

# 一种基于类 haar 小波的 MCTF 视频编码方案

刘兆广, 彭玉华, 杨 阳

(山东大学信息科学与工程学院, 山东济南 250100)

**摘要:** 在传统的基于运动补偿的时间滤波小波视频编码方案中, 帧组尺寸和低通帧位置都是固定的, 没有考虑到视频序列中运动的性质. 本文提出了一种类似 haar 小波的基于运动补偿的时间滤波编码方案, 利用互信息技术, 根据视频的运动特征自适应的选择帧组尺寸和低通帧位置. 同时, 根据选定的帧组尺寸和低通帧位置决定时间上的编码层次. 大量的实验比较了提出的方法和传统方法的压缩性能. 当视频序列运动较为剧烈时, 本文提出的自适应选择低通帧方法较固定低通帧方法的 PSNR 提高约 0.3~0.5dB; 当视频序列中运动缓急不一, 如频繁的画面切换, 本文提出的自适应的选择帧组尺寸的方法能取得更好的效果.

**关键词:** 关键帧; 互信息; 帧组; 基于运动补偿的时间滤波

**中图分类号:** TP391.41      **文献标识码:** A      **文章编号:** 0372-2112 (2008) 02-0386-06

## An Adaptive Haar-Like MCTF Based Wavelet Video Coding Scheme

LIU Zhaoguang, PENG Yuhua, YANG Yang

(School of Information Science and Engineering, Shandong University, Jinan, Shandong 250100, China)

**Abstract:** In conventional motion compensated temporal filtering based wavelet coding scheme, where the size of group of picture and low pass frames' positions are fixed, and the motion activities in video sequences are not considered. In this paper, we propose a haar like motion compensated temporal filtering based wavelet video coding scheme, which can adaptively select the size of group of picture and the low pass frames based on mutual information technology. Then, the temporal decomposition level is determined using the selected size of group of picture and the low pass frames' position. We compare the PSNR performance of the proposed scheme with the conventional scheme by carrying out a large amount of experimental work. When the video sequence has high motion activity, the proposed low pass frame selection can improve the compression quality by about 0.3~0.5 dB compared with the conventional scheme; And, the proposed adaptive approach for selecting the size of group of picture is more suitable for situations with uneven variation of motion activities, e. g. with frequent shot cuts.

**Key words:** key frame; mutual information; group of picture; Motion Compensated Temporal Filtering(MCTF)

## 1 引言

基于 MCTF(Motion Compensated Temporal Filtering) 的小波视频编码方案中采用了开环式编码, 因此避免了“漂移”效应. Ohm<sup>[1]</sup>首次用镜像滤波对时间上的滤波过程进行了全面解释. 提升方案的引入<sup>[2]</sup>解决了在 MCTF 中由于分数像素运动估计造成无法完全重构的问题. 在低比特流情况下, 运动矢量常耗用较大的比特流, 文献 [3] 提出了一种运动矢量的分层次编码方法, 称为分层的变块尺寸匹配(Hierarchical Variable Size Block Matching, HVSBM). 在基于 MCTF 的小波编码方案中, 当时间上滤波器类型和分解层次固定后, GOP(Group of Pictures, 帧组) 尺寸也就固定了. UMCTF (Unconstrained Motion Compensated Temporal Filtering)<sup>[4]</sup> 是一种更为灵活的

MCTF 编码方案. 该方案有一组控制参数, 用户可以根据不同的视频来进行设置. 但文章并没有给出获取这些参数的有效方法. 自适应的 GOP 尺寸在 MPEG-x 及 H. 26x 等编码中业已提出. 文献 [5] 就提出利用 HOD(histogram of frame difference) 来计算帧间的关联, 文献 [6] 则使用与宏块相关的一些参数来决定编码模式和 GOP 结构. 在基于 MCTF 的小波视频编码方案中, 自适应的 GOP 尺寸的使用见文献 [7], 在当前分解层次下, 非连接像素的个数作为评判的标准来决定是否进行下一层分解, 该方法只能得到不同 GOP 结构, 无法改变 GOP 尺寸.

互信息(MI) 可以度量帧间的变化量, 因此, 它可以作为评判镜头边界或关键帧提取的方法之一<sup>[8,9]</sup>. 文献 [8] 使用互信息检测镜头切换、图像渐入、渐出等视频段, 并且在分割得到各个帧簇中, 使用互信息技术提取

其中的关键帧. 而在文献[9]中, 则提出了一种基于互信息的自动设定阈值的方法.

本文利用互信息技术, 提出了一种自适应的选择 GOP 尺寸和低通帧的方法, 并且, MCTF 的编码方案根据选定的 GOP 尺寸和低通帧进行自适应的确定.

## 2 背景知识

### 2.1 基于 haar 的 MCTF 编码方案

在 MCTF 编码过程中, 原始的帧数据在时间上沿着运动轨迹进行滤波, 然后对滤波得到的低通帧和高通帧进行小波变换, 小波系数使用嵌入式方法进行编码, 如 MG-EZBC<sup>[7]</sup>. 图 1 给出了基于 haar 小波的 MCTF 方案, 具体的细节讨论请参考文献[1, 3, 7].

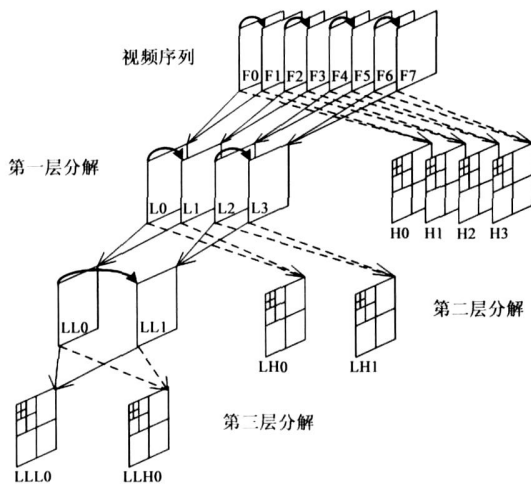


图 1 基于 haar 的 MCTF 编码过程

### 2.2 互信息(MI)

互信息可以度量帧间的信息传递, 因此可以用来检测镜头边界和关键帧提取<sup>[8]</sup>. 帧间的较大的变换对应着较小的互信息值, 反之亦然.

设  $X, Y$  是离散随机变量, 其取值范围为  $\{a_1, a_2, \dots, a_N\}$  及相应概率为:  $\{p_1, p_2, \dots, p_N\}$ , 则变量  $X$  和  $Y$  的互信息为:

$$I(X, Y) = - \sum_{x, y \in A_X, A_Y} p_{XY}(x, y) \log \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} \quad (1)$$

在 YUV 色彩模型的视频序列中, 亮度和色度的互信息值可以分别计算. 众所周知, 视频序列大部分能量集中在亮度分量上. 因此, 本文只考虑亮度分量的互信息计算.

设定视频序列的像素取值范围为: 0 到  $N-1$ , 其中  $N$  为最大灰度阶. 令  $F_t$  和  $F_{t+1}$  分别代表  $t$  时刻帧和  $t+1$  时刻帧,  $P_{t, t+1}(i, j) (0 \leq i, j \leq N-1)$  是  $F_t$  中的像素值为  $i$  而在  $F_{t+1}$  中改变为  $j$  的概率, 该数据可以统计帧数据  $F_t$  和  $F_{t+1}$  得出. 因此可以计算出这两帧间的亮度分量的互信息值为:

$$MI_{t, t+1} = - \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_{t, t+1}(i, j) \log \frac{P_{t, t+1}(i, j)}{P_t(i)P_{t+1}(j)} \quad (2)$$

互信息可以检测帧间的变化, 因此也能检测帧间的运动性质, 也就是说剧烈的运动对应着较低的互信息值, 而缓慢的运动则得出较高的互信息值.

## 3 类 haar 的 MCTF 编码方案

在本文提出的类 haar 的 MCTF 编码方案中, GOP 尺寸可以根据 MI 值自适应的选择, 在选定的 GOP 内部, 低通帧的位置同样根据帧间的 MI 值确定, 最后, 根据选定的 GOP 尺寸和低通帧位置来选择时间上的滤波方式, 下面将对这三个方面进行分别描述.

### 3.1 GOP 尺寸的选择

帧间剧烈的运动对应着较低的帧间 MI 值, 同时运动预测的消耗也较高, 反之亦然. 因此, 可以利用帧间的平均 MI 值来作为选择 GOP 尺寸的标准, 当帧间平均 MI 值较高时, 选择较长的 GOP 尺寸, 而较低的平均 MI 值对应着较短的 GOP 尺寸. 经过大量的实验, 表 1 给出平均 MI 值和 GOP 尺寸的关系, 其中参数  $low\_MI$ ,  $median\_MI$  和  $high\_MI$  是用于控制 GOP 尺寸的阈值.

表 1 平均 MI 值和 GOP 尺寸关系

平均 MI 值 (Average_MI)	GOP 尺寸
$Average\_MI < low\_MI$	4
$low\_MI < Average\_MI < median\_MI$	8
$median\_MI < Average\_MI < high\_MI$	16
$high\_MI < Average\_MI$	32

如果某段视频下 MI 值间的变化过大, 这表明该视频帧间的运动性质不一, 不宜划分到同一个 GOP 内. 因此, 在利用平均 MI 值来划分 GOP 的同时, 本文还使用帧间各 MI 值间的标准差来控制 GOP 的尺寸, 以防止同一个 GOP 内运动变化过快. 本文中所使用的标准差阈值为  $var\_T$ .

下面给出选择 GOP 尺寸的伪代码:

(1) 初始化:

$n = 0$ ;  
设置标准差阈值  $var\_T$ ;  
读取第一帧数据  $F_0$ ;

(2) 决定 GOP 尺寸

$n++$ ; 读取一帧数据  $F_n$ ;  
计算帧  $F_{n-1}$  和  $F_n$  的亮度分量的 MI 值  $MI_{n-1, n}$ ;  
计算 MI 值集合  $\{MI_{0,1}, MI_{1,2}, \dots, MI_{n-1, n}\}$  的平均 MI 值  $average\_MI$ ;

if ( $(average\_MI < low\_MI) \ \&\& \ (n > = 4)$ ) err\_code\_GOP();

else if ( $(low\_MI < = average\_MI < median\_MI) \ \&\&$

```

(n > 8) encode_GOP();
    else if ((median_MI <= average_MI < high_MI)
&& (n >= 16) encode_GOP();
    else if (average_MI >= high_MI) && (n >= 32))
encode_GOP();
    else
        计算MI值集合{MI0,1, MI1,2, ..., MIn-1,n}的标
        准差  $\sigma_n$ 
        if ( $\sigma_n >= var\_T$ ) encode_GOP();
        else goto (2);

```

所选择的 GOP 尺寸为  $n$ 。“encode\_GOP()”是对一个 GOP 进行编码的过程。

帧间 MI 值的亮度分量“ $MI_{n-1,n}$ ”可由等式(2)计算。

### 3.2 低通帧的选择

在传统的基于 MCTF 的小波编码方案中,低通帧的位置由时间上滤波类型决定。如图 1,在每个 GOP 内低通帧 LLL0 都在位置 F0。如果解码端只重构一层,则重构帧位置固定。但是,这些重构的帧时常并不是当前 GOP 的最佳代表。关键帧提取可以从一个视频段中提取一个或多个帧作为当前视频段的代表,该方法常用于视频检索中<sup>[8,9]</sup>。因此,关键帧提取技术同样可以用于 GOP 内低通帧的选择。本文引入文献[8]中从一个帧簇中提取关键帧的方法来从一个 GOP 内选择低通帧。在该方法中,某段视频段(GOP)内与其他帧间具有最大 MI 值的帧被选作为最具代表性的帧,见式(3):

$$F_{key} = \max_j \left( \frac{1}{N} \sum_{\substack{i=0 \\ i \neq j}}^{N-1} MI_{j,i} \right) \quad (3)$$

其中  $N$  是视频段的帧数目,在本文中,就是选定的 GOP。

$MI_{j,i}$  是帧  $F_j$  和  $F_i$  亮度分量的 MI 值,其中  $i$  和  $j$  是帧编号。

$F_{key}$  是从 GOP 选定的最具代表帧。

### 3.3 时间分解层次选择

在选择了 GOP 尺寸和低通帧位置后,由于 GOP 尺寸和低通帧都不再符合传统的基于 haar 的 MCTF 编码过程(图 1)。本文提出了一个类 haar 的 MCTF 编码方法,它可以根据选择的 GOP 尺寸和低通帧位置,自动确定时间上的编码层次和每层上运动预测的方向。下面以 GOP 尺寸为 14,低通帧位置是  $F_8$  举例说明,见图 2。

由图 2 可以看出,在该方法中位于低通帧前面的帧使用后向运动预测,而位于低通帧后面的帧使用前向运动预测。每层分解完毕后,剩余的帧都尽量与低通帧保持较短的距离,因此在下层继续分解时避免出现较大的预测残差,进而提高压缩性能。

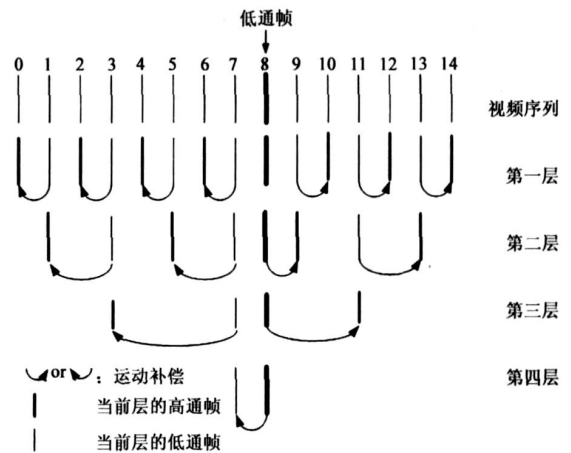


图 2 时间分解过程

## 4 试验结果及分析

本文针对不同运动类型的视频序列进行了大量的试验。本文选择 HVSBM<sup>[3]</sup> 中的变块范围为  $64 \times 64$  到  $4 \times 4$ 。MCTF 得到低通帧和高通帧由 MC-EZBC<sup>[7]</sup> 进行编码。另外,在本试验中使用了部分由文献[11]下载的参考软件。

表 3~7 给出了各个序列的试验结果,其中粗体部分为最佳的压缩性能。其中‘GOP8’和‘GOP16’分别表示 GOP 尺寸为 8 和 16 的传统 MCTF 编码方案的试验结果。试验中使用了两组 GOP 尺寸控制的阈值‘ADGOP1’和‘ADGOP2’,见表 2。

表 2 自适应选择 GOP 尺寸的参数集

名称	low_MI	median_MI	high_MI	var_T
ADGOP1	1.5	2.0	3.0	0.15
ADGOP2	1.4	1.9	3.1	0.14

符号‘XXX+KF’表示方法‘XXX’使用了低通帧选择

### 4.1 压缩性能比较

从试验结果可以看出,对于运动性质有明显变化的序列(Foran),或者有镜头切换的序列(Tennis),尤其对后者,使用自适应的选择 GOP 尺寸的方法能取得较好的压缩性能,如在相同的比特率下,Tennis 序列约能提高 0.8~1.0dB。而对运动性质变化不大的序列,如 Mobile, Stefan, Football, 自适应选择 GOP 尺寸的方法在压缩性能上略有降低。

相比于传统的 MCTF 编码方案,针对较为剧烈的运动序列,如 Stefan, Football, 本文提出的自适应低通帧选择方法可以取得更好的压缩效果,图像质量大约能提高 0.3~0.5dB。

图 3 给出了视频序列 Tennis 和 Football 在码率 1200kbps 情况下各帧的失真比较。在序列 Tennis 中,帧编号为 66 和 96 时对应着两次场景变换,因此也对应两个非常大的图像质量的下降。在本文提出的自适应的

选择 GOP 尺寸方法中, 图像质量得到了明显的提高, 见图 3(a) 箭头所指区域. 在图 3(b) 中比较了 GOP 尺寸同 为 8 时低通帧选择与固定低通帧方法的率失真情况, 经过选择低通帧, 整体图像质量提高了约 0.5dB.

表 3 Mobile 的率失真比较

码率 (kbps)	PSNR (dB)							
	GOP8	GOP8+ KF	<b>GOP16</b>	GOP16+ KF	ADGOP1	ADGOP1+ KF	ADGOP2	ADGOP2+ KF
600	26.99	26.96	<b>28.98</b>	28.95	28.76	28.75	28.98	28.95
800	29.00	28.96	<b>30.69</b>	30.65	30.50	30.48	30.69	30.64
1200	31.67	31.61	<b>32.99</b>	32.93	32.83	32.78	32.99	32.92
1600	33.52	33.46	<b>34.54</b>	34.41	34.41	34.32	34.54	34.41
2000	34.95	34.85	<b>35.93</b>	35.78	35.81	35.68	35.93	35.78

表 4 Foreman 的率失真比较

码率 (kbps)	PSNR (dB)							
	GOP8	GOP8+ KF	GOP16	GOP16+ KF	ADGOP1	<b>ADGOP1+ KF</b>	ADGOP2	<b>ADGOP2+ KF</b>
600	34.94	35.33	35.18	35.61	35.33	<b>35.76</b>	35.36	<b>35.80</b>
800	36.29	36.65	36.41	36.79	36.58	<b>37.02</b>	36.59	<b>37.03</b>
1200	38.28	38.59	38.32	38.63	38.48	<b>38.81</b>	38.47	<b>38.84</b>
1600	39.68	39.97	39.67	39.93	39.84	<b>40.16</b>	39.83	<b>40.17</b>
2000	40.87	41.11	40.84	41.10	40.98	<b>41.28</b>	40.97	<b>41.28</b>

表 5 Stefan 的率失真比较

码率 (kbps)	PSNR (dB)							
	GOP8	<b>GOP8+ KF</b>	GOP16	<b>GOP16+ KF</b>	ADGOP1	ADGOP1+ KF	ADGOP2	ADGOP2+ KF
600	28.45	<b>29.16</b>	28.63	<b>29.47</b>	28.54	29.04	28.55	29.14
800	30.30	<b>30.92</b>	30.40	<b>31.09</b>	30.32	30.80	30.33	30.89
1200	32.82	<b>33.38</b>	32.74	<b>33.29</b>	32.83	33.23	32.83	33.29
1600	34.58	<b>35.10</b>	34.46	<b>34.96</b>	34.62	35.03	34.61	35.07
2000	36.03	<b>36.52</b>	35.86	<b>36.32</b>	36.10	36.44	36.08	36.47

表 6 Football 的率失真比较

码率 (kbps)	PSNR (dB)							
	GOP8	<b>GOP8+ KF</b>	GOP16	GOP16+ KF	ADGOP1	ADGOP1+ KF	ADGOP2	ADGOP2+ KF
600	25.12	<b>25.67</b>	24.66	25.12	25.16	25.68	25.10	25.58
800	28.31	<b>28.83</b>	27.90	28.31	28.32	28.85	28.27	28.74
1200	31.38	<b>31.75</b>	30.89	31.29	31.21	31.76	31.18	31.64
2600	33.10	<b>33.59</b>	32.73	33.11	33.04	33.60	33.02	33.50
3000	34.21	<b>34.66</b>	33.92	34.19	34.15	34.64	34.16	34.58

表 7 Tennis 的率失真比较

码率 (kbps)	PSNR (dB)							
	GOP8	GOP8+ KF	GOP16	GOP16+ KF	ADGOP1	<b>ADGOP1+ KF</b>	ADGOP2	<b>ADGOP2+ KF</b>
400	29.84	30.02	30.38	30.64	31.34	<b>31.39</b>	31.31	<b>31.40</b>
800	32.96	33.16	33.51	33.72	34.50	<b>34.57</b>	34.48	<b>34.57</b>
1200	35.06	35.25	35.42	35.59	36.45	<b>36.57</b>	36.44	<b>36.56</b>
1600	36.71	36.85	37.05	37.16	37.99	<b>38.02</b>	37.98	<b>38.01</b>
2000	38.03	38.15	38.22	38.28	39.07	<b>39.13</b>	39.06	<b>39.13</b>

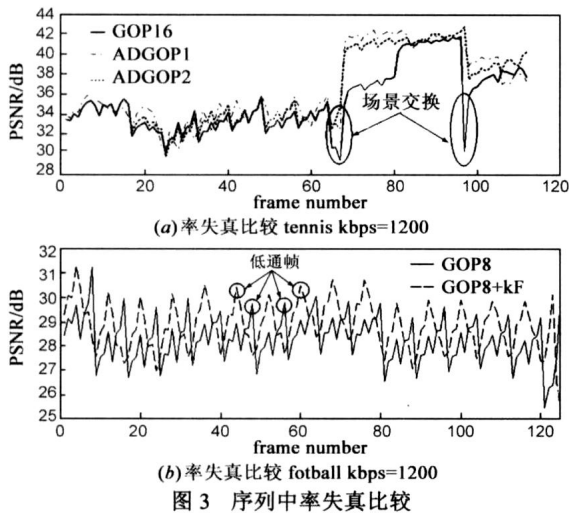


图3 序列中率失真比较

## 4.2 时间多层次分析

在本文的方法中, GOP 的选择根据运动情况自适应而定. 下面给出序列 Foreman 的试验结果. 使用表 2 给出的参数集 'ADGOP1', 可以得到 Foreman 序列的 GOP 尺寸顺序为{16, 13, 16, 16, 16, 16, 4, 32, 11, 16, 3, 10, 8, 8, 4, 4, 4, 4, 8, 8, 8, 14, 7, 16, 19, 15}. 当解码端只重构时间上一个层次, 图 4 中给出重构帧的比较.

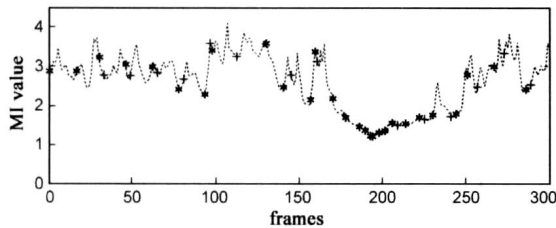


图4 固定 GOP 尺寸和自适应 GOP 尺寸的比较

在图 4 中, 曲线表示帧间的 MI 值, 即帧间的运动情况. '+' 标记了在 GOP 尺寸固定为 16 时可以重构的帧编号, 该方法重构的帧显然均匀分布在整个序列上. '\*' 标记了自适应的 GOP 方法重构的帧, 当然这些重构帧的分布是不均匀的. 运动平稳时(帧编号 110 左右)重构帧较少; 运动剧烈时(帧编号 200 左右)重构帧较多. 运动平缓的视频段帧间相差不大, 信息也较少, 因此较少的重构帧就能体现出原始序列的运动情况; 运动剧烈的视频段含有较多的信息, 因此使用较短 GOP 尺寸, 从而解码端可以在很低的时间层次上重构更多的帧数据, 尽可能的重现原始序列的运动过程. 因此, 在解码端重构层次较低时, 相对于固定 GOP 方法, 自适应的 GOP 更能体现出原始视频序列的运动过程, 不至于出现很大的帧间跳动, 从而视觉效果更流畅.

## 5 结论及后续工作

本文提出了一种类 haar 的 MCTF 编码方案, GOP 尺寸根据帧间的互信息的变换自适应的选择, 从而与视

频序列中的运动性质保持一致, 在选定的 GOP 内部, 根据互信息值来选择低通帧, 最后, 根据选定的 GOP 结构, 来确定时间上的分解层次和运动预测方向. 试验结果表明, 在运动性质变化较大的序列中, 本文提出的 GOP 尺寸方法可以在一定程度上提高压缩性能, 同时, 当解码端时间上重构的层次较少时, 本文的方法恢复的视频序列较传统方法更能体现出原始视频序列的运动情况. 另外, 在运动较为剧烈的视频序列中, 本文提出的低通帧的选择方法较之传统方法可以得到更好的压缩性能.

本文提出的计算 GOP 尺寸的参数集是经过大量的试验而得. 因此, 如何通过理论的方法来得到这些参数还需要进一步的探讨. 在本文提出的低通帧选择的方法中, 需要计算 GOP 内各帧间的互信息值, 计算量较大, 如何减少计算量以及使用其他的低通帧选择方法也需要进行研究.

## 参考文献:

- [1] Jens Rainer Ohm. Three dimensional subband coding with motion compensation[J]. IEEE Trans on image processing, 1994, 3 (5): 559- 571.
- [2] L Luo, J Li, S Li, Z Zhuang, Y Q Zhang. Motion compensated lifting wavelet and its application in video coding[A]. International conference on multimedia and Expo (ICME' 2001) [C]. Tokyo: IEEE, 2001. 365- 368.
- [3] S J Choi, J W Woods. Motion compensated 3 D subband coding of video[J]. IEEE Trans on image processing, 1999, 8(2): 155- 167.
- [4] D S Turaga, M van der Schaar, Y Andreopoulos, A Munteanu, P Schelkens. Unconstrained motion compensated temporal filtering (UMCTF) for efficient and flexible interframe wavelet video coding[J]. Signal processing on image communication, 2005, 20(1): 1- 19.
- [5] Hwangjun Song, Jongwon Kim, G-C Jay Kuo. Real time encoding frame rate control for H. 263+ video over the internet[J]. Signal processing on image communication, 1999, 15(1): 127 - 148.
- [6] Yir Lin Wang, Jing Xin Wang, Yerr wen Lai, Alvin W Y Su. Dynamic GOP structure determination for real time MPEG-4 advanced simple profile video encoder[A]. International conference on multimedia and Expo (ICME' 2005) [C]. Amsterdam: IEEE, 2005. 293- 296.
- [7] Peisong Chen, John W. Woods. Bidirectional MG-EZBC with lifting implementation[J]. IEEE Trans on circuits and system for video technology, 2004, 14(10): 982- 993.
- [8] Zuzana Cerneková, Ioannis Pitas, Christophoros Nikou. Information theory based shot cut/fade detection and video summarization[J]. IEEE Trans on image processing, 2004, 13(12): 2500- 2510.

rization[J]. IEEE Trans on circuits and system for video technology, 2006, 16(1): 82- 91.

- [9] Wengang Cheng, Yaniing Liu, De Xu. Shot boundary detection based on the knowledge of information theory[A]. International conference on neural networks and signal processing(ICNNSP' 2003)[C]. Nanjing: IEEE, 2003, 2(14): 1237- 1241.
- [10] 刘阳, 吴志美. 一种在 MPEG 压缩域中检测镜头扫换的

方法[J]. 电子学报, 2002, 30(5): 741- 744.

Liu Yang, Wu Zhi-Mei. A wipe transition detection method in MPEG compressed domain. Hsueh Pao/Acta Electronica Sinica, 2002, 30(5): 741- 744. (in Chinese)

- [11] Peisong Chen. Software package of MCGZBC wavelet coder is publicly available at <ftp://ftp.cipr.rpi.edu/personal/chen> [CP].

#### 作者简介:



刘兆广 男, 1977 年生于山东菏泽, 分别于 1998 年和 2001 年在西安理工大学获得工学学士学位和工学硕士学位, 现为山东大学信息科学与工程学院博士研究生. 主要研究方向为视频编解码、视频转码等.

E-mail: liuzhg1031@yahoo.com.cn



彭玉华 女, 1966 年生于山东郓城, 教授, 博士生导师, 中国电子学会会士、中国电子学会高级会员、电子测量与仪器学会微波毫米波测试专业委员会副主任委员、中国通信学会会员, 现为山东大学信号处理与计算机视觉研究所所

长, 研究兴趣包括多尺度分析、小波分析与数字信号处理、低信噪比信号分析、视频编解码技术、数字图像处理及在生物医学工程、生物特征识别领域的应用等.