值不能适应光照、摄像等环境的变化。为了提高复原图象的质量,可以把帧间变化检测阈值降低,但码率随之升高,若把阈值加

大,则容易出现漏判,这些错误严重影响复原图象质量。我们实践表明,采用空间域 CFR,图象容易出现假轮廓和方块效应,特别对于缓慢变化的平坦区,恢复图象的质量难以得到保证。

为了实现在 PC 平台上软件实时且高效的运动图象编码,本文提出一种基于 DCT 系数的条件帧间补偿法。该算法利用 DCT 变换的特点和人眼视觉特性,快速高效地对运动图象进行编码。由于其运算量低,特别适用于基于 PC 平台的可视电话、会议电视和可视多媒体通信系统。

2 基于频域的条件帧间补偿法

本文所提出的基于频域的条件帧间补偿法的基本思想是:由于空域的噪声经过变换后将分配到频域的各个系数上,因此可以在频域上进行条件帧间补偿检测,并根据人眼的视觉特性,引入视觉阈值矩阵进行误差屏蔽,从而可以很好地消除了图象噪声的干扰,并大大提高了帧间判决的准确性,从而提高了编码效率。该算法的运算量低,易于用软件实现。同时,编解码双方的运算量基本平衡,给应用带来了很大方便(图 3、4)。

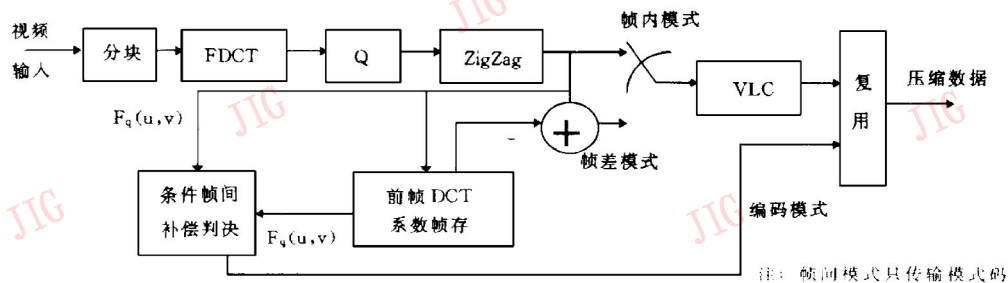


图 3 基于 DCT 频域的条件帧间补偿编码器框图

Fig. 3 Encoder block diagram of DCT frequency domain based CFR algorithm.

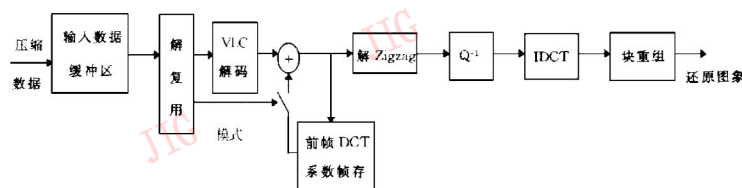


图 4 基于 DCT 频域的条件帧间补偿解码器框图

Fig. 4 Decoder block diagram of DCT frequency domain based CFR algorithm.

2.1 分块与块重组

分块是把输入图象分成 8×8 块,每一图象子块作为离散余弦变换(DCT)的输入。而块重组则把 8×8 图象块重新组成完整的画面。

2.2 FDCT 和 IDCT: 二维离散余弦变换对

我们注意到二维 DCT 变换系数 $F(i, j)$ 具有以下性质^[4]:

(1) 描述的是图象子块中不同频率下的振幅谱的表达,而且 F 矩阵的能量主要集中在 F 矩阵的左上角,即 F 的能量主要分布于直流分量和低频区。当图象子块各象素值为常数时, F 的能量 $A = F^2(0, 0)$ 。

(2) 各系数 $F(i, j)$ 几乎不相关。

(3) 图象子块中的噪声影响经过变换后分散到各个变换系数上。

因此,我们可以利用 DCT 系数进行条件补偿,并在频域利用视觉特性揪人眼对不同频率的响应不同,采用量化对噪声进行掩盖,从而既提高了判决的准确性,又提高了压缩比。

2.3 量化

量化器采用线性均匀量化器,量化定义为对 64 个 DCT 系数除以量化步长,四舍五入取整,如下式所示:

$$F_q(v, u) = \text{Integer Round}[F(v, u)/Q(v, u)]$$

$Q(u, v)$ 值随 DCT 系数的位置和处理分量 (Y, U, V) 的不同而有不同取值。

反量化表达式如下所示:

$$F_q(v, u) = F_q(v, u)Q(v, u)$$

量化矩阵的设计考虑了人的视觉特性

(1) 人眼的空间频率特性,人眼对空间频率呈带通特性,对高低频响应较低;

(2) 掩盖效应,图象内容剧烈变化的地方,即高频部分,允许相对较大的误差。并通过心理视觉实验,确定对应于不同频率的视觉阈值,即可确定对不同频率的量化器步长,我们采用了 JPEG 小组根据心理视觉实验而得的亮度 Y 和色度 Cr, Cb 的量化矩阵^[4,5]。

2.4 ZigZag 扫描

为了进行游程编码,首先要把二维的系数经扫描转换成一维排列,扫描顺序影响着零游程编码的

效率,因为 DCT 系数的能量通常集中在左上角,一般扫描顺序采用 Zig-Zag 方式(表 3)。

表 1 亮度信号的量化矩阵

Table 1 Quantization table for luminance component

| | | | | | | | |
|----|----|----|----|-----|-----|-----|-----|
| 16 | 11 | 10 | 16 | 24 | 40 | 51 | 61 |
| 12 | 12 | 14 | 19 | 26 | 58 | 60 | 55 |
| 14 | 13 | 16 | 24 | 40 | 57 | 69 | 56 |
| 14 | 17 | 22 | 29 | 51 | 87 | 80 | 62 |
| 18 | 22 | 37 | 56 | 68 | 109 | 103 | 77 |
| 24 | 35 | 55 | 64 | 81 | 104 | 113 | 92 |
| 49 | 64 | 78 | 87 | 103 | 121 | 120 | 101 |
| 72 | 92 | 95 | 98 | 113 | 100 | 103 | 99 |

表 2 色度信号的量化矩阵

Table 2 Quantization table for Chrominance

| | | | | | | | |
|----|----|----|----|----|----|----|----|
| 17 | 18 | 24 | 47 | 99 | 99 | 99 | 99 |
| 18 | 21 | 26 | 66 | 99 | 99 | 99 | 99 |
| 24 | 26 | 56 | 99 | 99 | 99 | 99 | 99 |
| 47 | 66 | 99 | 99 | 99 | 99 | 99 | 99 |
| 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 |
| 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 |
| 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 |
| 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 |
| 99 | 99 | 99 | 99 | 99 | 99 | 99 | 99 |

表 3 ZigZag 扫描顺序

Table 3 Zigzag scanning order

| | | | | | | | |
|----|----|----|----|----|----|----|----|
| 0 | 1 | 5 | 6 | 14 | 15 | 27 | 28 |
| 2 | 4 | 7 | 13 | 16 | 26 | 29 | 42 |
| 3 | 8 | 12 | 17 | 25 | 30 | 41 | 43 |
| 9 | 11 | 18 | 24 | 31 | 40 | 44 | 53 |
| 10 | 19 | 23 | 32 | 39 | 45 | 52 | 54 |
| 20 | 22 | 33 | 38 | 46 | 51 | 55 | 60 |
| 21 | 34 | 37 | 47 | 50 | 56 | 59 | 61 |
| 35 | 36 | 48 | 49 | 57 | 58 | 62 | 63 |

2.5 频域条件帧间补偿的判决算法

① 设 $\text{MODE} = \text{INTER}$;

② 比较 $F_q(u, v)$ 和前一帧对应子块 $F_q(u, v)$, 如果 $F_q(u, v) \neq F_q(u, v)$ 则设 $\text{MODE} = \text{INTRA}$, 转④;

③ 如果比较完 64 个系数, 转⑤; 否则转②;

④ 比较当前矩阵 $F_q(u, v)$ 编码长度的估计 $\text{Len}(F_q)$ 和块差 $F_q - F'_q$ 矩阵编码长度的估计 $\text{Len}(F_q - F'_q)$, 如果 $\text{Len}(F_q - F'_q) < \text{Len}(F_q)$ 则设 $\text{MODE} = \text{DIFF}$;

⑤ 其中, $\text{Len}()$ 首先对 DC 系数的编码长度, 若是帧内模式按 DPCM 编码方式进行估计; 对余下的

63 个系数,按零游程编码的方式进行计算。

如果 $MODE = INTER$, 指示此块与参考块相同,作帧间编码。

如果 $MODE = INTRA$, 指示此块作帧内编码。

如果 $MODE = DIFF$, 指示此块作帧差编码,对块差矩阵进行编码。

2.6 系数编码

图象子块经过 DCT 变换后,能量集中在 DCT 系数矩阵的左上角,经过量化后,出现零值的概率大大增加,因此,采用零游程编码将大大提高编码效率。

在帧内模式时,对直流分量 $F(0,0)$ 进行差分编码 (DPCM), 即对相邻块之间的 DC 系数的差值 $DIFF = DC_j - DC_{j-1}$ 编码,而对其他的 63 个交流系数 AC 进行零游程编码,而在帧差模式下,除了直接对 DC 差值 $F_q(0,0) - F'_q(0,0)$ 作编码,其余与帧内模式相同;帧间模式时,只需传送模式码。

最后,为了进一步提高编码效率,使用哈夫曼 (Huffman) 码对 DC 码和 AC 码进行编码。

我们从上述的编码器结构和算法可以看出,基于频域的条件帧间补偿编码器只须做一次 DCT 变换和量化运算,而且,其条件帧间补偿判决的算法是非常简单的。与 DCT/MC 的方案相比,不仅无需做复杂的 ME/MC,同时,可以省下了做 DCT 逆变换和逆量化的步骤,因此,经过优化和 DCT 采用快速算法,基于频域的条件帧间补偿编码器可以实现实时软件视频编码。

3 频域条件帧间补偿法实验结果

我们采用基于频域的条件帧间补偿法,对彩色序列图象进行实验。我们采用了标准图象序列 MISS AMERICAN, 彩色分量为 $Y:U:V = 4:1:1$, 图象大小为 352×288 , 假设帧率是 6fps, 跳 5 帧抽取序列。该实验是在配备 Intel 的 Pentium (120 的 PC 上进行的,软件平台是 Windows95, 采用 C 语言编程,测试程序是 32 位应用程序。时间测试只限于压缩编码运算。

从上面的数据表明:基于频域的条件帧间补偿法具有非常好的性能—具有较高的压缩比和良好的主观视觉效果。同时,编码运算时间是 0.227s。

作为比较,我们采用 MPEG-1 的视频压缩算法对上述图象序列进行压缩。由于 MPEG-1 输出

视频码率是固定的 1.2Mbit/s, 因此标准的 MPEG-1 视频编码器是带有码率控制的。为了检验 DCT/MC 算法的性能,我们在实现 MPEG-1 视频编码算法时,取消码率控制并设定量化参数,对 I、P 和 B 帧都固定 $Q = 30$, 运动估值采用全搜索算法。下表给出 MPEG 编码结果。

表 4 采用基于频域的条件帧间补偿法的编解码结果

Table 4 Result of new-cfr coding algorithm with "MissA"

| 当前帧 | 帧类型 | 编码长度 (byte) | 压缩比 (CR) | PSNR (dB) |
|------|-----|-------------|----------|-----------|
| M029 | I | 3689 | 41 | 38.34 |
| M035 | P | 1800 | 84 | 37.72 |
| M041 | P | 1853 | 82 | 37.96 |
| M047 | P | 2063 | 74 | 37.55 |
| M053 | P | 2080 | 73 | 37.68 |
| M059 | P | 2198 | 69 | 37.52 |

表 5 采用 MPEG 的视频编解码结果

Table 5 Result of MPEG encoding and decoding with "MissA"

| 当前帧 | 帧类型 | 编码长度 (byte) | 压缩比 (CR) | PSNR (dB) |
|------|-----|-------------|----------|-----------|
| M029 | I | 3785 | 40 | 38.12 |
| M035 | B | 1728 | 88 | 37.58 |
| M041 | B | 1742 | 87 | 37.71 |
| M047 | P | 1845 | 82 | 37.84 |
| M053 | B | 1747 | 87 | 38.05 |
| M059 | B | 1805 | 84 | 37.86 |

编码运算时间是 1.953s。在同样平台下,这两种视频编码方案的执行时间相差近 9 倍。MPEG 编码时间比基于 DCT 系数的 CFR 长许多的主要原因是 MPEG 除了 ME/MC 运算复杂,耗时很多外,逆 DCT 和逆量化也需不少的运算量。

4 结论

采用基于 DCT 的条件补偿法后,压缩比达到与 MPEG-1 相近的水平,复原图象具有良好的主观质量,基本看不到明显的方块效应、伪轮廓,消除了因为编码时帧间检测错误造成的图象损伤。同时,运算时间和 MPEG 相比,具有明显的优势。

该方法具有较强的抗输入(光照、摄像等引起

的)噪声的能力,判决准确性大为提高,因而提高了编码效率。和 MC/ME 相比,复杂度低,经过优化,该算法可望在配备 Pentium(100 的 PC 实现实时编解码,为实现 PC 平台上基于软件的低成本的可视电话、会议电视系统和可视多媒体通信系统开辟了新途径。

参考文献

1 Recommendation H 261, Video codec for audiovisual services at Px64kbit/s. Geneva:CCIT, 1990.

- 2 MPEG (ISO/IEC JTC1/SC29/WG11). Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s. ISO CD 11172, 1991.
- 3 肖自美. 图象信息与压缩编码技术. 广州:中山大学出版社,1996. 182~183.
- 4 Wellace G K. The JPEG still picture compression standard. Commun:ACM,1991. 31~43.
- 5 JPEG (ISO/IEC JTC1). Information technology-digital compression and coding of continuous-tone still images. ISO/IEC DIS 10918-1, 1991.



郑伟国 1992 年中山大学电子系毕业,1995 年获硕士学位,现正在攻读博士学位,曾参加包括国家自然科学基金、广东省自然科学基金、广东省重点科技攻关项目等课题的研究,1993 年度国家教委科技进步二等奖获得者,主要研究领域是语音和图象压缩编码,ATM 与分组视频,可视多媒体通信系统等。

A Novel Video Coding Scheme with DCT Frequency Domain Based Conditional Interframe Replenishment Algorithm and Its Implementation

Zheng Weiguo Xiao Zimei

(Department of Electronics, Zhongshan University, Guangzhou 510275)

Abstract A Novel Video Coding Scheme with conditional interframe replenishment (CFR) algorithm based on DCT frequency domain is proposed in this paper. Utilizing the characteristics of DCT and visual perception of human, this DCT-based CFR algorithm can encode and decode motion pictures quickly and efficiently. Its compression ratio and reconstructed picture quality are similar with DCT/MC coding scheme such as H. 261 and MPEG but with much lower complexity. Both simulation and practical video communication application have shown that this algorithm has high performance and ease of implementation. Because of its low computing complexity, it can be real-time realized by PC software. It can also work with H. 261/MPEG by replacing ME/MC module (the motion vector is fixed as zero) after some little modifications on coding scheme. This DCT-based CFR algorithm and the video coding scheme are very suitable for low cost full PC software videophone, video conference and multimedia communication system.

Keywords Motion picture coding, Conditional interframe replenishment in frequency-domain, Video coding