

一种在 H. 261 算法压缩域中实现 四画面合成的算法和性能分析

黄廷学¹⁾ 朱志祥²⁾ 戴冠中¹⁾

¹⁾(西北工业大学自动控制系, 西安 710072) ²⁾(西安邮电学院通信技术研究, 西安 710061)

摘要: 鉴于 H. 261 是多媒体通信领域中图象编码的重要算法, 因此实现 H. 261 算法的四画面合成具有重要意义。因传统四画面合成是在模拟域中完成, 故存在成本较高和质量较差的问题, 为了降低其实现成本及提高合成质量, 首先提出了一种在压缩域中实现四画面合成的算法; 然后在详细地论述 H. 320 和 H. 323 两种视频会议框架下的四画面合成方法(包括合成原理)的基础上, 给出了该合成算法以及程序流程; 最后给出了该算法的实验结果, 并和模拟域的实现方法进行了比较, 实验结果表明, 不仅合成质量有了提高, 而且实现成本也有较大的降低。

关键词: H. 261 算法 压缩域 四画面合成

中图分类号: TN919.81 **文献标识码:** A **文章编号:** 1006-8961(2002)07-0699-05

A Kind of Four-Image Composition Implementation and Performance Analysis in H. 261 Algorithm Compressed Domain

HUANG Ting-xue¹⁾, ZHU Zhi-xiang²⁾, DAI Guan-zhong¹⁾

¹⁾(Automatic Control Department, Northwest Polytechnical University, Xi'an 710072)

²⁾(Research Institute of Communication Technology, Xi'an Institute of Posts & Telecommunication, Xi'an 710061)

Abstract H. 261 is an important algorithm for video composition in multimedia communication field. Therefore, it is very significant to implement four-image composition of H. 261 algorithm. The traditional four-image is in analog domain. So there is a problem that the cost of implementation is high and its quality isn't very satisfying. In order to lower the cost of implementation and enhance its quality, this paper provides a kind of algorithm of four-image composition which is completed in compressed domain. This paper discusses in detail four-image composition methods in the two kinds of videoconference systems of both H. 320 recommendation and H. 323 recommendation. First, the composition principle is given. Then based on this principle, the composition algorithm and program flow chart are provided. Finally, its experiment's result is given and compared with that of implementation in analog domain. According to the result we can find not only its quality is enhanced but also its cost is lowered largely.

Keywords H. 261 Algorithm, Compressed domain, Four-image composition

0 引言

在多媒体通信领域中, 四画面合成技术是研究热点之一。大家知道, 四画面合成就是让多点通信业务下的业务终端, 能够同时看到其他多个终端的视频图象, 在理想情况下, 还要能够根据不同的要求配

置合成画面的个数以及画面的大小。由于 H. 261^[1] 是多媒体通信中, 压缩与编码视频图象的重要算法, 也是 H. 320^[2] 视频会议系统和 H. 323^[3] 视频会议系统必备的视频功能, 另外, 由于 H. 261 算法还可以应用到远程监控, 远程医疗以及远程教学等多媒体业务中, 因此, 研究基于 H. 261 算法的四画面合成技术具有十分重要的意义。

传统的四画面合成是在模拟域中完成,即将多个终端的图象数据先传输到一个汇接设备上,其中一路图象数据对应使用一个编解码器来把压缩数据转变成模拟图象;然后在模拟域中,将多路模拟图象合成为一路图象数据;最后用一个编解码器把合成后的模拟图象编码成压缩数据进行传输,但这种传统实现方法有两个重要缺点:一是由于需要完全解码和编码,因此产生较大的时延,二是需要多个编解码器,因而系统的成本比较高,不利于实用。

本文提出了一种在压缩域中实现四画面合成的方法,从而解决了传统方法的缺点,因它不仅具有较短时延,并且只使用一个编解码器,故成本较低。

目前,通信网络的交换模式可以归为电路交换和包交换两类,下面分别就这两种网络环境下,对基于 H. 261 算法的四画面合成技术加以探讨。

1 压缩域内电路交换下,基于 H. 261 算法的四画面合成

1.1 四画面合成原理

在电路交换网络(SCN)下的多媒体通信,其四画面合成处理,如果选择在通信终端来完成,则需要把多个视频流同时传输到一个终端上,可是由于电路交换网络(如 DDN 专线网或 ISDN 网络)的每个终端是通过一个电路通路接入网络来实现通信,且由于该业务终端只接受一个网络媒体流(譬如 H. 320 会议系统使用 H. 221 帧结构复合的视频、音频以及数据媒体流来进行传输),因此,多个终端同时向一个终端传送视频流不能够实现。于是,四画面合成一般选择在通信系统的某一中间设备上实现,例如 H. 320 会议系统的多点控制设备 Multipoint Control Unit(MCU)。并且,四画面的合成是选择在压缩域内,而不选择在像素域内处理,因为在压缩域内处理所耗的 CPU 及内存等资源要小得多,且处理时延也要小得多。

H. 261 算法有 QCIF 和 CIF 两种图象格式,如表 1 所示。

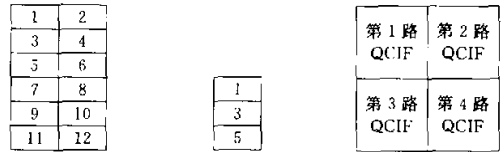
表 1 H. 261 算法的两种图象格式比较

像素纵横比 均为 3:1	CIF		QCIF	
	pixels/行	行/帧	pixels/行	行/帧
亮度 Y	352	288	176	144
色度 C _u , C _v	176	144	88	72

通过研究, QCIF 格式图象的大小是 CIF 格式

图象的 1/4。如果深入到图象的层次编码就可以发现,两种格式的像素层、宏块层以及块组层具有相同的结构,只是图象层结构不同,即 CIF 格式的图象包括了 12 个块组,而 QCIF 格式的图象只有 3 个块组(如图 1 所示)。

这样就可以把 4 路 QCIF 视频流合成为一个 CIF 视频图象,其合成示意图如图 2 所示。



(a) CIF 图象格式 (b) QCIF 图象格式

图 1 H. 261 算法的图象层组成



图 2 4 路 QCIF 视频流合成一路 CIF 图象的示意图

1.2 四画面合成的系统结构

四画面合成在通信系统的中间设备实现,该独立设备需要先从通信网络上接受 4 路 QCIF 视频媒体流,然后合成一个 CIF 格式的编码媒体流,最后输入通信网络中,这样,四画面合成设备需要先对视频编码数据进行一定程度的解码,然后合成,再进行完全编码。合成设备的功能模块组成如图 3 所示。

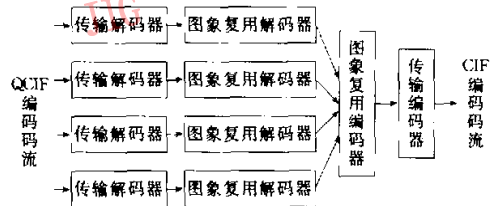


图 3 四画面合成设备方框图

传输编码器的主要功能是插入 BCH 前向纠错码,以便解码器能够检测和纠正错误的码字,另外,传输编码器为了能够满足在每个有效时钟周期内,电路交换网络(SCN)都必须有视频数据,其还必须使用纠错帧块中的填充比特来实现。

传输解码器的功能是执行如下与传输编码器相反的操作:一是检测和纠正错误码字,并且对纠错帧结构解码,剔除掉 BCH 纠错码字以及定帧的辅助数据;二是丢掉填充比特。

图象复用编码器的任务是将每帧图象数据编排成 4 个层次的数据结构,以便在各层次中插入必要的辅助数据,图象复用解码器则执行与图象复用编码器相反的操作,即对具有 4 个层次的图象复用数

据结构解码。由于四画面合成是在压缩域内进行，并且只在图象层进行处理，因此，就保留了原来宏块和像块层的图象数据，并只需要解码到块组层结构。

4 路 QCIF 格式图象的编码码流首先经过传输解码器来解除纠误帧结构，进而获得视频流的复用编码数据，再进入图象复用解码器，只要解开到块组层就可以了，这是因为从块组层以下的各个层次，CIF 和 QCIF 两种格式是完全相同的；然后把分析出的 QCIF 块组送入 CIF 的图象复用编码器进行 CIF 图象的合成，进而送入传输编码器进行纠误成帧，再插入 BCH 纠误码，最后就生成了 CIF 格式的编码码流，并被送入传输网络中。

1.3 四画面合成的算法实现

从四画面合成的系统结构可以看出，用算法实现编解码有如下两个过程：一是传输编码器和传输解码器对纠误帧结构的解析；二是图象复用编码器和图象复用解码器对图象的 4 层复用编码数据结构进行处理。纠误帧结构如图 4 所示。处理时，传输编码器则根据纠误帧序列 $S_1, S_2 \dots S_7, S_8$ 来解除纠误帧结构，进而得到完整的纯粹的视频复用编码数据，而传输编码器则按该结构实施纠误帧的结构封装。传输次序 $\rightarrow (S_1, S_2 \dots S_7, S_8) = (00011011)$

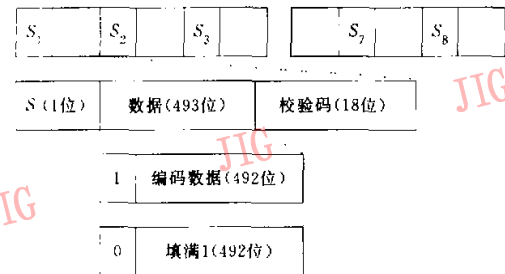


图4 纠误帧的安排

所谓图象复用编码就是把各个层次的数据按一定方式连接起来，构成一帧数据流。一帧数据的安排方式如图 5 所示。

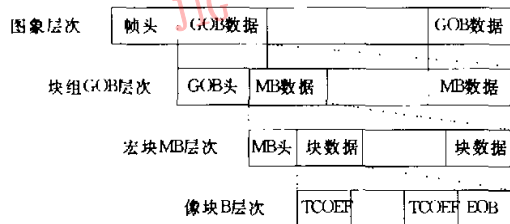


图5 图象复用编码

当传输解码器把网络送来的编码流解成复用视

频数据结构后，四画面合成则需把 QCIF 图象的压缩域解除到块组层，而只保留第 1 路 QCIF 视频流的图象开始码 (PSC)，并把第 1 路 QCIF 中图象层帧头的其他域也拷贝入合成 CIF 视频流中。在这些域中，首先要修改一位，即类型信息域 (PTYPE) 的第 4 位，这样图象格式标示只需把“0”去掉，由“0”值改变成“1”，即由 QCIF 格式变成 CIF 格式；然后定位匹配各路 QCIF 视频流中的块组标示头，也即定位各个块组，接着修改各个 QCIF 中相应块组的组号，再把各个块组放到 CIF 格式图象中，这样一个 CIF 格式的复用编码数据结构就被合成了。其四画面合成中，复用结构操作示意图如图 6 所示。

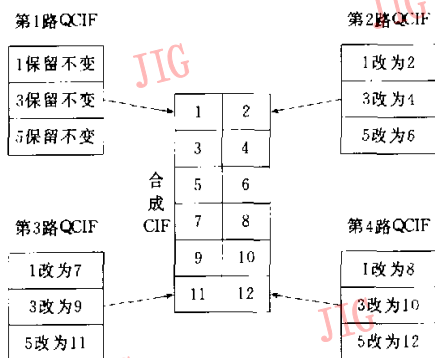


图6 四画面合成中图象复用结构操作的示意图

这里，合成实现的一个重要问题是如何处理以不同速率送来的 QCIF 数据流。其解决办法有以最快速率为准和以最慢速率为准两种方案。对于多路数据，在合成之前被放在一个缓冲器中，如果按照最快速率为准的话，只要其中一路数据被更新，就使用其他路原来的数据进行合成处理；如果按照最慢速率为准，则让多路数据都被更新之后，才进行合成处理。对于后一种，需要同步慢速通路，因为这样会使整个系统的视频合成帧速降低，而对前一种，由于快速数据仍然可保持原来的帧速，即不会降低整个系统的视频合成帧速，因此合成处理以最快速率数据为准。

合成实现的另一个重要问题是，当有多路图象数据的时候，如何进行被合成图象源的选择，本算法是通过终端的语音激活或申请发言来选择最近发言的 4 个图象源进行合成。如前所述，合成是在中间设备上实现的，因为该中间设备可同时处理 H. 323 或 H. 320 的信令信号和媒体流信号。这样，当合成设备接受到终端将要发言的信号后，就用其替代当前 4 个图象源中最长时间没有发言者，而加入到合成图象中。

1.4 四画面合成的算法流程

该算法有如下两个主要流程:一个是传输解码器和传输编码器解除和生成纠错帧的流程,另一个是图象复用编码器和图象复用解码器解除和生成复用编码结构所用的流程。

影响流程的性能有如下两个关键因素:一是对位的操作,另一个是匹配操作。从前面的讨论可知,算法的操作对象是以位为单位,而在一般的高级语言,如 Visual C++、Visual Basic,则没有将“位”作为数据类型,由于对位操作比较烦琐,因此,为了提高效率,在高级语言中是通过嵌入汇编来专门实现对位的操作,也就是说,对于上层的用户界面可用高级语言来实现;由于匹配操作是算法的主要操作,如图象开始码(PSC)和块组开始码(GBSC)的匹配定位以及纠错帧的定位等,因此匹配效率的提高就对整个算法有较大优化。为了提高匹配效率,可以采用 Kunth 与 Pratt 和 Morris 开发的 KMP 算法,这是模式匹配的一种改进算法,该算法在文献[4]中有详细讨论。

2 压缩域内包交换网络下的 H. 261 四画面合成

ITU-T 出台了 H. 323 协议,大大促进了基于包交换的 IP 网络的多媒体通信业务。因为包交换网络与电路交换网络的通信方式不同,其势必要求通信业务的实现机制发生改变。在 SCN 下,视频、音频和数据被复合成到 H. 221 帧结构里进行传输,媒体流之间是紧密聚合的,它们被安排在 H. 221 帧结构里相应的位置;而在 IP 网络通信环境下,视频、音频以及数据等媒体流是通过 H. 225 协议的 RTP^[5]来传输,由于媒体流之间在传输中是松散的,不像在 SCN 环境下那样,被严格地捆绑在一个通信信道上,因此,在 IP 多媒体通信业务下,多个不同的源端可以通过多个信道同时向一个目标端传输媒体流。这样在 IP 网络环境下实现四画面合成,即把几路视频信号复合成到一路视频信号中进行传输,就没有像在 SCN 网络下那样必须,因为这种环境下可以先在目标端为每个源端分配一个 IP 通信端口,然后使用多线程机制来同时接受多个源端的视频流,再施用软解压进行显示,就能在一个屏幕上看到多个源端的画面了。

但是,如 H. 323 视频会议等包交换网络的多媒体业务,一般采用集中控制机制,如果通过多个终端同时向某一终端发送视频流来完成四画面合成功

能,则实现集中控制功能不便利,并且,对于多点对多点的多媒体通信业务,若使用上述方案,则需要每个通信终端都给其他终端传输视频媒体流,这将要消耗较多的网络带宽资源,由此可见,在目前带宽资源紧张的现状下,这种做法不适宜。另外,上述方案必须对视频图象进行软解压,因为只有软解压才能方便地使用多线程机制,这样不仅要求终端的性能较高,并且由于软解压的图象质量没有硬解压高,而对于需要高质量图象的多媒体业务来说,仍需硬解压,所以,在 SCN 下的四画面合成算法同样适合于基于包交换的 IP 网络,还可以做如下简化。

在 IP 网络下,由于是基于包交换的,没有如 SCN 下有效时钟周期的概念,即不仅可实现有数据就发送,没有就不发送,而且可以实施每个数据包只传输一帧图象数据,另外,还有 RTP 和 RSVP^[6]协议提供业务质量保证,因此可以取消掉传输解码器和传输编码器,这种基于 H. 261 的四画面合成系统示意如图 7 所示。

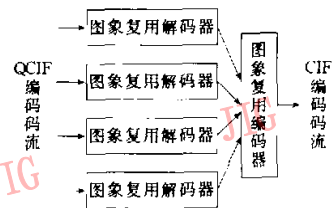


图7 IP网络下的四画面合成系统方框图

在基于包交换的 IP 通信网络环境下,进行四画面合成过程就简化了,其不需要传输解码器和传输编码器的算法,而只保留图象复用解码器和图象复用编码器的算法即可。

3 实验结果及分析

本算法的实验是在 Windows 98 环境下,使用 VC++ 和汇编语言开发,在一台主频为 733MHz、内存为 128M 的 PC 机上进行的。利用上面所介绍的原理和算法分别开发了在包交换和在电路交换下的四画面合成软件,实验获得的四画面合成结果如图 8 所示。西安邮电学院通信技术研究所研发的 IP 视频会议系统 MZH21,就采用包交换网络下的 H. 261 算法实现了压缩域内的四画面合成。

另外,本文还利用传统的方法进行了四画面合成实验,它使用 5 块 Zydacron 的编解码卡 Z350 和



图 8 H. 261 压缩域内的
四画面合成实验结果



图 9 H. 261 传统的
四画面合成结果

一个模拟的四画面合成器, 图象合成结果如图 9 所示。由图 8、图 9 可见, 如果只凭视觉已分辨不出二者在图象质量上有什么差别。使用传统方法和 H. 261 压缩域算法的性能指标比较如表 2 所示。

表 2 合成算法的技术指标

合成方法	编解码器	最大帧速(帧/s)	时延(ms)	保真率
传统的模拟合成	5 个	30	650	0.9
电路交换下的 压缩域合成	0 个	20	400~550	1
包交换下的 压缩域合成	0 个	20	380~530	1

从表 2 可见, 压缩域合成算法的成本大大低于传统的模拟合成。由于传统的模拟合成主要用硬件来实现, 所以对合成的帧速基本没有限制, 只跟源帧速有关; 而压缩域合成主要由软件实现, 它随着源帧速的提高和处理任务增加, 所耗的 CPU 和内存资源也将增大, 因此对最大处理帧速有限制。对于合成时延, 传统的模拟合成需要完全解码和编码, 还要进行数据域和模拟域的互相转变, 所以时延要长些, 又由于用硬件实现, 所以时延是恒定的; 对于压缩域内合成, 由于只进行不完全编解码, 所以时延要短些, 并且由于是用软件实现, 因此受源帧速的影响, 其时延有一定的范围。对于图象合成处理的前后保真率, 由于传统的模拟合成需要通过将图象转变成模拟图象来实现, 故图象质量有一定的下降, 然而对于压缩域的合成处理由于没有改变原来图象数据, 所以图象质量没有下降。

由此可见, 压缩域中的合成算法虽然在源帧速率不高的情况下, 使用较高性能的主机(比实验所用的主机性能更佳为好)能够得到满意的效果, 例如远程教育和桌面型的视频会议, 但在功能要求较高, 运动场较大的情况下, 传统算法仍然具有较强的优势, 例如远程医疗和足球等运动项目的转播。

对于 H. 261 压缩域内的合成算法改进的方案为: 可以综合传统算法以及压缩域中算法的各自优点, 采用软硬件相结合的方法来提高效率, 例如可以使用压缩域的合成算法开发一个专用板卡插入主机

内, 这样尽管在实现成本上没有多大的变化, 可是在算法性能上却会有提高。

4 结 论

本文提出了一种基于 H. 261 压缩域的四画面合成算法, 并且把合成原理和合成算法应用于实践, 实验结果表明, 合成图象的质量以及合成处理的时延都十分满意。与传统的模拟域中四画面处理方法比较, 不仅成本降低, 且时延和保真率等指标提高, 至于最大帧速的限制, 一方面随着 PC 处理速度提高会有改善, 另一方面对于如视频会议, 远程教学等运动视频不高的业务也能完全适用。

根据通信环境的不同, 基于 H. 261 的四画面合成采用的具体算法也有些差异。由于 H. 320 系统要求各种媒体流被复合到 H. 221 帧结构中, 而 H. 323 系统则使用 RTP 来松散传输各种媒体流, 所以, H. 320 系统下的四画面合成比在 H. 323 下的合成相对要复杂一些。

参 考 文 献

- 1 ITU-T Recommendation H. 261, Video codec for audiovisual services at $p \times 64$ kbit/s[S], 1993.
- 2 ITU-T Recommendation H. 320, Narrow based visual telephone systems and terminal equipment[S], 1997.
- 3 ITU-T Recommendation H. 323, Visual telephone systems and equipment for local area networks which provide a non-guaranteed quality of service[S], 1999.
- 4 严蔚敏, 吴伟民, 数据结构[M]. 北京: 清华大学出版社, 1992.
- 5 ietf-avt-rtp-05, JCCITT, RTP: A transport protocol for real-time application[S]. Internet Draft, draft-ietf-avt-rtp-05.
- 6 Braden R, Zhang L, Berson S et al. Resource reservation protocol (RSVP)-version 1 functional specification[S]. IETF, RFC-2205, September 1997.



黄廷学 1973 年生, 现为西北工业大学博士研究生, 现主要研究 IP 网络上的多媒体通信, 包括新电话体系研究, 统一消息系统以及视频和音频等媒体处理, 发表论文 10 余篇。

朱志祥 1959 年生, 工学博士, 教授, 现任西安邮电学院通信技术研究所所长。研究兴趣主要在中多媒体通信领域, 包括基于 ISDN, DDN 以及 IP 网络的视频会议, 新电话体系, 统一消息系统以及嵌入式通信系统。

戴冠中 1937 年生, 西北工业大学教授, 博士生导师, 研究领域为大系统估计与控制, 智能控制, 控制系统中的并行处理理论、算法与并行计算机, 网络信息安全, 虚拟现实和多媒体通信。