

一种基于 BP 网络的 H.264 快速帧内模式选择算法

季国飞,王庆

Ji Guo-fei, WANG Qing

西北工业大学 计算机学院,西安 710072

School of Computer, Northwestern Polytechnical University, Xi'an 710072, China

E-mail: jiguofei@gmail.com

Ji Guo-fei, WANG Qing, BP network based fast mode decision algorithm for intra prediction in H.264 video coding. Computer Engineering and Applications, 2008, 44(13): 100-102.

Abstract: H.264/AVC provides 17 modes for intra prediction, and the optimal mode is selected by Rate Distortion Optimization (RDO) with high coding complexity. Although most of the current algorithms of intra prediction mode decision can reduce coding complexity greatly, there still exists high computational complexity. A new fast intra prediction mode decision algorithm based on BP neural network is proposed. The algorithm firstly judges the mode of the macroblock as either Intra16×16 or Intra4×4 mode, using the training result of the BP neural network with lots of video sequence. Then the optimal mode is decided from 17 modes by Pan Feng's method. Experimental results show that the proposed algorithm can efficiently reduce the computational complexity up to 75%~83% compared to JM86 and 30%~70% compared to Pan Feng's method, and can maintain the coding performance (the reduction of PSNR is only 0.01 dB on average).

Key words: H.264; intra prediction; mode decision; BP network

摘要: H.264/AVC 采用 17 种预测模式和率失真优化模型来进行帧内模式的选择, 复杂度很高。目前的快速帧内模式选择算法能有效减少模式选择的复杂度, 但还存在很大的不足。提出了一种基于 BP 神经网络的快速帧内模式选择算法。该算法使用 BP 神经网络对大量视频序列预先训练的结果对宏块进行预判, 得出当前宏块采用 Intra16×16 或 Intra4×4 模式, 然后使用 Pan Feng 的方法判断具体选择何种模式。实验结果表明, 该算法在保证视频质量几乎不变 (PSNR 平均降低约 0.01 dB) 的同时, 可以有效减少计算复杂度, 与 JM86 相比减少 75%~83%, 与 Pan Feng 的方法相比减少 30%~70%。

关键词: H.264; 帧内编码; 模式选择; BP 神经网络

DOI: 10.3778/j.issn.1002-8331.2008.13.030 **文章编号:** 1002-8331(2008)13-0100-03 **文献标识码:** A **中图分类号:** TP393

1 引言

H.264/AVC^[1] 视频编码标准由联合视频开发小组 (JVT) 于 2003 年 5 月颁布, 与以往的视频编码标准相比, H.264/AVC 在码率上减少了 30%~70%^[2]。H.264/AVC 采用了帧内预测、整数变换、4×4 块的运动估计/运动补偿、环路滤波、率失真优化技术 (RDO, Rate Distortion Optimization) 等新的编码技术, 使其具有压缩效率高、网络性能好等优点, 现已受到越来越广泛的关注。其中的帧内预测技术利用图像的空间相关性, 先使用周围的像素进行适当的插值运算, 得到原图像的预测图像, 再对原图像和预测图像的残差进行变换编码。在以往标准中则是采用直接对帧内图像块做变换编码的方法, 与之相比, H.264/AVC 做帧内预测后能够明显提高编码效率和降低编码比特率。但是与此同时也大大增加了计算的复杂度, 很难用到实时性要求较高的应用中去, 因此研究相关快速算法具有重要意义。

在 H.264/AVC 标准^[3]中, 亮度块可以有 9 种 4×4 块和 4 种

16×16 块帧内预测模式, 色度 8×8 块有 4 种预测模式, 和亮度 16×16 块相同。H.264/AVC 进行模式选择的传统方法是首先分别计算出一个宏块的 Intra 16×16 的 4 种模式以及该块的 16 个 Intra 4×4 块的 9 种模式的率失真代价, 通过比较 Intra 16×16 和 16 个 Intra 4×4 块率失真代价之和的大小决定使用 Intra 16×16 或者 Intra 4×4 模式进行预测, 所以总共要进行率失真代价函数的计算次数为 592 次^[4], 而率失真代价本身的计算过程也十分耗时, 在编码时预测模式生成和选择的时间要占到总计算时间的 77%^[5]。

到目前为止, 已有许多学者提出了很多快速帧内模式选择算法。文献[6]通过简化 RDO 算法并结合硬件实现来减少计算复杂度。文献[8, 9]在变换域进行帧内预测, 改变传统编码结构来优化帧内模式选择过程。不过在变换域内进行帧内预测实现较为复杂, 而且视频质量损失较多。

文献[3, 4]是由 Pan Feng 提出的基于 Sobel 边缘检测 (Edge

基金项目: 教育部“新世纪优秀人才支持计划” (NCET, Supported by Program for New Century Excellent Talents in University); 西北工业大学研究生创业种子基金 (Graduate Starting Seed Fund of Northwestern Polytechnical University No. Z200754)。

作者简介: 季国飞 (1984-), 男, 硕士研究生, 研究方向为: 视频压缩编码; 王庆 (1969-), 男, 教授, 博士生导师, 研究方向为: 图像处理、结构模式识别。

收稿日期: 2007-08-29 **修回日期:** 2007-10-24

Detection)的快速帧内模式选择算法,它利用当前块及其周围像素相关性,通过计算幅值和角度信息选择最有可能的模式作为候选模式进行预测,大大降低了帧内模式选择的复杂度。该算法有很好的效果,RDO 计算次数从传统的 592 次降到了 132 次^[4],平均可以节省 50%的计算时间。

文献[7]与 Pan Feng 方法类似,不同的是它使用整数变换获得边缘信息,然后再根据计算出来的角度信息来缩小候选模式的范围,以此减少计算复杂度。

文献[3,4]和文献[7]中的方法虽然取得了很好的效果,不过它们只是减少了 Intra 16×16 或者 Intra 4×4 块预测时的预选模式个数,仍然需要分别计算 Intra 16×16 以及 Intra 4×4 的率失真代价并比较选出最小者作为最优模式,所以计算复杂度仍然很高。

本文在充分考虑了图像宏块特征以及 Pan Feng 的 Sobel 边缘检测模式选择算法^[3,4](简称 Pan Feng 的方法,下同)的基础上,提出了一种基于 BP 神经网络的快速帧内模式选择算法。该算法首先利用大量视频序列对 BP 网络进行训练,基于训练结果对 Intra 16×16 和 Intra 4×4 模式进行预判决,从而在保证视频质量几乎不变的情况下使编码时间减少 30%~70%。

2 基于 BP 神经网络的帧内模式选择算法

Intra 16×16 模式以整个宏块为处理单元,Intra 4×4 模式则把一个宏块分成 16 个 4×4 子块进行处理。注意到编码后的视频序列,采用 Intra 16×16 编码的块大多属于比较平滑、色彩单一、纹理简单的块,而那些采用 Intra 4×4 中 9 种模式进行编码的块则属于细节较多、纹理较复杂的块。

基于以上观察到的事实考虑,在编码过程中,如果能够依据宏块的某种特征,对于特征明显的宏块提前决定采用 Intra 16×16 或者 Intra 4×4 预测模式,避免对亮度 Intra 16×16 和 Intra 4×4 两种模式的全搜索,这样的话就可以减少大量不必要的计算。

考虑到时间复杂度不能过高,因此就增加了特征选取的难度。我们选取相对比较简单的平均绝对误差和 S 来作为特征。首先计算当前 16×16 块的平均像素值 Avg ,如式(1)所示,其中 $p_{i,j}$ 表示当前 16×16 块中 (i,j) 位置亮度分量(Y 分量)的值(由于色度分量的影响较小,所以本文方法中只使用亮度分量作为判断标准):

$$Avg = \frac{\sum_{i=0}^{15} \sum_{j=0}^{15} p_{i,j}}{16 * 16} \quad (1)$$

然后计算每一点上亮度分量值和 Avg 的差值的绝对值,相加得到平均绝对误差和作为特征,公式如下:

$$S = \sum_{i=0}^{15} \sum_{j=0}^{15} |Avg - p_{i,j}| \quad (2)$$

图 1 和 2 为使用标准测试模型 JM86 分别编码测试序列 Stefan 和 Foreman 前 20 帧图像的编码模式分布图,纵轴为 S ,横轴为 Avg 。

从分布图可以看出,使用特征 S 并结合 Avg 即可较为明显地区分出采用 Intra 16×16 或者 Intra 4×4 模式预测的块。在这里可以采用阈值判断的方法进行预判,不过由于视频序列本身的差异性,如果选取固定的阈值,会出现某些视频序列预测准确,某些视频序列误判率高的情况,不具有通用性。

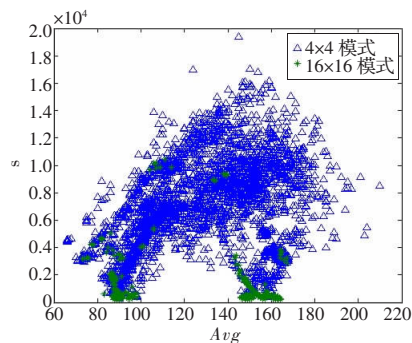


图 1 Stefan 前 20 帧模式分布

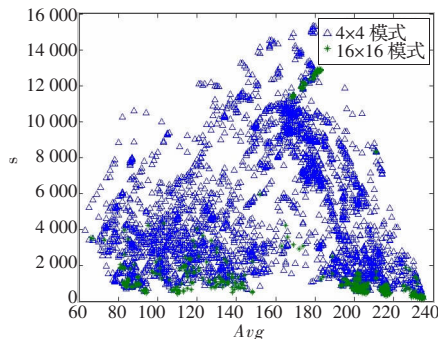


图 2 Foreman 前 20 帧模式分布

考虑到 BP 神经网络^[10]具有很强的自学习和自适应特征,并已在图像处理、模式识别等领域取得了很好的效果,在这里把 BP 神经网络引入到算法中。本文设计的网络结构是 3 层结构,输入层有 2 个节点,结构中只有 1 个隐含层,有 3 个节点,输出层有 1 个节点,此网络可表示为 $BP(2,3,1)$ 。

第 p 组样本的输入为 $\mu_p = \{Avg, S_p\}$, 相应的网络输出是 d_p 。节点 i 在第 p 组样本输入时,输出为 y_{ip}

$$y_{ip} = f(x_{ip}) = f\left(\sum w_{ij} I_{jp} + b_i\right) \quad (3)$$

其中, I_{jp} 是第 p 组输入时,节点 i 的第 j 个输出, w_{ij} 为上一层节点 i 和下一层节点 j 之间的连接权值, b_i 为节点 i 的偏移值,隐含层和输出层的激活函数采用 Sigmoid 型变换函数,即:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

网络训练目标是使误差函数 E 最小,由于本网络只有一个输出节点,所以 E 的定义可简化为:

$$E = \frac{1}{2} \sum (t_p - d_p)^2 \quad (5)$$

其中, t_p 为第 p 个样本的期望输出, d_p 为该样本的实际输出。

训练时,选取 ε 为网络允许误差, $\varepsilon > 0$ 。如果 $E < \varepsilon$,则说明网络误差满足要求,计算结束。否则,依据误差的梯度下降算法^[10],逐步调整权值和偏移值。

本文采用的实验样本来自于大量的视频测试序列,样本数为 200 000 组,迭代次数选择 2 000 000 次,网络允许误差 ε 设为 0.01。

表 1 是使用 JM86 以及本文设计的 BP 网络训练的结果对 Foreman 和 Stefan 测试序列前 20 帧(7 920 组样本)进行模式判决的结果。

图 3 和图 4 是 Foreman 测试序列第一帧分别使用 JM86 编码以及上述 BP 网络模型进行预测得到的各宏块编码模式分

表1 本文方法得到的模式预测结果

| 序列 | 算法 | Intra 16×16 | Intra 4×4 | 误判率/% |
|---------|-------|-------------|-----------|-------|
| Foreman | JM86 | 1 497 | 6 423 | |
| | BP 算法 | 1 440 | 6 480 | 10.2 |
| Stefan | JM86 | 1 586 | 6 334 | |
| | BP 算法 | 1 614 | 6 306 | 4.9 |

布图。其中大块表示采用 Intra 16×16 预测模式,小块表示采用 Intra 4×4 预测模式。

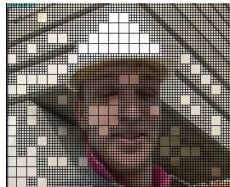


图3 Foreman 序列第一帧采用 JM86 预测结果

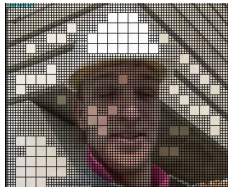


图4 Foreman 序列第一帧采用本文算法预测结果

虽然对 Foreman 等测试序列,会有 10%左右的模式误判,但是从图 3 和图 4 可以看出,判决结果基本符合当前宏块的特征,平滑、细节较少的宏块大部分被划为 Intra 16×16 预测模式,细节较多的宏块则多被划为 Intra 4×4 预测模式。而且由后边的实验数据也可以验证,虽然 Foreman 会有 10%左右的模式误判,但是编码后的 PSNR 的下降几乎可以忽略不计,不会影响视频质量。可见,使用上面设计的 BP 神经网络,可以对宏块的模式进行很好的预判,应用到编码过程中,可以有效的减少编码时间。

3 本文算法的计算流程

步骤 1 在编码前从大量视频序列中提取出上面介绍的特征 S 和 Av_g ,并使用本文设计的 BP 神经网络进行训练,得到各层间的权值及偏移值信息。实验参数可以参考第 2 章中的介绍。

步骤 2 与 Pan Feng 的方法^[3,4]相同,对该帧使用 Sobel 边缘算子生成其对应的边缘图,该边缘图中的每个像素关联一个边缘矢量:

$$D_{i,j}=\{dx_{i,j},dy_{i,j}\} \quad (6)$$

其中, $dx_{i,j}$ 和 $dy_{i,j}$ 表示水平和垂直方向上的差分。从边缘矢量可以计算出幅值和角度,即:

$$Amp(D_{i,j})=|dx_{i,j}|+|dy_{i,j}| \quad (7)$$

$$Ang(D_{i,j})=\frac{180^\circ}{\pi} \times \arctan\left(\frac{dx_{i,j}}{dy_{i,j}}\right), Ang(D_{i,j}) < 90^\circ \quad (8)$$

步骤 3 根据公式(1)、(2)计算当前宏块的 Av_g 以及 S 特征,使用上一步训练结果中的权值及偏移值信息对其进行验证,当输出值大于 δ ,则为 Intra 4×4 模式,转到步骤 4;否则为 Intra 16×16 模式,转到步骤 5。为简化计算,此处 δ 取 0.5。

步骤 4 已判决当前宏块使用 Intra 4×4 模式,根据步骤 2 中的幅值信息统计得出图像块的累计方向直方图,取直方图中最大的方向及其两个相邻方向模式以及平均模式(DC)共 4 种模式作为候选模式,而不是传统方法中的 9 种模式^[4]。转到步骤 6。

步骤 5 已判决当前宏块使用 Intra 16×16 模式,根据步骤 2 中的幅值信息统计得出图像块的累计方向直方图,取直方图中最大的方向和平均模式(DC)共 2 个模式中进行,而不是传统方法中的 4 种模式^[4]。转到步骤 6。

步骤 6 对色度分量 8×8 块,和 Intra 16×16 模式的处理相同^[4]。

步骤 7 对各候选模式分别计算率失真代价(RDO_Cost),率失真代价最小者则为最佳模式。

4 实验结果和分析

实验基于 JM86 参考模型完成,本文选择 Foreman、Stefan 以及 Football 三种不同复杂度的 CIF 格式测试序列。编码参数为:采用 CAVLC 熵编码;编码帧数分别为 300,300,90;编码序列为全 I 帧;量化参数分别是:24,28,32,36;通过修改 JM86 参考模型中的参数分别实现 Pan Feng 的算法和本文算法,可以得到两个算法相应的编码结果,如表 2 所示。

表 2 中 QP 表示量化参数, $\Delta psnr_y$ 和 $\Delta psnr_u, \Delta psnr_v$ 分别为亮度和色度的峰值信噪比的平均增量,单位为 dB。 $\Delta time$ 为时间的增量百分比。由于在算法中未对色度分量(U, V 分量)进行修改,因此本文算法得出的 $psnr_u, psnr_v$ 与 Pan Feng 的方法的结果一样,故表中不予列出。

表2 本文方法和 JM86 以及 Pan Feng 方法的结果对比

| 序列 | QP | 与 JM86 的方法比较 | | | | 与 Pan Feng 方法比较 | |
|----------|------|---------------------------|---------------------------|---------------------------|------------------|---------------------------|------------------|
| | | $\Delta psnr_y/\text{dB}$ | $\Delta psnr_u/\text{dB}$ | $\Delta psnr_v/\text{dB}$ | $\Delta time/\%$ | $\Delta psnr_y/\text{dB}$ | $\Delta time/\%$ |
| Football | 24 | -0.108 6 | -0.081 8 | -0.118 4 | -80.59 | -0.050 0 | -47.03 |
| | 28 | -0.054 8 | -0.043 0 | -0.107 4 | -80.20 | -0.023 8 | -45.45 |
| | 32 | -0.043 1 | 0.050 5 | -0.064 1 | -85.27 | -0.012 2 | -69.32 |
| | 36 | -0.040 2 | 0.103 8 | -0.040 7 | -80.63 | -0.018 3 | -49.64 |
| Foreman | 24 | -0.077 8 | -0.050 9 | -0.090 8 | -75.82 | -0.002 5 | -34.65 |
| | 28 | -0.050 7 | -0.054 8 | -0.093 9 | -82.88 | -0.000 5 | -45.61 |
| | 32 | -0.048 2 | -0.024 1 | -0.055 3 | -79.59 | 0.000 0 | -61.83 |
| | 36 | -0.031 7 | 0.022 3 | -0.047 0 | -76.80 | 0.005 0 | -37.90 |
| Stefan | 24 | -0.138 1 | -0.051 7 | -0.051 9 | -75.41 | -0.001 0 | -32.10 |
| | 28 | -0.120 6 | -0.043 7 | -0.056 1 | -75.22 | -0.002 1 | -30.43 |
| | 32 | -0.116 5 | 0.004 8 | -0.010 9 | -78.23 | -0.006 2 | -43.14 |
| | 36 | -0.109 9 | 0.030 4 | 0.061 6 | -76.02 | -0.010 8 | -35.08 |

从不同序列不同 QP 的编码数据可以看出,本文算法对复杂度较高(大量采用 Intra 4×4 模式)的视频序列还是对比较低平滑、复杂度较低(大量采用 Intra 16×16 模式)的视频序列,速度上都有很大的提升,与 JM86 的结果相比,可提高 75%~83%,与 Pan Feng 的方法相比,提高 30%~70%。不过,在速度提升的同时,码率会略有增加,本文算法实现的码率与 JM86 相比平均约有 6%左右的增幅。理想情况下,如果预判出来的模式和 JM86 标准算法一致,则和标准算法的码率相差不大,但是任何改进的算法都不可能预测得完全准确,实验中的码率增加就是由于对模式的误判造成的。较低的误判率虽然会增加码率,但是对视频质量的影响几乎可以忽略不计。如果增加 BP 神经网络中训练样本的数量,则可以提高判断准确度,减少误判,同时也使码率增加率降低。

5 结论

本文提出了一种基于 BP 神经网络的 H.264/AVC 快速帧内模式选择算法。该算法首先使用一个 BP 神经网络对大量视频序列进行训练,基于训练结果预先判断出当前宏块使用 Intra 16×16 或者 Intra 4×4 模式,然后使用 Pan Feng 的 Sobel 边缘检测模式选择算法^[3,4]对 Intra 16×16 或者 Intra 4×4 的子模式进行选择,得出最优模式进行编码。实验结果表明,该算法在保

(下转 133 页)