不确定时延环境的图像序列预测理论

张　珩　周后权　许　滨　潘力均　(中国科学院力学所,100080)

摘要　该文提出的"不确定时延环境的图像序列预测理论"从图像通信系统目端的角度出发,可以将接收端的不连续序列图像构造成连续的视频数据。即使图像源端发送出帧速率小于 24 帧每秒的图像序列(如每秒 15 帧),在复现端仍然能够看到连续的视频图像。

关键词　图像序列　图像预测　不确定时延　运动估算　图像通信

中图分类号　TN911.73　　**文献标识码**　A

1　前言

随着图像通信业务的迅猛增长,带宽问题成了一个制约图像通信发展的主要因素。为了解决通信的带宽问题,常见的方法主要从以下两个方面考虑:①在图像被传输之前,采取各种各样的压缩编码方法对源图像数据进行处理,用尽可能少的比特数来表示源图像,以减少被传输的数据量;②不断应用各种新技术来提高传输系统的数据传输率。图像通信系统包括图像采集、图像预处理、图像信源编解码、信道编解码、调制解调、图像显示等环节。其实,整个系统可以简化图像源、数据传输和系统目端为三个部分。上面提到的两种方法分别从图像源和数据传输方面着手的。

本文试图从"系统目端"的角度出发,探索一种解决该问题的新技术,即不确定时延环境的图像序列预测技术,在图像显示端对接收到的图像数据处理,进一步减少图像数据的传输量,缓解带宽问题。

2　不确定时延环境的图像序列预测技术

不确定时延环境图像序列预测的主要思想可以简述如下。一个连续视频是由时间上连续的图像帧 1、2、3、4、5、6、7、8、9 ……构成的,并且其帧速率不能小于 24 帧每秒。而由时间上不连续的帧图像 1、3、5、7、9 ……构成的视频则是不连续的,给人的感觉是动画效果,图像内容是跳跃运动的,在这里称之为非连续视频,或者非连续图像序列。该预测理论正是以这种非连续视频为研究对象,运用有效的预测技术将非连续视频恢复成连续视频。显然,传输非连续视频所必需的数据量要少于连续视频,能够起到减少数据量的作用。在预测的过程中,有两种不同情况。以非连续视频 1、3、5、7、9 ……为例,一种情况是对有已知信息可以利用的帧图像进行预测,例如 1'、3'、5'、7' 等;另一种是在没有任何已知信息可以利用的情况下预测出帧图像,例如 2'、4'、6'、8' 等,也就是说根据所接收到的帧图像——参考图像——来"构造"出这些预测图像。所有的这些预测图像帧按时间顺序排列就可以构成时间上连续的预测视频 1'、2'、3'、4'、5'、6'、7'、8'、9' ……,从而将非连续视频恢复成连续视频。由于连续视频中图像内容的复杂程度和变化程度的不同,因此非连续视频中的帧图像也是有变化的,可以由 1、3、5、7、9 ……构成,也可以 1、4、7、8、10、12 ……构成。从这两个非连续图像序列可以看出,"1 和 3"与"1 和 4"之间的时间间隔是不相同的。如果时间上相邻的源图像之间内容变化很小,那么传送的参考图像之间的时间间隔就可以适当地增大;如果源图像间内容差别很大,则应当缩短参考图像之间的时间间隔,以避免在系统目端无法构造出预测图像。在极端条件下可以考虑按连续视频传送,不跳过任何源图像帧。在系统目端(图像显示端)看来,这种时间间隔表现为非连续视频中帧图像接收过程中的时间延迟,时间间隔的可变性则体现为时延的不确定性。不确定时延环境图像序列预测就是针对不确定时延环境,基于所接收到的帧图像来预测出不同数目的预测图像帧——指那些没有已知信息可以利用的预测图像帧,最终将非连续视频恢复成连续的预测视频。

与 MPEG 标准或 H261 标准的图像序列预测不同之处在于,不确定时延环境图像序列预测中有一部分预测图像仅仅依据现有的参考图像信息,在没有任何信息可以利用的情况下预测得到的。MPEG 标准或 H261 标准的预测算法实际上是用尽可能少的比特数来表达源图像,然后在解码端复现其预测图像(标准中的 I 图像帧或 P 图像帧),而不确定时延环境的图像序列预测技术在某些情况下是完全不利用预测图像的任何信息来预测出像素值,是在构造预测图像。

3　预测算法及其实现

序列图像中的图像数据是高度相关的。一帧图像内部以及视频序列中相邻图像之间有大量的冗余信息。基于序列图像的强时间相关性和强空间相关性,并且每帧图像都是由前景和背景构成的,因此在不确定时延环境图像序列预测的实现过程中,分两个阶段完成,即粗预测和精预测。粗预测先估算出图像背景的运动矢量,构造出预测图像的背景部分;而精预测操作的对象则是图像的前景部分。

粗预测是整个预测过程的第一阶段,分两步进行,即背

收稿日期:2000 - 03 - 14

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

景运动估算和背景运动补偿。背景运动估算的目的是估算出背景运动的形式和运动矢量的大小。常见的背景运动有平移(translation)、旋转(rotation)和缩放(zooming)三种形式。在视频图像中,“平移”运动是比较普遍的一种形式,也是比较容易实现和判断的一种运动形式。在本文提到的预测算法中,目前只针对“平移运动”进行处理。背景运动估算产生的结果是背景的平移矢量在 X 和 Y 方向上的分量 V_x 和 V_y,紧接着进行背景运动补偿。背景运动估算操作的对象永远是相邻的两帧参考图像,其结果直接作用于后面背景运动补偿。

如果依据三幅参考图像进行前向预测,则能够预测出物体的非匀速运动和匀速运动。例如,由三帧参考图像 S_{k-3}、S_{k-2} 和 S_{k-1} 来预测图像(下标仅表示它们的时间顺序),在背景运动估算时先估计出 S_{k-3} 和 S_{k-2} 之间的运动矢量 V_0,以及 S_{k-2} 和 S_{k-1} 之间的运动矢量 V_1。背景运动补偿时则依据 V_0 和 V_1 计算得到 S_{k-1} 和 S'_k 之间的运动矢量 V',然后将帧图像 S_{k-1} 的内容平移 V' 后构造出预测图像 S'_k 的背景部分。

运动矢量 V 可以用速度场 $[V_x \quad V_y]^T$ 表示,也可以用位移 $[d_x \quad d_y]^T$ 来表示。由于参考图像间的时间间隔比较小,因此可以用位移表示运动矢量,即 $V_0 = [d_{x0} \quad d_{y0}]^T$, $V_1 = [d_{x1} \quad d_{y1}]^T$, $V' = [d'_x \quad d'_y]^T$。假设参考图像 S_{k-3}、S_{k-2}、S_{k-1}、S'_k 之间的时间间隔分别为 t_0、t_1 和 t_2。计算参考图像与预测像之间运动矢量的预测公式(根据抛物线原理)可以描述为:

$$\begin{cases} d_{x0} = a_x t_0^2 + b_x t_0 \\ d_{x0} + d_{x1} = a_x (t_0 + t_1)^2 + b_x (t_0 + t_1) \\ d_{y0} = a_y t_0^2 + b_y t_0 \\ d_{y0} + d_{y1} = a_y (t_0 + t_1)^2 + b_y (t_0 + t_1) \end{cases} \tag{1}$$

由式(1)求出系数 a_x、a_y、b_x、b_y,然后代入下式即可得到 d'_x 和 d'_y:

$$\begin{cases} d'_x = a_x t_2 (t_2 + 2t_1 + 2t_0) + b_x t_2 \\ d'_y = a_y t_2 (t_2 + 2t_1 + 2t_0) + b_y t_2 \end{cases} \tag{2}$$

如果采用的是双向预测法,即由原始图像 S_{k-1}、S_{k+1} 来预测图像 S'_k(下标仅仅表示显示顺序),并假设物体作匀速运动。因此,从 S_{k-1} 和 S_{k+1} 之间估算运动矢量 $V_1 = [d_{x1} \quad d_{y1}]^T$ 到 $V' = [d'_x \quad d'_y]^T$ 的计算要比前向预测时简单,预测公式如下:

$$\begin{cases} d_x = d_{x1} * t_1 / t_0 \\ d_y = d_1 * t_1 / t_0 \end{cases} \tag{3}$$

在式(3)中,t_1 和 t_0 分别表示 S_{k-1}、和 S_{k-1}、S_{k+1} 之间的时间间隔。

背景运动补偿时,前向预测和双向预测除了 S'_k 在计算运动矢量时使用了不同公式之外,还有一个不同点,在进行前向预测时,会遇到一个“图像新区”的问题(见图3)。图中的阴影部分即“图像新区”,前向预测时新区内的像素点无法由原始图像 S_{k-1} 平移得到。在这里,用 $S_{k-2}(x, y)$ 和 $S_{k-1}(x, y)$ 的平均值来代替 $S'_k(x, y)$。如果是双向预测,则可以有 S_{k+1} 中的像素点平移得到。平移矢量与 S'_k 到 S_{k+1} 的运动矢量方向相反,大小相等。

在估算参考图像间的运动矢量(如 V_0 和 V_1)时,采用的是基于图像块的匹配法。用到了 MPC 准则和变体 MAD 准则。其中变体 MAD 准则为:

$$MAD(V_x, V_y) = \sum_{(x,y) \in B} | S(x, y, k+1) - S(x + V_x, y + V_y, k - 1) |$$

进行块匹配计算时,为了提高匹配矢量的搜索速度,更好的利用图像序列的强时间相关和强空间相关,提出了局部振动搜索法。在匹配一个块运动矢量时,先以其相邻快的运动矢量为起点,一个一个像素地逐渐往外推。让预测矢量在一个较小的范围内振动。如果存在一个矢量能满足上述两判据之一即可以停止该数据块的匹配计算。否则,对该数据块进行全局搜索,选择变体 MAD 准则值中最小的一个矢量作为估算矢量。

图1显示了粗预测完成后的各个图像块的预测矢量。参考图像分别为图2中的图(a)和图(c)。

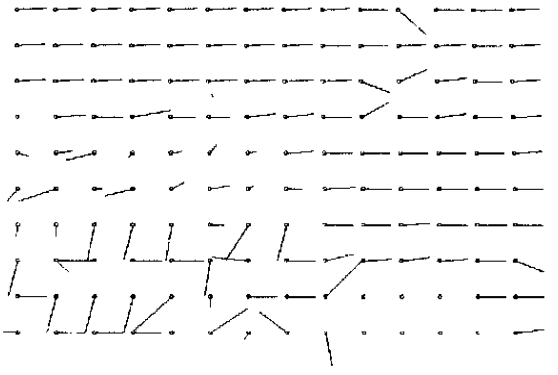

图1 粗预测后的块运动矢量图

精预测针对的是前景部分,也就是说该部分的运动与背景运动是不一致的。例如图2所示图像序列中的“字幕部分”和“正在运动的动物”。其中,面对各种不同运动形式的前景物体,需要分情况对待。某些前景物体预测产生的误差对图像是比较敏感的,称之为第二类区域;而某些物体的误差则是可以忽略的,称之为第一类区域。可以想象为这样一种情况,在同一桌面上,有一个乒乓球和一只蚂蚁都在作不同的运动。两个前景物体中,蚂蚁属于第一类区域,而乒乓球属于第二类区域。进行预测时,显然可以不管蚂蚁的运动,因为蚂蚁太小对整幅图像的影响很小。由于预测是不可能完全精确的,因此忽略了第一类区域。精预测操作的对象是第二类区域的像素点。将第二类区域从整个图像中抽出来,进行单独地预测和计算。

粗预测用的是块匹配法,精预测中运动估算时则采用了相减法,对物体轮廓线进行运动估算。精预测阶段是分以下几个步骤进行的,标识差值区、区域连通判断、求区域边界并

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

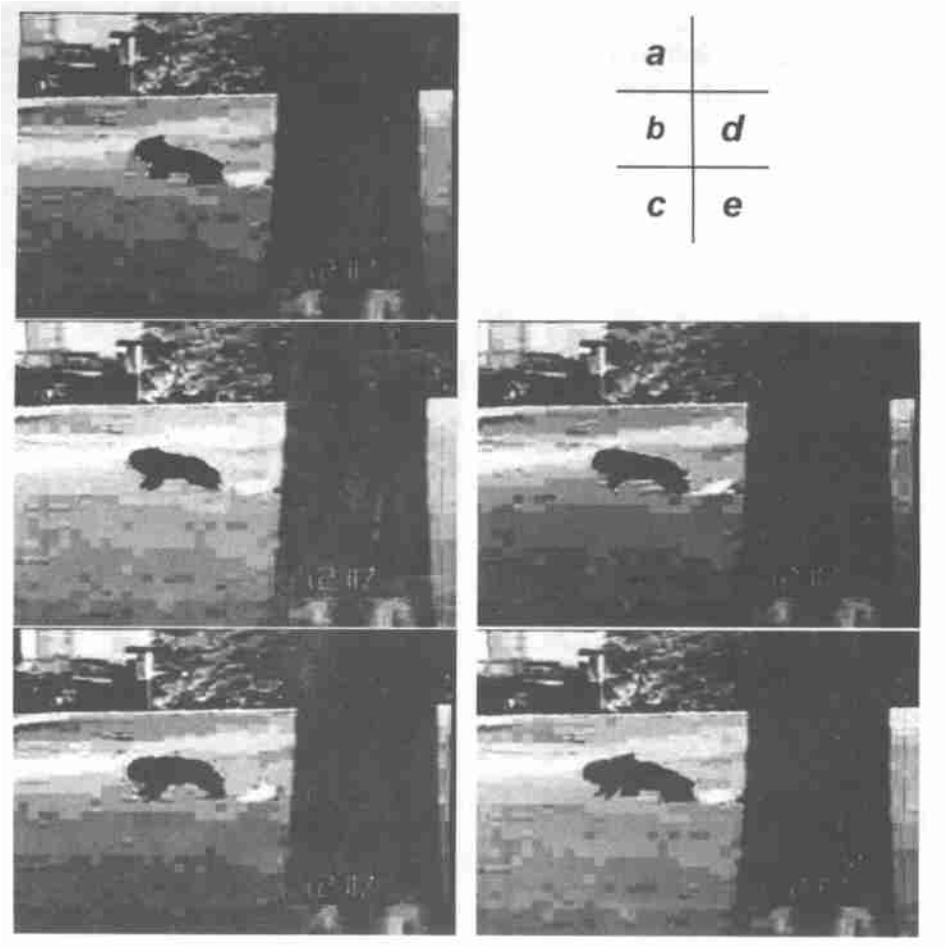

图 2　预测结果

图(a)和图(c)前后两帧原始图像;

图(b)粗预测和精预测完成后的预测图像;

图(d)原始图像序列中排在图(a)和图(b)中间的源图像

图(e)粗预测完成后的预测图像;

滤掉小区域、求分界线将区域分成两部分、对不同部分进行像素填充。首先用相减法求得连通区域,区域的边缘包含了前景物体的运动信息,是物体在前后参考图像中的轮廓线。然后,对这些区域进行处理,求出前景物体在预测图像中的轮廓线,该轮廓线把各连通区域分割成两部分。接下来要做的是如何填充被分割开的两个不同的部分,分别记为 Z1 区域和 Z2 区域。总的 策略是复制参考图像中对应部分的像素值。是复制参考图像 S_{k-1} 还是复制参考图像 S_{k+1} 呢,则依据下面的判据来决定。

$$\begin{cases} DIF1_{k-1} = \sum_{(x,y)\in Z1} | S(x,y,k-1) - S(x,y+5,k-1) | \\ DIF1_{k+1} = \sum_{(x,y)\in Z1} | S(x,y,k+1) - S(x,y+5,k+1) | \\ DIF2_{k-1} = \sum_{(x,y)\in Z2} | S(x,y,k-1) - S(x,y-5,k-1) | \\ DIF2_{k+1} = \sum_{(x,y)\in Z2} | S(x,y,k+1) - S(x,y-5,k+1) | \end{cases} \tag{4}$$

计算式(4) 中的四个数值,然后作出判断。

①对于 Z1 区域的像素点有:

$$\text{if}(\text{Dif}1_{k-1} <= \text{Dif}1_{k+1}) \quad \text{then} \quad S_k(x,y) = S_{k-1}(x,y)$$
$$\text{Else} \quad S_k(x,y) = S_{k+1}(x,y)$$

②对于 Z2 区域的像素点有:

$$\text{if}(\text{Dif}2_{k-1} <= \text{Dif}2_{k+1}) \quad \text{then} \quad S_k(x,y) = S_{k-1}(x,y)$$
$$\text{Else} \quad S_k(x,y) = S_{k+1}(x,y)$$

填充完第二类区域后,然后进行平滑处理,整个预测过程结束。其粗预测和精预测完成后的最后结果见图 2。

4　结束语

不确定时延环境图像序列预测仅作用于系统目端,其效果使在图像源部分只需对非连续视频进行处理并进行传送,降低了整个图像通信系统的数据负荷,缓解了图像通信的带宽问题。该预测算法分粗预测和精预测两个阶段来进行,符合序列图像的自然规律 ——序列图像中内容的变化部分比较

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

少。因此,将序列图像的预测分成背景和前景进行。背景的变化相当于图像的整体变化,前景是图像中与整体运动不一致的部分。由图 2 可以看出,粗预测完成后,图像的大致运动可以满足要求了。精预测后,使预测图像的质量又得到了提高。本算法在粗预测时没有对旋转和缩放等其余的运动形式进行估算。精预测算法对刚性物体的预测效果还可以。不过,目前对柔性物体的运动预测有些比较好的方法,例如光流计算法。

参考文献

[1] A Murat Tekalp. (美) 著. 崔之祜,江春,陈丽鑫译. 数字视频处理. 电子工业出版社,1998

[2] Heeger D J. "Optical Flow from Spatiotemporal Filters". In: Proceedings First International Conference on Computer Vision, London,1987: 181 - 190

[3] S Uras, F Girosi, A Verri, and V Torre. "A Computational Approach to Motion Perception", Biol. Cybern., vol. 60:79 - 97,1988

[4] Michael J. Black, and Allan D. Jepson, "Estimating Optical Flow in Segmented Images Using Variable - Order Parametric Models With Local Deformations", IEEE Trans. Patt. Anal. Nack. Intel., vol. 18, No. 10:972 - 986,1996

[5] Tretiak O, and Pastor L. "Velocity estimation from image sequences with second order differential operators". In: Proceedings of the International Conferences on Pattern Recognition, Montreal, Que,1984:16 - 19

[6] Jong - Il Park, and Choong Woong Lee. "Robust Estimation of Camera Parameters from Image Sequence for Video Composition", Signal Processing: Image Communication 9:43 - 53, 1996

[7] 吴乐南编著. 数据压缩的原理与应用. 电子工业出版社,1995

作者简介:

张　珩,男,1961 年生,博士,研究员。主要从事航天遥科学技术、微小卫星技术、自动控制及电子工程技术的研究。现任中国科学院高技术与研究发展局副局长兼中国科学院军工办公室主任,中国科学院力学研究所研究员,国家 863 航天领域遥科学及空间机器人专家组副组长、微小卫星技术专家组成员、921 - 2 系统副总指挥。

周后权,中科院力学所硕士生。

许　滨,副研究员。现任中科院力学所应用气动力学与飞行器控制研究室副主任。主要从事航天领域遥科学技术。

潘力均,助理研究员,现在中国科学院力学研究所工作。

The Image Sequence Prediction Theory Under Uncertain Delay

Zhang Heng　Zhou Houquan　Xu Bin　Pan Lijun

(Institute of Mechanics, Chinese Academy of Science,100080)

ABSTRACT　With the fast increasing of new image communication services, the problem of bandwidth restricts the application and development of image communication more and more. Considering from the terminal of a image communication system, we proposed the image sequence prediction theory under uncertain delay. The application of this theory can make a video from temporal discontinuous image sequence which has fewer than 24 image frames in a second. We still can make image communication system into really clairvoyance and clairaudient, even if image frames sent from the system source is fewer than 24, eg 15 image frames per second. The data that needs to be transfer is decreased for the lessen image frames sent from the system source, then we can reduce the necessary system transfer bandwidth and solve the problem of bandwidth shortage in a certain extent.

KEYWORDS　Image sequence　Image prediction　Uncertain delay　Motion estimation　Image communication

(上接第 48 页)

Implementation Techniques of 3D Objects Scene Graph Structure based on Virtual Reality

Wu Yun　Zhou Tingmei　Wang Zhongfan　(Wuhan Automotive Polytechnic University)

ABSTRACT　VR is a new application - driven technology which is integrated in partly by the technologies of computer science, sensoring and measurement, simulation, micro - electronics. within the area of computer science。In this paper, the author discussed the 3D object scene graph structure theory, system design and implementation techniques and application case. The authors construct automobile scene graph structure based on Virtual Reality.

KEYWORDS　Virtual reality　Scene graph structure　Automobile body

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net