

一种用于视频对象编码的运动模式识别算法

黎洪松, 许保华

(北京师范大学信息科学与技术学院, 北京 100875)

摘 要: 针对目前视频编码中广泛采用的块匹配运动估计补偿 (ME + MC) 算法的不足, 提出一种基于自组织映射 (SOM) 的运动模式识别 (MPR) 算法, 并将其应用于会议电视的视频对象编码中. 为了改善 SOM 算法的性能, 提出一种频率敏感的自组织映射算法 (FSOM). 实验表明, 与 ME + MC 算法相比, FSOM-MPR 算法具有更好的预测编码性能. 对 Claire 视频测试序列, 当压缩比为 170:1 时, 重建视频图像的平均峰值信噪比 (PSNR) 有 2.7dB 的改善.

关键词: 运动模式识别; 基于对象的视频编码; 自组织映射

中图分类号: TP183; TN919.8 **文献标识码:** A **文章编号:** 0372-2112 (2007) 12-2324-05

A Motion Recognition Algorithm for Object-Based Video Coding

LI Hong-song, XU Bao-hua

(College of Information Science and Technology, Beijing Normal University, Beijing 100875, China)

Abstract: To improve the interframe prediction capability of motion estimation and motion complement (ME + MC) in video coding, a motion pattern recognition (MPR) algorithm based on self-organizing feature maps (SOM) was proposed, and it was tested in object-based conference video coding. The frequency sensitive self-organizing feature maps (FSOM) algorithm was also proposed to overcome the disadvantages of SOM algorithm. Experimental results show that FSOM-MPR algorithm has better coding performances than ME + MC algorithm. When the compression ratio is 170:1, the improvement of PSNR is about 2.7dB using Claire conference video sequence.

Key words: motion pattern recognition; object-based video coding; self-organizing feature maps

1 引言

基于对象的视频编码方法充分利用了人类视觉系统的相关特性, 在较高压缩比下仍能保持较好的主观感觉^[1~3]. 但即使是最新的国际视频编码标准 MPEG-4, 视频序列的帧间编码仍然采用较简单的块匹配运动估计补偿 (ME + MC) 算法^[4]. ME + MC 算法假定物体只做平移运动, 认为当前帧中的每个图像子块由其匹配块平移而来. 由于没有考虑物体的旋转、缩放、映射以及运动遮挡和环境照明变化等, 其预测性能会受到一定影响.

针对 ME + MC 算法的不足, Huang 等人^[5]将视频图像分为刚性和非刚性运动区域两类, 刚性运动区域采用运动估计补偿编码, 非刚性运动区域则采用矢量量化编码, 由于两类区域都是不规则形状, 区域分割和运动估计比较困难. 作为一种新的探索, 本文提出一种基于自组织映射 (SOM) 算法的运动模式识别 (MPR) 算法, 并将其应用于基于对象的视频编码. SOM 算法是一种非常有效的聚类方法^[6~8], 通过对样本的无监督训练学习,

可以得到与训练信源相匹配的模式库. 由于 SOM 算法存在训练效率低、无效模式矢量过多等问题, 本文提出频率敏感自组织映射 (FSOM) 算法以改善其性能. 应用于三组会议电视序列的编码实验表明, FSOM-MPR 算法比 ME + MC 算法有更佳的预测编码性能, 重建图像的主客观评价均有不同程度的提高.

2 运动模式识别与视频编码方案

2.1 运动模式识别

运动模式识别的基本思想是: 编码前利用 FSOM 算法建立视频运动模式库; 编码时则利用运动模式识别对视频图像进行预测补偿编码. 会议电视的视频背景通常是固定不变的, 可只针对视频对象建立对象运动模式库, 也就是人体头肩图像的模式库, 考虑到压缩比的要求, 模式矢量采用 8 × 8 的图像块. 通常用于拍摄会议电视的摄像机是固定的 (偶尔会有小幅度抖动), 视频对象的运动也相对比较平缓, 这些因素对设置对象运动模式库的大小有影响.

收稿日期: 2006-09-12; 修回日期: 2007-04-08

基金项目: 国家自然科学基金资助项目 (No. 6027101)

2.2 视频编码方案

基于运动模式识别的视频编码方案如图 1 所示,其具体步骤如下:

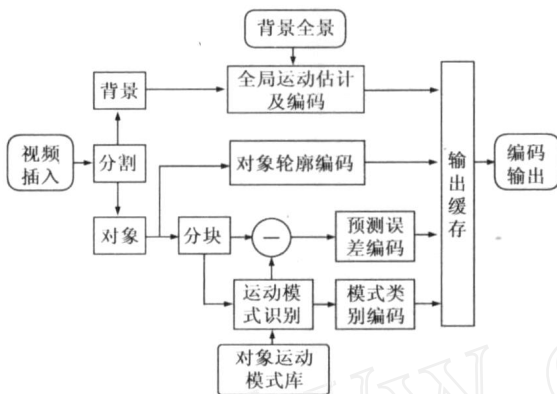


图 1 基于运动模式识别的视频编码方案

步骤 1:编码前,对所有视频图像进行对象分割,得到视频对象区域和视频背景区域.视频对象区域采用 FSOM 算法训练得到对象运动模式库.视频背景区域进行全局运动估计和加权求和,形成视频背景全景.

步骤 2:编码时,对当前视频图像进行视频对象分割,然后分别对视频背景和视频对象编码.

步骤 3:视频背景编码.根据视频背景全景进行全局运动估计,并对运动参数编码;解码时,视频背景由视频背景全景及其全局运动参数恢复.

步骤 4:视频对象编码.对象轮廓采用微分链编码;对象纹理采用运动模式识别编码.即在对象运动模式库中为每个对象子块搜索匹配模式,然后分别对预测误差和模式类别编码.

3 视频对象分割

视频对象分割是基于对象的视频编码基础,也是其关键部分.为了实时应用, Kim 等人提出一种基于块的快速对象生成算法^[9],并将其用于电视序列,得到较好的效果.通过定义对象边界块并将分割操作限制在边界块上,其算法可以降低对象分割的计算量.但是其边界块定义和分割方法复杂,本文给出一个较为简单的基于块匹配的对象分割算法,算法流程如图 2 所示,其中 I_k 和 O_k 分别表示第 k 帧视频图像和对象模板.

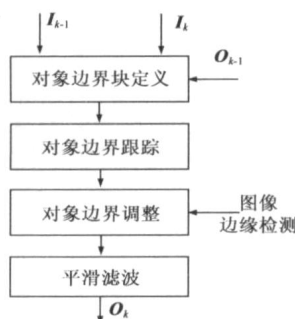


图 2 视频对象分割流程图

(1)对象边界块定义.为了降低计算量,首先定义对象边界块,接下来的操作将只在对象边界块上操作.对于当前帧的每个块,如果它在参考帧中的搜索区域属于背景区域,那么此块定义为背景块(图 3(a)中的黑色

块);如果搜索区域属于对象区域,那么此块定义为对象块(图 3(a)中的白色块);其他块属于对象边界块.

(2)对象边界跟踪.基于块匹配法估计对象边界块在参考帧中的匹配块,并将匹配块对应的二值模板复制到当前帧,得到的对象模板如图 3(b)所示.

(3)对象边界调整.使用 Canny 算子计算当前帧的图像边缘.然后联合边缘信息调整对象模板的边界:如果对象模板边界附近有边缘,将模板边界向边缘靠近,结果如图 3(c)所示.

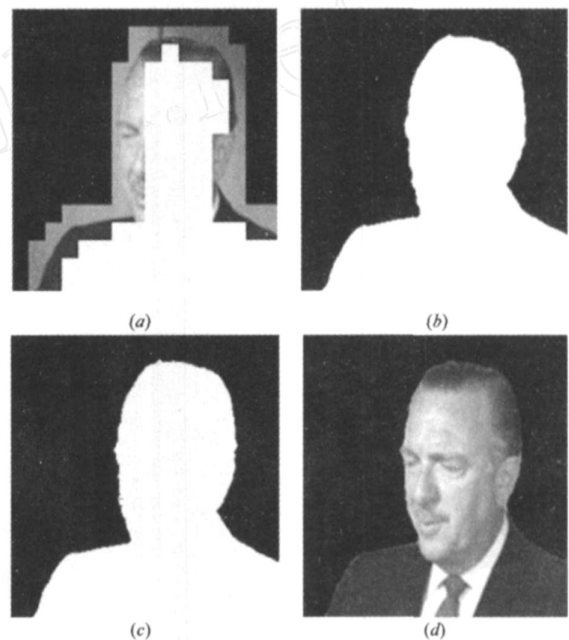


图 3 视频对象分割示意图

(4)平滑滤波:利用形态学滤波和中值滤波平滑对象模板,最终得到的视频对象如图 3(d)所示.

4 对象运动模式库的训练算法

4.1 SOM 算法存在的问题

得到视频对象后,采用 SOM 算法建立对象运动模式库.但是 SOM 算法存在两个主要问题:一是如何提高学习训练的效率;一是如何减少无效模式矢量数量,提高对信源的适应性.究其原因,主要有以下两个方面:

(1)SOM 算法是以调整邻域内的模式矢量来试图减少无效模式矢量的,尽管每个模式矢量都有可能被调整,但每个模式矢量成为获胜模式矢量的概率是不一样的,模式矢量获胜概率越大,则该模式矢量及其邻域内的模式矢量调整的机会也就越大;反之,调整的机会就越小.

(2)SOM 最佳映射希望每个模式矢量都能得到充分地利用,要求每个模式矢量能有效地适应信源的输入矢量,这就要求训练机制能体现模式矢量对训练矢量的自适应性,有必要对每个模式矢量的调整频率进

行“公平”调节,以适应各种可能的输入矢量。

4.2 FSOM 算法基本思想

针对 SOM 算法存在的公平性问题,使用频率敏感的 SOM 算法来提高训练效率,其基本思想如下:

(1) 为每个模式矢量设置一个获胜频率计数器 c_j ($j = 0, 1, \dots, N-1$), 模式矢量每获胜一次, c_j 加 1。

(2) 采用下式计算输入矢量与模式库中模式矢量的失真

$$\bar{D}_j^{(n)} = f(c_j) D_j^{(n)} \quad (1)$$

式中 $D_j^{(n)}$ 为均方失真, 频率敏感函数 $f(c_j) = c_j^s$, $s(0 < s < 1)$ 为频率敏感参数, s 越大, 模式矢量的失真对频率越敏感。显然, 当某个模式矢量获胜次数增加时, 其失真测度将增大, 其再次获胜的概率将降低, 达到人为的“公平”。

4.3 FSOM 算法步骤

步骤 1: 设置对象运动模式库的大小为 (N, M) , 其中 N 为模式矢量的数目, M 每个模式矢量的大小; 选定训练样本集 $\{X_n, n = 0, 1, \dots, L-1\}$, 其中 L 为训练样本数目; 设置初始模式库 $\{W_j^{(0)}, j = 0, 1, \dots, N-1\}$ 和初始邻域 $NE_j^{(0)}$; 令响应计数器 $c_j = 1$, 算法迭代次数 $n = 0$ 。

4.3 FSOM 算法步骤

步骤 2: 输入一个新的训练样本矢量 X_n 。

步骤 3: 按照式(1) 分别计算训练样本矢量 X_n 与对象运动模式库中各模式矢量的失真 $\bar{D}_j^{(n)}$, 选择具有最小失真的模式矢量 $W_j^{(n)}$ 为获胜模式矢量, 即

$$D_j^{(n)} = \min_{j \in N-1} \bar{D}_j^{(n)} \quad (2)$$

步骤 4: 按下式调整 $W_j^{(n)}$ 及其拓扑邻域 $NE_j^{(n)}$ 内的模式矢量, 即

$$W_j^{(n+1)} = \begin{cases} W_j^{(n)} + (c_j) [X_n - W_j^{(n)}], & j \in NE_j^{(n)} \\ W_j^{(n)}, & \text{其他} \end{cases} \quad (3)$$

式中, 学习函数 $(c_j) = A_1 e^{-c_j/T_1}$, $A_1 = 1$ 为训练初的最大学习速度, T_1 为衰减常数, T_1 越大, 学习函数下降越慢, 反之下降越快, 实验时应该合理选择。 $NE_j^{(n)} = 1 + A_2 e^{-n/T_2}$, A_2 为最大邻域, T_2 为衰减常数, 与 T_1 的选值规律一样。

步骤 5: 令 $c_j = c_j + 1$, 返回步骤 2, 直到训练结束。

5 视频背景编码

视频背景采用静态 Sprite 编码方法。针对视频背景本身没有运动的特点, Sprite 编码通过拼接技术生成一个视频背景全景, 解码时再由视频背景全景恢复每帧图像的背景。

假设视频背景图像表示为 $B_i, i = 0, \dots, n-1$, 则视频背景全景(表示为 S) 的生成算法为:

步骤 1: 令 $S = B_0, i = 1$;

步骤 2: 估计视频背景 B_i 到 S 的全局运动参数

MV_i , 并根据 MV_i 将 B_i 向 S 变形为 B_i , 然后使用 $S = (i \times S + B_i) / (i + 1)$ 更新视频背景全景;

步骤 3: 令 $i = i + 1$, 返回步骤 2, 直到生成视频背景全景图像。

视频背景全景对所有视频背景图像进行加权求和, 可以有效地消除单幅图像的噪声。视频编码时, 视频背景全景使用 DCT 编码。每帧图像的背景则进行全局运动估计并对运动参数编码。

6 视频对象编码

人类视觉系统对物体的边缘和轮廓非常敏感, 纹理信息次之。基于对象的视频编码充分利用了人类视觉系统的特性, 可以在低码率下获得较好的视觉效果。

6.1 视频对象轮廓编码

链编码方法^[10]可精确地对对象轮廓进行编码, 其基本方法是在确定对象轮廓的起点之后, 对下一个轮廓像素所处位置的方向进行编码。

除了对轮廓的方向进行编码外, 链编码还可以对轮廓方向的变化进行编码(即微分链码), 由于它利用了轮廓方向的统计相关性, 因此编码效率更高。

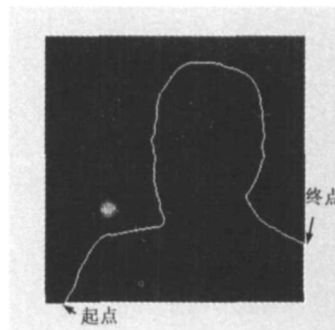


图 4 视频对象轮廓

针对会议电视中视频对象轮廓的特点, 我们首先确定对象轮廓的起点和终点(如图 4 所示), 然后对从起点到终点的轮廓曲线采用微分链码编码。

6.2 视频对象纹理编码

视频对象的纹理信息采用运动模式识别编码(如图 1 所示), 视频对象分块后, 为每个子块搜索匹配模式, 然后分别对子块的模式类别和预测误差进行编码。

会议电视序列具有很强的时间和空间相关性, 相邻帧中位于相同位置的对象子块(称为时间相邻块)以及同一帧中位置相邻的对象子块(称为空间相邻块)的匹配模式相近的概率较大, 为此可采用索引补偿算法^[11]对对象子块的模式类别进行编码。其基本思想是: 得到当前对象子块的匹配模式后, 先计算该模式和时间相邻块的匹配模式之间的均方误差, 如果误差小于预置的门限值, 则编码为‘0’, 否则编码为‘1’, 同时对匹配模式的类别进行编码。由于视频对象的运动, 对象子块可能不存在时间相邻块, 此时可采用空间相邻块作为其参考块。

预测误差采用 DCT 编码, 即 DCT 变换后对系数进行量化、“Z”形扫描和熵编码。为了适应不同压缩比的

要求,预测误差的编码分为两类:(1)中低压缩比,根据预测误差的大小判断是否对其编码;(2)高压压缩比,所有子块的预测误差均不编码。

7 实验结果

采用三组会议电视序列(Walter, Claire 和 Akiyo)来模拟本文提出的编码方法.视频图像均为 256 × 256 灰度图像,共 16 帧;视频对象区域的图像质量测度采用平均峰值信噪比(PSNR).对象运动模式库的训练参数:模式库大小为 2048, $s = 0.2$, $T_1 = 20$, $A_2 = 4$, $T_2 = 1024$, 初

始模式库采用随机抽取加排序的方法。

图 5 给出三个视频序列的编码结果.可以看到,与 ME+MC 算法相比,FSOM-MPR 算法得到的重建图像的 PSNR 均有不同程度的改善,Walter 序列的 PSNR 改善值小于另两个序列,这是因为 Walter 序列中对象的运动幅度较大,这意味着它需要更多的比特数传输模式类别和预测误差.图 6 则给出两种算法在相似压缩比下的重建图像(视频序列第 15 帧),显然 FSOM-MPR 算法得到的重建图像具有较好的主观效果,图像中的细节区域(如眼睛、嘴巴等)更为清晰。

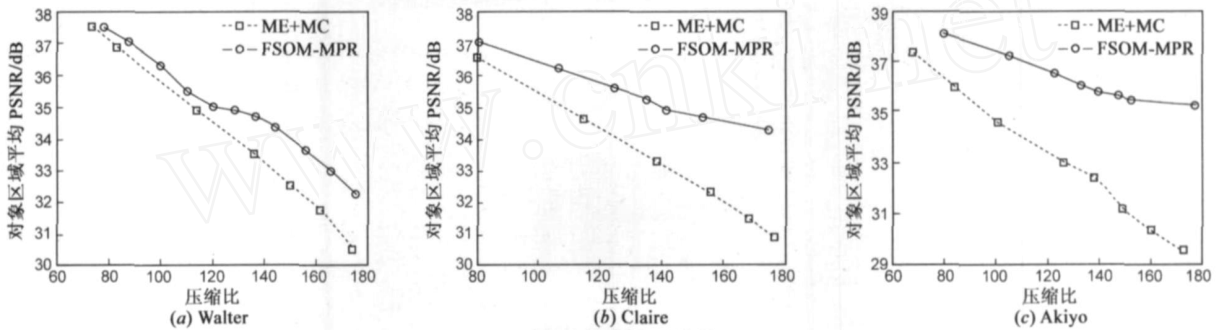


图 5 视频编码结果



图 6 视频重建图像 (CR 表示压缩比)

8 结论

作为一种新的探索,本文提出一种基于自组织映

射的运动模式识别算法,并用于会议电视的对象编码.实验表明,在高压压缩比的情况下,基于自组织映射的运动模式识别算法较块匹配运动估计补偿算法具有更好

的预测编码性能.

参考文献:

- [1] Ebrahimi T, Horne C. MPEG-4 natural video coding-an overview [J]. Signal Processing: Image Communications, 2000, 15(4-5):365-385.
- [2] Shamim A, Robinson J A. Object-based video coding by global-to-local motion segmentation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(12):1106-1116.
- [3] Chaumon M, Pateux S, Nicoias H. Object-based video coding using a dynamic coding approach[A]. 2004 International Conference on Image Processing (ICIP) [C]. Rennes:IRISA, Campus de Beaulieu, 2004. 1105-1108.
- [4] Hang H M, Chou Y M, Cheng S C. Motion estimation for video coding standards[J]. Journal of VLSI Signal Processing, 1997, 17(2-3):113-136.
- [5] Huang C S, Lin Y, Sun M T. Region-based video coding using a geometric motion compensation[J]. Journal of Visual Communication and Image Representation, 2000, 11(3):279-301.
- [6] Kohonen T. An introduction to neural computing[J]. Neural Networks, 1988, 1(1):3-16.
- [7] Seiffert U, Michaelis B. Estimating motion parameters with three-dimensional self-organizing maps [J]. Information Science, 1997, 101(3-4):187-201.
- [8] Cirrincione G, Cirrincione M. A novel self-organizing neural network for motion segmentation [J]. Applied Intelligence, 2003, 18(1):27-35.
- [9] Kim B C, Park R H. A fast automatic VOP generation using boundary block segmentation[J]. Real-Time Imaging, 2004, 10(2):117-125.
- [10] Wang Y, Osetermann J, Zhang Y Q. Video Processing and Communications[M]. Beijing: Publishing House of Electronics Industry, 2003. 245-246.
- [11] Nicolaos B, Karayiannis, Li Y. A replenishment technique for low bit-rate video compression based on wavelets and vector quantization [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2001, 11(5):658-663.

作者简介:

黎洪松 男, 1963年12月出生于湖北省监利县, 现为北京师范大学信息科学与技术学院副教授, 出版著作7本, 在国内外发表学术论文40余篇. 主要研究方向有: 模式识别、智能信息处理、数字视频处理等. E-mail: hongsongli@yahoo.com

许保华 男, 1981年2月出生于江苏省大丰市, 研究生. 主要研究方向为数字图像与视频处理. E-mail: xubaohua@tom.com