

多媒体会议系统中视频图像的组合*

王 朴

(国防科技大学计算机系 长沙 410073)

摘 要 在多媒体会议系统中,同时显示多方会议参加者的动态图像,是提高现有的多媒体会议系统性能的重要方向之一。本文全面地讨论了多个视频图像的组合方式、策略和方法,提出了视频图像重迭级的概念和按视频图像重迭级组合图像的算法,从网络传输数据量和图像组合工作量两方面,分析比较了各种图像组合方式的性能。

关键词 多媒体会议系统, 视频图像, 图像重迭, 图像组合

分类号 TP391. 41

Composition of Video Image in Multimedia Conference

Wang Pu

(Department of Computer, NUDT, Changsha, 410073)

Abstract In multipart multimedia conferences, each participant may desire to view all of the other participants and other contributed images and video streams or subset of them. And so video composing is an important research topic. This paper discusses combining modes and combining strategies and combining methods of video images. Combining algorithms of video images are given based on relay level of video images. The performance of three combining modes is analysed and compared as well.

Key words multimedia conference system, video image, image relay, image composition

多媒体会议系统为与会人员提供了良好的视频信息、音频信息和文本信息的交互手段,具有广泛的应用前景。但是,在现有的诸多多媒体会议系统和产品中,音频和视频的交互大都只提供一个音频和一个视频图像,这就与实际的面面对面的会议实况有所差别。实现同时可以听到多个声音,看到多个视频图像的多媒体会议系统,已成为许多研制工作者的一个追求目标。文献[1]实现了一个动态图像和四个静态图像的组合,本文比较全面地讨论了多媒体会议系统中非压缩的多视频图像的组合问题,分析了几种不同的组合方式和策略,提出了有关的组合算法,分析了不同组合方式的性能。对于压缩的视频图像组合,必须首先对视频图像进行解压缩,然后才能利用上面讨论的各种方法。对于帧内压缩的图像,利用解压缩软件可以边接收边交替地解出各个视频图像帧,这要求服务器的性能很高,而且视频图像采样频率较低。对于帧间压缩的图像,必须采用多个解压缩硬件分别对每个视频图像解压缩后,才能组合。这种方法十分复杂,代价也很高,要有特殊的硬件支持^[2,3]。

1 多视频图像的组合方式

根据组合图像的个数、组合图像中被组合图像的类型、个数和大小,多视频图像的组合方式有三种:固定方式、全动态方式和半动态方式。

* 1997年9月15日收稿

第一作者:王朴,男,1938年生,教授

1.1 固定方式

在这种方式下，根据视频选择模型选择几个视频图像组合成一个图像。各个视频图像的大小与相对位置不变，多个与会者所看到的是同一个组合视频图像（图1）。这种方式虽然不太灵活，但是实现简单，适用于任意点数的多媒体会议系统。

1.2 半动态方式

在这种方式下，各个与会者可以选择他（她）感兴趣的视频图像进行组合，但是各个视频图像的大小和被组合的视频图像个数不变（图2）。这种方式也很灵活，但复杂性介于固定方式和全动态方式之间，在点数较多的多媒体会议系统中也可采用。

1.3 全动态方式

在这种方式下，各个客户机根据自己的需要选择视频图像进行组合，因而组合的视频图像类型、个数、各个视频图像的大小和相对位置都可由各个与会者任意指定。各个与会者所看到的视频图像都不相同（图3）。这种方式非常灵活，但是，实现十分复杂，只适用于点数不多的多媒体会议系统。

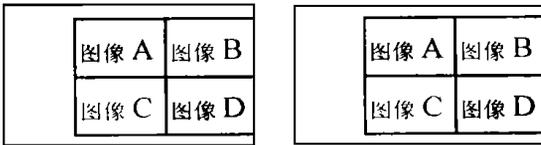


图1 固定方式

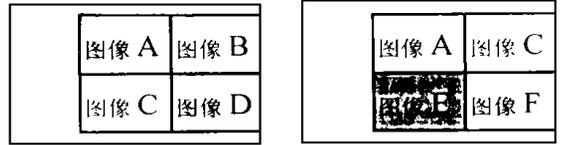


图2 半动态方式

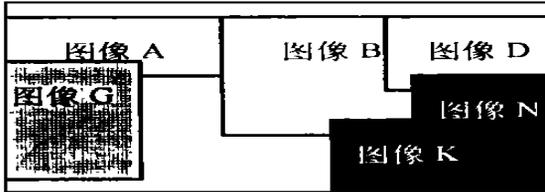


图3 全动态方式：各个客户屏幕上的视频图像个数和组合视频图像都不同

2 多视频图像的组合策略

一个多媒体会议系统一般由多个客户机和1个或多个服务器（或多点控制器）组成，多视频图像的组合，既可以在各个客户机上实现，也可以在服务器上实现。根据在各客户机上实现还是在服务器上实现，有三种处理策略。

2.1 集中处理策略

在这种策略下，由中央服务器实现视频图像的组合。这种策略能够很好地处理固定方式组合，视频图像数据传输量比分布方式低。但是，在处理半动态方式组合和全动态方式组合时，中央服务器要实现所有客户机的视频图像的组合，当会议的点数增多时，其性能要求很高。

2.2 分布处理策略

在这种策略下，视频图像的组合在各个客户机上实现。采用固定组合方式可用这种策略。采用半动态和动态方式用此策略将带来以下问题：

- (1) 由于客户机方选择显示的视频图像种类和数量事先不知道，会议所有各方的视频图像都必须传送到每一个客户方，图像数据传输量很大。
- (2) 每个客户机都要处理视频图像的组合问题，对客户机的性能要求较高。

2.3 混合处理策略

在这种策略下，将选择视频图像种类的工作交由中央服务器完成，从而减小视频数据传输量。将视频图像的组合工作交由客户机完成。这种方式适合于处理半动态和动态两种组合方式（表1）。

表 1

	固定方式	半动态方式	动态方式
集中策略	很好	好	不好
分布策略	好	不好	不好
混合策略		很好	

3 视频图像的组合方法

视频图像数据一般存放在帧存储器中, 存放的方式对应于视频图像在屏幕的显示方式。视频图像的组合可以归结为帧存储器中视频数据的组合。视频图像的组合可以采用窗口重叠方法, 也可以采用色键控制方法。采用窗口重叠方法容易在帧存储器中实现, 所以, 下面只讨论窗口重叠方法。

设客户 i 用于显示视频图像组合的窗口坐标为 $\{(x_1(i), y_1(i)), (x_2(i), y_2(i))\}$, 其中 $(x_1(i), y_1(i))$ 为窗口左上角的坐标, $(x_2(i), y_2(i))$ 为窗口右下角的坐标。该窗口坐标对应于帧存储器 VRAM $(0, 0) \sim \text{VRAM}(x_2(i) - x_1(i), y_2(i) - y_1(i))$ 的矩形区。假定视频图像 A, B, C 分别在组合窗口内的窗口 a, b, c 中显示, 窗口 a, b, c 的坐标分别为 $[(x_1(a), y_1(a)), (x_2(a), y_2(a))]$, $[(x_1(b), y_1(b)), (x_2(b), y_2(b))]$, $[(x_1(c), y_1(c)), (x_2(c), y_2(c))]$, 则其对应的视频图像 A, B, C 的数据分别存放在 VRAM $[(x_1(a) - x_1(i), (y_1(a) - y_1(i))) \sim \text{VRAM}[(x_2(a) - x_1(i), (y_2(a) - y_1(i))]$, VRAM $[(x_1(b) - x_1(i), (y_1(b) - y_1(i))) \sim \text{VRAM}[(x_2(b) - x_1(i), (y_2(b) - y_1(i))]$ 和 VRAM $[(x_1(c) - x_1(i), (y_1(c) - y_1(i))) \sim \text{VRAM}[(x_2(c) - x_1(i), (y_2(c) - y_1(i))]$ 区中。若视频图像 B 重叠视频图像 A 的一部分, 则 VRAM 中的重叠区域的数据应为视频图像 B 的数据。为了保证重叠区数据的正确性, 下面引入重叠和重叠级的概念。

定义 1 重叠: 两个图像重叠, 当且仅当一个图像覆盖另一个图像的一部分或全部时。

定义 2 重叠级: 当两个图像重叠时, 重叠图像的重叠级等于被重叠图像的重叠级加 1。当一个图像重叠多个图像时, 其重叠级等于被重叠图像中的最大重叠级加 1。例如, 图像 C 同时重叠图像 A 和 B , 其中 A 的重叠级为 0, B 的重叠级为 1, 则 C 的重叠级为 2。重叠级为 0 的图像是最底层的图像。

算法 图像重叠算法

设客户 i 希望在他的屏幕上看到 $n+1$ 个视频图像, 各个视频图像的重叠级分别为 $d_0, d_1, d_2, d_3, \dots, d_n$, 其对应的视频图像的窗口坐标为 $[(x_1(K), y_1(K)), (x_2(K), y_2(K))]$, $K=0, 1, 2, 3, \dots, n$ 。在存储器中设置一个缓冲区作为帧存储器 VRAM, 其大小等于组合图像的大小, 各客户机都按帧采集和传送。

算法 1 VRAM 重叠写算法

- (1) 按重叠级递增顺序排列, 重叠级相同的可任意排列;
- (2) FOR $K=0, n, 1$ DO

{ 将视频图像 K 的数据写入 VRAM $[x_1(K), y_1(K)] \sim \text{VRAM}[x_2(K), y_2(K)]$ }

本算法利用图像的重叠级控制帧存储 VRAM 的写入, 先写底层被重叠的图像数据, 后写上层重叠图像的数据, 使上层图像的数据覆盖其底层图像的数据, 从而保证视频图像组合的正确性。

算法 2 VRAM 只写一次算法

- (1) 按重叠级递减顺序排列, 重叠级相同的可任意排列;
- (2) 定义一个一位数组 $F(x, y)$, 其中 $x = \max[x_2(K)]$, $y = \max[y_2(K)]$, $K=0, 1, 2, 3, \dots, n$;
- (3) 令 $F(x, y)$ 的初值为 0;
- (4) FOR $K=0, n, 1$ DO
 - { $l_y = y_1(K)$;

```

 $m_y = y^2 (K);$ 
 $l_x = x_1 (K);$ 
 $m_x = x_2 (K);$ 
FOR  $I = l_y, \quad m_y, \quad 1$  DO
  FOR  $J = l_x, \quad m_x, \quad 1$  DO
    IF  $F (I, J) = 0$ 
      THEN { 将视频图像  $K$  的一个像素数据写入 VRAM ( $I, J$ );
               $F (I, J) = 1$ ; }

```

算法1用于固定组合方式较好, 算法2用于半动态或全动态方式较好。

当采用固定或半动态组合方式时, 被组合的视频图像的大小事先可以算出, 由客户机按要求采集即可。在采用全动态方式时, 为了减小图像数据传输量, 一般传送较小的图像, 所以必须在组合时对每个视频图像按各个客户机的要求进行缩放。

假定每个客户机采集的视频图像都是一样大小, 设为 $x \times y$, 在中央服务器设置一组 BUFF (x, y), 视频图像 K 的水平缩放率为 M_k , 垂直缩放率为 N_k , 考虑实际使用情况和实现简单起见, 限定 M_k (N_k) 只取下列值:

4, 3, 2, 1.75, 1.5, 1.25, 0.75, 0.5, 0.25

同时限定图像放大或缩小是指同时在水平和垂直两个方向放大或缩小。这样, 在读取图像时, 可根据 M_k (N_k) 的值去掉或增加某些列 (行) 的数据。具体算法如下:

算法3 图像缩放算法

(1) 将图像 K 的数据写入 BUFF $_k$ (x, y);

(2) 设视频图像 K 写入 VRAM 的起始坐标为 $x^0 \times y^0$;

(3) 根据 M_k (N_k) 的值选择 BUFF $_k$ (x, y) 中的行 (列) 数据送到 VRAM:

如果 $M_k > 1, N_k > 1$: 每2行 (列) 之间利用内插法增加 M_k 行 (N_k 列);

如果 $M_k = 1.75, N_k = 1.75$: 每4行 (列) 一组中, 在第1、2行 (列)、2、3行 (列) 和3、4行 (列) 之间利用内插法增加1行 (列);

如果 $M_k = 1.5, N_k = 1.5$: 每4行 (列) 一组中, 在第1、2行 (列) 和3、4行 (列) 之间利用内插法增加1行 (列);

如果 $M_k = 1.25, N_k = 1.25$: 每4行 (列) 一组中, 在第1、2行 (列) 之间利用内插法增加1行 (列);

如果 $M_k = 0.75, N_k = 0.75$: 每4行 (列) 一组中, 不送第4行 (列);

如果 $M_k = 0.5, N_k = 0.5$: 每4行 (列) 一组中, 不送第2行 (列) 和第4行 (列);

如果 $M_k = 0.25, N_k = 0.25$: 每4行 (列) 一组中, 只送第1行 (列);

4 性能分析

设客户机数为 N , 组合的视频图像为 $x \times y$, 每帧组合视频图像的数据量为 D_h , 组合的客户机视频图像个数为 m , 客户机 k 的视频图像大小为 $X_k \times Y_k$, 每帧数据量为 D_k 。

在固定方式下, m 个客户机的视频图像大小一样, $X_k \times Y_k = (X \times Y) / m$, 每个客户机的组合视频图像的内容和大小一样, 从服务器传送到客户机的图像数据量为 D_h , 从 m 个客户机传送到服务器的图像数据量为 $m \times D_k = D_h$ 。每帧图像总的网络传输数据量为 $2D_h$ 。由于只需组合一个视频图像, 图像组合工作量 W_1 最小。

在半动态方式方式下, 最坏情况下, 每个客户机都需要采集图像, 要为每个客户机组合不同的视频图像。因此, 从 N 个客户机传送到服务器的图像数据量为 $N \times D_k = N \times D_h / m$ 。从服务器传送到客户机的图像数据量为 $N \times D_h$, 每帧图像总的网络传输数据量为 $N \times (1 + 1/m) \times D_h$ 。组合图像的工作量为 $W_2 = N \times W_1$ 。

在全动态方式下, 每个客户机都需要采集图像, 其图像大小 $X_k \times Y_k$ $X \times Y$ 。服务器要为每个客户机组不同的视频图像, 在组合过程中还可能需要对图像进行缩放。因此, 从 N 个客户机传送到服务器的图像数据量 $N \times D_h$, 从服务器传送到客户机的图像数据量为 $N \times D_h$, 每帧图像总的网络传输数据量 $2 \times N \times D_h$, 组合图像的工作量 $W_3 = N \times (W_1 + W_s)$, 其中 W_s 为一个图像缩放的工作量。

从上面的分析可以得出以下结论: 采用固定方式, 组合的视频图像单一, 不够灵活, 但是网络数据传输量和组合工作量都最小, 容易实现。采用动态方式非常灵活, 但是网络传输数据量和组合工作量都很大, 没有专门的硬件支持, 是很难实现的。采用半动态方式, 其灵活性和网络传输数据量、组合工作量都介于固定方式和动态方式之间。

参考文献

- 1 王 朴, 朱海滨, 沈清等. 多媒体会议与合著系统 M CCS. 中国图像图形学报, 1996, 1 (5, 6)
- 2 Boyer D G, Lukacs M E. The Personal Presence as a system—A Wide Area Network Resource for the Real Time Composition of Multipoint Multimedia Communications. Proceeding ACM Multimedia'94, 1994, 10
- 3 Lukacs M, Bellecore. The Personal Presence System—Hardware Architecture. Proceeding ACM Multimedia'94, 1994, 10