

规范式边缘模型基运动图象压缩编码方法

高磊 张兆扬 张文俊

(上海大学通信与信息工程学院, 上海 200072)

摘要 在图象边缘提取与优化的基础上建立了一种新的运动图象信源模型, 并通过计算机模拟实现了编解码系统。其基本思想是: 根据主观视觉特性, 以边缘信息为表达运动图象的空间与时间特征的核心。这种模型基编码方法不需要图象的先验知识, 模型建立及方案实现较为简单。针对信源模型, 引入基于几何约束的图象匹配方法作为运动分析的基础, 并建立了广义运动概念, 提出并实现了广义运动分析, 丰富了帧间预测编码方法。

关键词 图象压缩, 模型基编码, 边缘检测, 图象匹配, 运动分析

1 引言

近几年在低码率图象通信领域^[1], 数字图象压缩方法的研究有了令人瞩目的发展, 出现了许多突破经典信源编码理论框架的新方法, 开辟了一些新的研究领域。其中模型基(Model-Based)图象压缩编码作为一种新的现代数据压缩技术是很值得研究和探讨的。其基本特征是建立更加符合实际的图象信源模型(Source Model), 方法上引入了计算机图形学、计算机视觉等学科的知识。语义基(Semantic-Based)^[2]和物体基(Object-Based)^[3]编码是两种典型的模型基编码方法。本文建立的基于边缘信息的模型基信源模型及相应的编码方案不需要图象的先验知识; 景物模型的构造简洁可靠; 不必恢复景物三维模型, 也能达到较高的压缩比。这就是基于边缘的模型基图象压缩编码。

在本文的方案中灰度分布都用多项式作规范描述, 故可称之为图象的规范式表达^[4]。与经典方法相比, 这里的信源模型不同于离散随机场, 而是由边缘信息建立的新信源模型; 编码方法也突破了传统的波形编码范畴。从广义上说, 它属于模型基编码方法。

2 基于边缘信源模型的编码原理

2.1 基于边缘的信源模型

图象边缘信息的重要性体现在两个方面: 景物中三维物体轮廓投影在二维图象平面上表现为边缘, 故边缘是理解静态图象的关键; 三维物体轮廓的空间位置变化反映了物体的运动, 故运动图象中相继帧边缘的变化体现了三维物体运动。本文的信源模型是在边缘信息基础上建立的, 称之为基于边缘的信源模型。

在此, 关于运动图象提出两点假设:

- (1) 光源是恒定非时变的;
- (2) 景物中物体的运动必然伴随着相继帧之间的边缘变化。

这里的边缘变化既包括物体整体运动产生的位置变化, 也包含物体局部运动或形变。显然, 大多数运动图象在特定时间段内是满足这两点假设的。

2.2 编码器结构描述

图1是系统的编码器框图。信号的流程大致可分为3路。第1路为初始帧的编码。由初始帧(记为 s_k)与其激励区 E_k 之差得到缓变区, 然后进行缓变

• 本项研究由国家自然科学基金(项目编号: 69572023)与上海市科技启明星计划资助

收稿日期: 1997-01-27; 收到修改稿日期: 1997-04-14

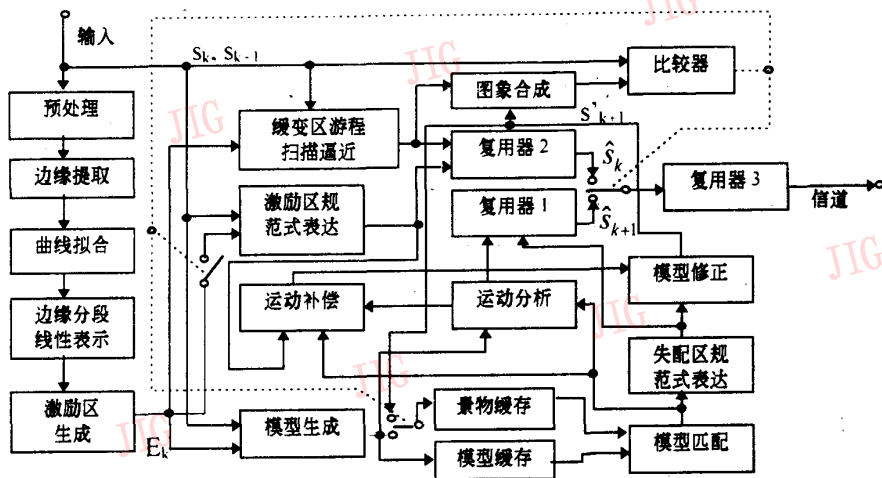


图 1 编码器框图

Fig. 1 Block diagram of the proposed coder

区游程扫描逼近。第 2 路为帧间预测编码。 S_k 与 S_{k+1} (当前帧) 经模型生成后得到对象与标记信息, 以进行模型匹配。根据模型与匹配信息进行运动分析、提取运动参数。那些小部分失配的激励区需要重新传输。第 3 路是产生使整个系统协调工作的控制信息。由匹配信息、运动参数和初始帧的激励区进行运动补偿。补偿后的激励区受失配区的规范表达修正后代替初始帧, 用来预测下一帧 (同时它也是当前帧激励区的表达, 与缓变区合成当前帧的解码图象 S'_{k+1} 所以这部分相当于解码器)。比较与 S'_{k+1} 与 S_{k+1} , 产生控制信号。当 S'_{k+1} 与 S_{k+1} 差异过大时表示必须重新传输一个初始帧。图中的 3 个受控开关协调这一工作。

3 初始帧编码与规范式

激励区的提取是在边缘检测的基础上进行的。首先对图象进行预处理 (如滤波与适当的灰度均衡) 以提高边缘检测的性能。为了进行激励区结构的聚合、分段和表达, 必须首先得到一个清晰而完整的激励区框架, 本文用 Markov 随机场^[5] (Markov-Random-Field MRF) 迭代方法来做到。激励区结构通过链码形式聚合成若干条线状结构。然后再沿着链码将这些线状结构进一步细分, 用更多的直线段来近似表达。最后激励区结构的位置是由这些直线段的端点坐标近似表达的。对于分段后得到的每段结构, 通过侧向展宽构造一个矩形拟合表达区域并用称之为规范三阶多项式来拟合该区域内的空间灰度分

布。矩形内的所有 M 个像素点给出一个三元线性超定方程组, 拟合参数的求解最后归结到求一个三元线性超定方程组的最小二乘解问题。

所有未收入激励区矩形表达区域内的像素点构成缓变区。为便于描述, 将缓变区划分为若干个任意边界形状的子块, 缓变区中的像素灰度值可以表示为图象点阵坐标的函数。缓变区子块通过游程扫描来记录和拟合。

4 帧间预测编码

本文中, 用预测编码技术进行帧间编码。帧间预测的目的在于提取相继两帧的相关性。如前所述激励区与图象边缘分段直线表示密切相关。图象中的一条边缘用一条分段折线表达。相继两帧的分段折线有对应关系。仅有少量边缘在下一帧中消失或第一次出现。基于分段折线的空间形态 (包括空间位置和折线内部各小段的几何关系) 可以找出这种对应关系, 称作模型匹配。激励区的变化主要取决于当前帧与前一帧对应边缘在几何上的差异。差异由物体的运动引起。在匹配信息的基础上, 利用运动分析技术提取运动参数可以有效地表达这种差异。

4.1 模型匹配^[6,7]

这里的算法着眼于建立边素之间的几何关系。匹配算法有点类似于两帧图象互相“滑过”, 找出最佳匹配位置。首先涉及基本概念的定义:

对象: 预测图象 (前一帧) 在分段折线中的基本段。用 a_i 表示, $1 \leq i \leq n$;

标记: 被预测图象 (当前帧) 在分段折线中的基本段。用 l_j 表示, $1 \leq j \leq n$;

景物: 对象所表示的分段折线。用集合 A 表示, $A = \{a_i | 1 \leq i \leq n\}$;

模型: 标记所表示的分段折线。用集合 L 表示, $L = \{l_j | 1 \leq j \leq n\}$ 。

其中对象与标记所含信息有空间位置和方向

(与水平方向的夹角)。

我们的兴趣在于确定 a_i 与 l_j 的匹配函数 $p(i, j)$ ($p(i, j) \in p(\{0, 1\})$); 取 1 表示匹配, 取 0 表示不匹配。匹配是基于几何约束的, 当我们确定对象 a_i 的标记为 l_j 时, 希望以此为基础, 发现另一个对象 a_k ($a_k \in A$), 其对应的标记为 l_k ($l_k \in L$)。寻找对象 a_k 是在由 i, j, k 所定义的区域中进行。将这个区域记为窗口 $w(i, j, k)$ (图 2)。每个对象 a_i 可以画成一个二维矢量 $\overrightarrow{A_i B_i}$, 每个标记 $l_j = \overrightarrow{P_j Q_j}$ 当我们由 l_k ($\overrightarrow{P_k Q_k}$) 去寻找时, 可以将 l_j “滑向” a_i , 在此过程中 l_k 扫过的区域就是 $w(i, j, k)$ 。最后定义 (i, j) 与 (n, k) 之间的相容关系“C”:

$$(i, j) \text{ 与 } (h, k) \text{ 相容} \Leftrightarrow (i, j)C(h, k) \\ \Leftrightarrow a_h \in w(i, j, k) \& a_i \in w(h, k, j)$$

相容关系通过叠代运算确定。首先根据任意两个对象与标记之间的角度来确定初始匹配函数 $p(i, j)$ 。然后在每次叠代中, 在给定 a_i 与 l_j 匹配时, 看其它对象与标记的匹配状况。如果模型 L 的 m 个标记中有 q 个与景物 A 中的对象匹配, 就认为 A 与 L 是同一边缘对应的激励区。

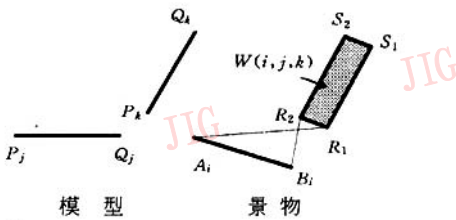


图 2 相容窗口定义
Fig. 2 Example of window design

4.2 广义运动分析

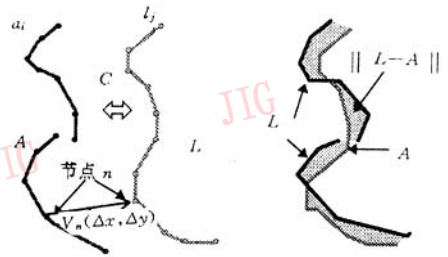
对于帧间预测, 必须更精密地描述 A 到 L 的变化, 仅根据匹配信息是不够的。提取 L 与 A 之间几何特性的差异是运动分析的过程。此处的运动分析与语义基或物体基方法中的运动分析不同, 其并不企图直接去计算位移矢量和旋转矩阵; 它也与 MPEG 标准中完全由宏块匹配所确定的二维运动矢量不同, 它所描述的是若干条平面曲线 (A 可能由几条曲线组成) 如何变化为一条曲线 L 。称之为广义运动分析 (图 3a)。

L 和 A 由分段直线来表示, 每一小段都有端点“·”, 我们称之为节点 n 。设 A 有 M 个节点、 L 有 N 个节点, 则 A 的 M 个节点的有序运动就可以表达 A 的变化。每个节点 n 联系着一个二维运动矢量 V_n

($\Delta x, \Delta y$), 用一个 M 维复向量 $V = [V_{n1}, V_{n2}, \dots, V_{nM}]$ 就可以恰当地描述 M 个节点的有序运动。如果定义了 M 维赋范复空间 C^M , 其范数“ $\| \cdot \|$ ”能反映 L 和 A 之间的几何特性差异 (比如图 3(b), 表示为 L 与 A 所夹区域的面积)。就可以提出求广义运动问题: 每个 M 维复向量 V 联系着一个范数 $\|L - A\|_v$, 找出其中一个向量 V_{min} , 使得范数 $\|L - A\|$ 最小, 即满足下式:

$$V_{min}: \|L - A\|_{v_{min}} = \min_{v \in C^M} (\|L - A\|_v)$$

这是一个多目标数学优化问题, 其求解是很困难的。我们不拟这样做, 转而求助于匹配信息。如果 $p(i, j)$ 反映的匹配信息是一一对应的, 显然很容易得到, 只需求出对应节点位置之差即可。可以将求解 V_{min} 的数学优化问题转化为如何通过 A 与 L 的重组构造使得匹配关系变为一一映射。实际的算法大致可分为 2 步: (1) 找出匹配关系中 2 个对象 (a_{i1}, a_{i2}) 对应着同一个标记 l_j 的情况, 拆分 l_j 为 (l_{j1}, l_{j2}) , 这样每个标记只对应着一个对象; (2) 找出 2 个对象标记 (l_{j1}, l_{j2}) 对应着同一个对象 a_i 的情况, 合并 (l_{j1}, l_{j2}) 为一个标记 l_j , 这样就可以变匹配关系为一一映射, 可以方便地得到 V_{min} 。实际处理中, 一对三以上的匹配极为罕见。由匹配叠代算法也可以看出这一点。但若确实出现这种情况, 可以通过简单的合并方法去除。



(a) L 与 A 几何特性关系 (b) 以面积定义范数

图 3 广义运动分析示意

Fig. 3 General motion analysis

模型匹配与广义运动分析希望相邻帧的变化 (运动) 不至于太大, 但并无绝对的限制。变化增大时只是增加模型失配区, 影响压缩率。当模型失配的比例达到一定的阈值 (比如 15%) 时, 表明必须传输一个新的初始帧。实际连续的 20 帧图象, 模型失配区一般界于 (7%~10%), 可以较好地与模型匹配和广义运动分析相适应。另外, 系统在利用 MRF 模型提

取图象边缘时,将会检出物体高光部位的形状,所以高光位置的运动也将纳入相同的模型匹配与广义运动分析处理。所不同的是,产生这部分变化机理不同于物体本身结构的形变或运动。如果高光部位的形状变化较大,系统则将其归入模型失配区处理。

5 参数编码

为简便起见,采用均匀量化、等字长编码。

缓变区是用“游程扫描—规范式拟合”的方法来表达图象灰度值的。所用信息有:子块起始坐标、游程长度、规范式系数等。在适当的图象质量下,对每个子块而言,记录起始坐标需 17bits、规范式 26bits、游程长度 1080bits。整幅图象约有 20 个缓变区子块(有任意形状),故缓变区数据量约为 $(17 + 26 + 1080) \times 20 = 22460(\text{bits}) = 2.8\text{kbytes}$ 。关于初始帧,每个激励区对应一条边缘,边缘用分段直线表示,其坐标采用 DPCM 方式编码。所用信息有:激励区首坐标、坐标变化量、规范式系数等。整幅图象约有 60 个激励区子块,故激励区数据量约为 $(17 + 64 + 184) \times 60 = 15900(\text{bits}) = 2.0\text{kbytes}$ 。

帧间预测参数的编码与初始帧类似。大量的规范式系数由预测所得,不必传输。需重新计算的规范式系数约占 8% 左右。广义运动参数的动态范围与初始帧中激励区的坐标变化量相同,但统计特性表现出明显的规律性,这里也只用了定字长编码。加上少量的匹配信息,每帧约需 0.6kbytes。

6 实验结果

实验所用为电视电话中常用的人物头肩像(Miss America)的灰度图象为 $320 \times 200\text{pels}$ 、8bits 量化、共 20 帧图象。

初始帧素图象(图 4(b))是根据文献^[8]中的算法得到。它经 MRF 随机叠代^[9]优化后(图 4(c)),收效明显:边缘清晰、定位精度提高。激励区的构造是在边缘的分段直线近似表示(图 4.(d))的基础上进行的。解码的帧间图象(图 4(f)、图 4(g))出现了稍可觉察的畸变。第 2 帧质量优于第 19 帧,这是因为它与初始帧的相关性更强。系统中后继帧直接利用初始帧的缓变区。由于激励区的运动,两部分叠加后,原被遮掩的缓变区部分应该显现,但此信息丢失了(对应于图 4(e)中的条、点状黑色区域)。可以用附近的缓变区进行补偿(图 4(f)就是补偿后的解码

图象)。该处理过程只需在解码端进行。如果遮掩部分新出现了激励区,系统将会保证重传这些新的结构信息。

影响帧间压缩比的关键因素是失配激励区数量,因为失配部分必须重新传输规范式系数,代价很高。图 5 是各帧(初始帧除外)失配激励区所占的比例,一般在 7%~9% 之间。文献^[9,10]用物体基方



(a)



(b)



(c)



(d)



(e)



(f)



(g)

(a)初始帧原图象;(b)经 MRF 叠代后得到的边缘结构;(c)分段折线逼近边缘;(d)第二帧激励区;(e)遮掩区部分;;(f)第二帧解码图象;(g)第十九帧解码图象;
 (a)Original image of initial frame;(b)Edge structure extracted by MRF iteration;(c)Edge approximated by linear segments;(d)Exciting area of 2nd frame;(e)Covered region;(f)Reconstructed image of 2nd frame;(g)Reconstructed image of 19th frame.

图 4 编码过程中的各种图象

Fig. 4 Each kind of images during the procession of coding and decoding

法对同样的 20 帧(Miss America)图象作了压缩,所选用的信源模型是“三维运动刚体”,模型生成和运动分析基本采用了 Ostermann^[3]的方法,可将每像素比特数压缩到 0.08bit/pel~0.15bits/pel 之间。以本文的方法在尚未采用较适当的熵编码情况下,每像素比特数已可达到 0.075bit/pel 左右。图 6 是两种方法压缩效率的比较。

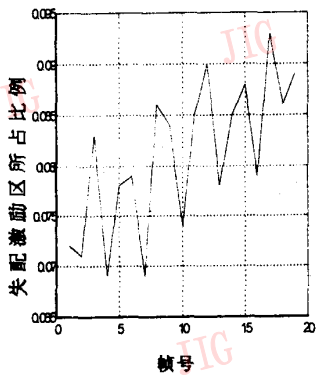


图 5 失配激励区比例

Fig. 5 Ratio of model failures

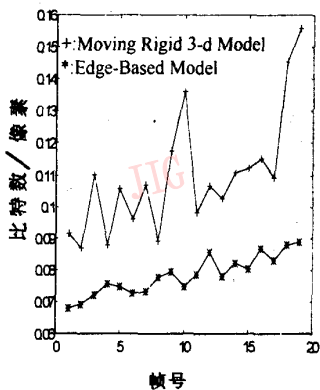


图 6 与 MR3D 信源模型的物体基压缩方法比较

Fig. 6 The proposed scheme compared to the object based coding with the source model MR3D

H. 261/3 和 MPEG-1/2 压缩后的每像素比特数一般在 0.2~0.4 bits/pel 之间。若进一步降低,则解码图象会出现明显的“块效应”和“散粒噪声”。本文中,基于边缘信息建立的信源模型按照图象本身的边缘特性对图象进行处理,而不象 H. 261/3 或 MPEG-1/2 那样硬性将图象分成 8×8 的像素块。这样,相继帧之间的运动分析与运动补偿效率更高,压缩率可提高 2.7~5.3 倍。此外,采用本文方法的解码图象具有一般模型基解码图象失真的特点^[1]: 主要是人眼不敏感的“几何失真”和“物理失真”(图

6)。所以在相同比特率时解码图象的质量也优于 H. 261/3 或 MPEG-1/2。

7 结 论

本文在图象边缘提取与优化的基础上建立了一种新的运动图象信源模型,并且实现了系统编码的计算机模拟。目前已达到的压缩率约为 0.075bits/pel,与物体基方法相当。若采用适当的参数编码技术(如矢量量化、霍夫曼编码等),压缩率可望达到 0.04bits/pel 以下。这种模型基编码方法有以下特点:

- (1) 不需要图象的任何先验知识;
- (2) 根据主观视觉特性,以边缘信息作为表达运动图象的空间与时间特征的核心,从而模型建立较为简单;
- (3) 引入基于几何特征的图象匹配方法,作为运动分析的基础;
- (4) 建立了广义运动概念,并提出广义运动分析方法,丰富了帧间预测编码方法,收到较好效果。

参 考 文 献

- 1 Li H, Lunkmark A, Forchheimer R. Image Sequence Coding at Very Low Bitrates: A Review IEEE Trans. Image Processing, 1994, 3(5).
- 2 Forchheimer R, Fahlander O, Kronander T. A semantic approach to the transmission of face images. Picture Coding Symp. (PCS'84), Cesson-Sevigne, 1984, (10).
- 3 Ostermann J. Object-Based Analysis-Synthesis Coding Based on the Source Model of Rigid 3D Objects. Signal Processing: Image Communication, 1984, 6.
- 4 Yosi Yomdin, Yoram Elichai. Normal Forms Representation-A Technology for Image Compression. SPIE, Image and Video Processing, 1993, 1903.
- 5 Geman S. et al. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Trans. Pattern Analysis and Intelligence, 1984, Pami-6(b).
- 6 Nevatia R. et al. Linear Feature Extraction and Description. Comput. Graphics and Image Processing, 1980, 13.
- 7 Mmedoni G. et al. Matching Image Using Linear Features. IEEE Trans. Pattern Analysis and Intelligence, 1984, Pami-6, (6).
- 8 王勇, 吴立德. 图象恢复和边缘提取的后验均值方法. 电子学报, 1994, 2.
- 9 蔡军. 物体基图象编码中的 MR3D 模型建立[硕士论文]. 上海大学, 1996.
- 10 钟峻. 物体基图象编码中 MR3D 模型的运动分析[硕士论文]. 上海大学, 1996.



高磊 博士研究生,1970年4月生于合肥,1992年毕业于安徽大学电子工程与信息科学系。同年进上海大学通信与电子工程系攻读硕士学位,现为上海大学通信与信息工程学院博士研究生。近年来主要从事运动图象压缩编码、瞬变电磁场和数字信号处理技术等领域的科研工作。

张兆扬,教授,博士生导师,1938年9月生,现任上海大学通信与信息工程学院副院长。长期从事数字视频、图象压缩编码、图象处理和多媒体的研究和教学工作。曾完成多项国家自然科学基金、863和国家部委项目的研究,获电子工业部和上海市科技进步二等奖2项、三等奖4项,《数字电视原理》一书获全国电子类优秀著作一等奖。在上述领域内已发表论文60余篇。

张文俊,博士,教授,博士生导师。1959年生于上海,1978年考入上海复旦大学数学系,次年公派赴外在贝尔格莱德大学留学,先后获工程师、硕士和博士学位。1989年回国后在上海大学从事教学和科研工作,作为项目负责人承担及均匀成了国家、部委及上海市下达的纵向项目八项,取得一批科研成果,先后发表学术论文六十余篇。主要研究领域有:通信网技术、图象处理与压缩、电磁场与微波技术等。

Normal Forms Edge-Model Based Compression Coding on Image Sequence

Gao Lei, Zhang Zhaoyang, Zhang Wenjun

(School of Communication & Information Engineering, Shanghai University, Shanghai 200072)

Abstract A new source model for image sequence based on detecting and optimization image edge established as well as the coder system is simulated by computer. On the basis of subjective vision characters, the basic idea of this paper is; the edge information of an image sequence is the most significant factor which can describe the image's characters both in space and time domain. By this scheme, neither any priori knowledge of an image is needed nor too much complicated techniques must be applied for model's construction or realization. According to the source model, the method of image matching using geometrical constrains is achieved for motion analysis; means of prediction between frames were extended by the concept of general motion introduced; and the algorithm of general motion analysis is completed. The scheme is also characterized by fitting gray value distribution of images using normal polynomials.

Keywords Image compression, Model-based coding, Edge detecting, Image matching, Motion analysis