

跳帧转码的运动矢量合成研究

曹 岗 张勇东 李锦涛 林守勋

(中国科学院计算技术研究所数字化技术研究室 北京 100080)
(caog@ict.ac.cn)

摘 要 由于终端设备、通信网络的异构性,视频服务需要提供跳帧转码功能,而直接使用输入码流运动矢量的视频跳帧算法导致视频质量下降严重.在前向主导运动矢量选择算法的基础上,讨论了运动矢量合成时运动矢量越界和主导帧内块不同处理方法对视频质量的影响,提出基于帧内刷新结构的前向主导运动矢量选择算法.实验结果表明,该算法可以有效阻断错误漂移,降低码率,提高视频质量.

关键词 视频转码;运动矢量合成;前向主导运动矢量选择;帧内刷新

中图分类号 TP391

Motion Vector Composition for Frame-Skipping Transcoder

CAO Gang, ZHANG Yong-Dong, LI Jin-Tao, and LIN Shou-Xun

(Digital Technology Laboratory, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080)

Abstract Video server should afford video frame-skipping transcoding for users because of clients' and network's heterogeneity. However, the video frame-skipping algorithm that directly reuses the motion vector from the incoming bit stream results in degraded video quality. Based on a forward dominant vector selection (FDVS) algorithm, discussed is the impact of different processing methods for motion vector out-of-range and forward intra dominant macroblock after the motion vector composition. An intra-refresh FDVS video transcoding algorithm is proposed. Experimental results show that the algorithm can stop the error propagation efficiently, decrease the bit rate of video stream, and increase the video quality.

Key words video transcoding; motion vector composition; FDVS; intra refresh

1 引 言

随着视频压缩标准、网络技术的发展,视频应用逐渐普及.采用的终端设备、网络数量和类型不断增加,网络状况不稳定,终端设备的异构性严重制约视频服务发展,无法为视频应用提供服务质量保证.视频服务需要转码技术^[1]屏蔽通信网络、终端设备的异构性为用户提供通用媒体存取接口 UMA (universal media access)^[2].把某种编码格式的压缩码流转换为另一种编码格式的压缩码流的操作称为视频转换编码(video transcoding,简称为视频转码),完成视频转码的软件或者设备称为视频转换编

码器(video transcoder,简称为转码器),转码器部署在编码器和解码器之间,如图 1 所示:

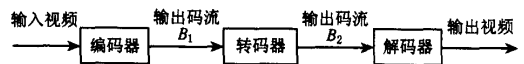


图 1 转码器结构图

码器的直接实现方式就是像素域级联转码器^[3]:直接解码输入码流,然后对解码结果再重新编码,就是一个标准解码器串联一个标准编码器.像素域级联转码非常灵活,没有错误传播,作用同编码器相当,可以实现视频编码器的任何功能.缺点是由于需要重新进行运动估计,计算复杂度大,不适合实时转码.所以,有学者提出通过复用输入码流

收稿日期:2003-04-06;修回日期:2003-07-14

基金项目:国家自然科学基金项目(60302028);国家“八六三”高技术研究发展计划重点课题基金项目(2001AA114060)

中的运动矢量,降低像素域级联转码器的计算复杂度^[4~6],而对视频质量的影响是有限的。

对于码率调整型转码器,如果传输网络带宽变化剧烈或者带宽极低,转码器必须通过跳帧技术略过部分输入编码帧,以保证输出码率符合信道要求。跳帧技术是指转码器对部分输入编码帧没有输出。对于异构客户端,如果客户端不支持高帧率,转码器也必须跳帧。由于时域编码预测技术,跳帧使参考帧发生变化,简单复用输入码流的运动矢量导致视频质量下降严重^[6],所以转码器必须重新计算运动矢量。

双线性插值(bilinear interpolation)^[4,6]运动矢量合成算法用双线性插值的方法,利用参考帧中同当前块相关的4个块的运动矢量合成新的运动矢量。在时域上,双线性差值算法从后向前处理,依次合成每个丢弃参考帧的运动矢量,直到最后一个没有被跳过的数据帧。双线性插值算法需要保留每个被跳过的参考帧。前向主导运动矢量选择算法(forward dominant vector selection,FDVS)^[7]从参考帧中选择与当前块重叠区域最大的块作为主导宏块,主导宏块的运动矢量同当前宏块的运动矢量相加,结果作为当前块的运动矢量。实验结果表明,FDVS算法的视频质量优于双线性插值算法,并且从前向后处理,不需要保存每个被丢弃的参考帧。

FDVS算法存在两个问题:第1,越界问题(out-of-range problem)。在运动矢量的合成中,运动矢量完全有可能越界,对于越界的处理算法不同,则结果必然有很大差异,原FDVS算法未提及如何处理越界问题;第2,主导帧内块问题(forward intra macroblock)。所谓主导帧内块是指主导宏块为帧内编码块,同样,对主导帧内块的处理方法不同,结果也必然有很大

差异,原FDVS算法将帧内编码块的运动矢量设为(0,0),使用FDVS算法合成运动矢量,然后重新计算宏块编码类型。本文通过分析越界问题和主导帧内块问题的各种解决方法,指出矢量越界的不同处理方法对视频质量的影响程度远大于主导帧内块处理方法,提出基于帧内刷新结构的前向主导运动矢量选择算法IR-FDVS(intra-refresh FDVS)。实验结果表明,该算法可以有效阻断错误漂移,降低视频码流的码率,提高视频质量。

本文第2节是前向主导运动矢量选择算法的详细分析,给出运动矢量合成时运动矢量越界以及主导帧内块的各种处理方法,并给出不同处理方法影响视频质量、码率的实验结果;第3节通过分析实验数据,给出基于帧内刷新结构的前向主导运动矢量选择算法,最后是结论及将来的工作。

2 前向主导运动矢量选择算法分析

2.1 前向主导运动矢量选择算法介绍

在混合离散余弦变换(discrete cosine transform, DCT)和运动补偿(motion compensation, MC)技术的分块视频编码算法中,运动补偿降低时域相关度,提高压缩码率。对于预测编码帧,编码器通过在参考帧的某个范围内搜索,按照某种规则,得到同当前块最相似的参考块,当前块减去参考块即得到预测残差块;按照率失真理论,选择帧内、帧间编码模式。对于帧间编码宏块,解码器通过运动矢量,找到参考帧中的对应参考块,残差块加上参考块就得到当前块的重建块。

在极低码率下,或者终端不支持高帧率时,部分数据帧必须被跳过。由于运动补偿技术,参考帧被

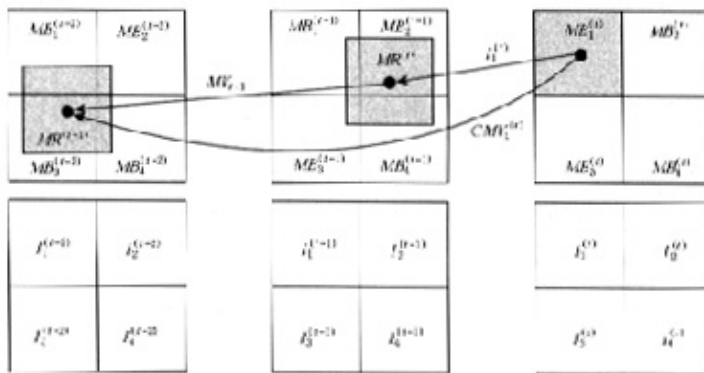


图2 跳帧示意图

跳过,当前块的运动矢量不再是最优的,必须通过运动矢量合成技术重新合成当前块的运动矢量,如图2所示.定义 $E^s = \{1, 2, \dots, s\}$, s 是一个正整数.其中 $MB_i^{(j)}$ 是第 j 帧中的宏块 i , 其对应的运动矢量为 $I_i^{(j)}$, $i \in E^s$, $j \in \{t-2, t-1, t\}$. 显然,当前块 $MB_1^{(t)}$ 可以通过 $I_1^{(t)}$ 确定其参考块 $MR^{(t)}$. 假设 $MR^{(t)}$ 的运动矢量为 MV_{t-1} , $MB_1^{(t)}$ 的合成运动矢量 $CMV_1^{(t)}$ 为

$$CMV_1^{(t)} = MV_{t-1} + I_1^{(t)}. \quad (1)$$

如果 $MR^{(t)}$ 同宏块边界不重叠,则 MV_{t-1} 不存在,无法合成 $CMV_1^{(t)}$. 文献[7]中提出前向导运动矢量选择算法 FDVS. FDVS 从同参考块相关的4个参考宏块中选取主导宏块,将该宏块的运动矢量作为主导运动矢量,再合成 $CMV_1^{(t)}$. 主导宏块是指同参考块重叠区域最大的宏块. 例如如图2,主导宏块是 $MB_2^{(t-1)}$. 所以, $MV_{t-1} = I_2^{(t-1)}$, $CMV_1^{(t)} = I_2^{(t-1)} + I_1^{(t)}$.

2.2 运动矢量越界

运动估计在一定范围中进行,所以运动矢量的取值范围有限^[8-10]. 例如,对于 H.263^[8] 编码标准,在缺省情况下,运动矢量的合法取值范围是 $[-16, 15.5]$; 如果允许使用非受限运动矢量 (unrestricted motion vector), 对于 QCIF 和 CIF 图像,运动矢量的合法取值范围是 $[-32, 31.5]$. 在 FDVS 算法中,运动矢量合成结果 $CMV_1^{(t)}$ 不可避免地存在越界情况,不在合法取值范围内,对于这种情况我们称为运动矢量越界问题.

原 FDVS 算法未讨论运动矢量越界问题,没有给出运动矢量越界的处理方法. 针对运动矢量越界问题,本文给出如下几种处理方法.

(1) C: 裁剪 (clip) 越界运动矢量,合成运动矢量设为运动矢量取值范围的边界值,作为当前块的运动矢量,宏块编码类型不变. 例如对于运动矢量的取值范围为 $[-16, 15.5]$, 如果合成后的运动矢量为 -18 , 则设置为 -16 ;

(2) CI: 如方法1的裁剪操作,重新计算宏块编码类型,按照帧内编码和帧间编码的模式选择算法,例如绝对帧差是否小于某个阈值,判断是否修改宏块编码类型(下同);

(3) CFI: 强制帧内,将当前块的编码模式设置为帧内编码;

(4) CO: 不进行越界处理,直接使用当前块的输入运动矢量,不重新计算宏块编码类型;

(5) COI: 不进行越界处理,直接使用当前块的输入运动矢量,重新计算宏块编码类型.

2.3 主导帧内块

前向导块可能为帧内编码块,对于这种情况我们称为主导帧内块问题. 如果前向导块为帧内编码块,FDVS 将帧内编码块的运动矢量设置为 $\langle 0, 0 \rangle$, 合成运动矢量,重新计算当前块的编码模式. 实际上,在预测帧中存在帧内编码块的重要原因是预测编码模式已经无法更好预测原始数据,需要的码率高于帧内编码模式,或者视频质量比帧内编码模式差. 而预测编码的性质决定了在一定搜索范围内,参考块与当前块最相似. 所以本文对原文算法进行进一步研究.

对于前向导块是帧内编码块,本文提供5种处理方法:

(1) I: 帧内编码块的运动矢量设为 $\langle 0, 0 \rangle$, 使用 FDVS 合成运动矢量,合成运动矢量作为当前块的运动矢量,不重新计算宏块类型;

(2) II: 帧内编码块的运动矢量设为 $\langle 0, 0 \rangle$, 使用 FDVS 合成运动矢量,合成运动矢量作为当前块的运动矢量,重新计算宏块类型. 该方法与文献[7]所示处理方法相同;

(3) IFI: 强制帧内,将当前块的编码模式设置为帧内编码;

(4) IO: 直接使用当前块的输入运动矢量,不重新计算宏块类型;

(5) IOI: 直接使用当前块的输入运动矢量,重新计算宏块类型.

2.4 处理方法

组合矢量越界问题和主导帧内块的不同处理方法,定义“方法” i 如下:

$$i = n \times 5 + m, \quad (2)$$

$n \in \{0 \cup E^4\}$, $n = 0, 1, 2, 3, 4$ 分别对应运动矢量越界的 C, CI, CFI, CO, COI 处理方法; $m \in E^5$, $m = 1, 2, 3, 4, 5$ 分别对应主导帧内块的 I, II, IFI, IO, IOI 处理方法. 其中 $n \times 5 + 2$ 是采用 n 对应的矢量越界处理方法,原文中提出的主导帧内块处理方法.

定义特殊处理方法“0”. 当 $i = 0$ 时,表示直接使用当前块的运动矢量. 如图2所示,对于块 $MB_1^{(t)}$,合成运动矢量 $CMV_1^{(t)} = I_1^{(t)}$,从参考帧 $(t-2)$ 确定其参考块. 虽然像素域视频转码器没有错误漂移^[9],但预测块不是当前块的最佳匹配块,所以引入较大误差.

2.5 实验结果

从 A, B, C 三类^[10] 测试序列中各选取一测试

序列. A 类测试序列是简单纹理,低运动视频;B 类测试序列是中度复杂纹理,低运动视频,或者低复杂度纹理,中度运动视频;C 类测试序列是复杂纹理,中度运动视频,或者中度复杂纹理,复杂运动视频. 本文分别测试了 QCIF 格式(176×144)的 A 类测试序列 Hall Monitor, B 类测试序列 Foreman, C 类测试序列 Table Tennis. 对于上述 3 个序列,各取 300 帧数据,用 H.263 编码器[11]编码,量化参数 QP (quantization parameter)为 7,运动矢量取值范围是 $[-16, 15.5]$,帧率为 30fps(frame per second);将上述码流转码为 10fps 的码流,即每编码一帧,跳过两帧. 跳帧算法分别采用 $i, i \in \{0 \cup E^{25}\}$ 对应的矢量越界和主导帧内块处理方法.

表 1 给出矢量越界块和主导帧内块数量. 对于 A 类序列 Hall Monitor, 转码时仅有 5 个块的矢量越界, 16 个主导帧内块,这是同 A 类序列纹理简单、运动比较小的特点一致,由于越界块和主导帧内块数量较小,所以无论采用什么方法,平均每帧视频质量差别不超过 0.01dB,码率差异不超过 50 比特. 对于中复杂度序列(B 类序列)Foreman, 转码时有 710 个块的矢量越界, 130 个主导帧内块,越界块和主导帧内块数量较多,不同实现方法对视频质量、码率影响较大;对于复杂序列(C 类序列)Table Tennis, 转码时有 222 个块的运动矢量越界和 17 个主导帧内块,运动矢量越界块数量大,而主导帧内块数量较小,同样,矢量越界的不同实现方法严重影响

视频质量.

表 1 矢量越界块和主导帧内块数量

序列	矢量越界块	主导帧内块
Hall Monitor	5	16
Foreman	710	130
Table Tennis	222	17

定义 $Rate(i)$ 和 $PSNR(i)$ 分别表示对应 i 方法的所有帧平均码率、平均峰值信噪比 PSNR(peak signal-to-noise ratio). 图 3 给出采用对应 $i \in E^{25}$ 运动矢量越界和主导帧内块处理方法的 Foreman 序列测试结果. 码率图 3(a)纵坐标表示 $Rate(i)$;峰值信噪比图 3(b)纵坐标表示 $PSNR(i)$. 码率变化范围在 13060.85~13627.23 bit, PSNR 变化范围在 32.11~32.31 dB,可见不同的实现方法对视频质量、码率影响巨大. 定义:

$$I\text{-PSNR} = \max_{k \in \{0 \cup E^4\}} (\max_{j \in k \times 5 + l} (PSNR(j)) - \min_{j \in k \times 5 + l} (PSNR(j))), l \in E^5, \quad (3)$$

$$C\text{-PSNR} = \max_{k \in \{0 \cup E^4\}} \left(\sum_{j=k \times 5 + 1}^{k \times 5 + 5} PSNR(k) \right), \quad (4)$$

$C\text{-PSNR}$ 和 $I\text{-PSNR}$ 分别反映了在运动矢量越界和主导帧内块处理方法相同时, $PSNR(i)$ 的变化距离程度. $C\text{-PSNR} = 0.17$ dB, $I\text{-PSNR} = 0.04$ dB. 显然,矢量越界的不同处理方法对视频质量的影响程度大于主导帧内块处理方法.

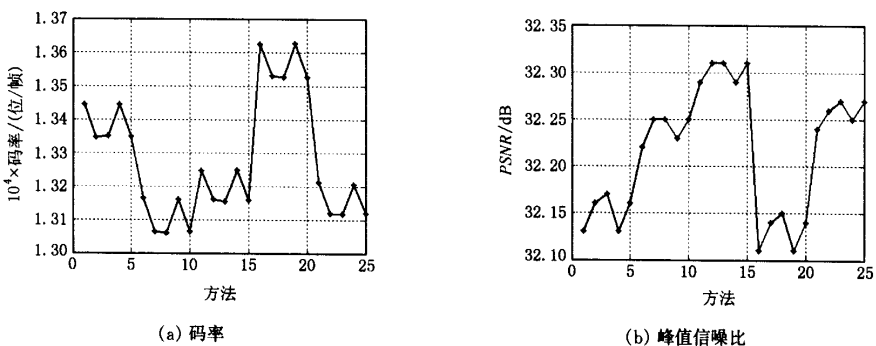


图 3 不同处理方法的测试结果

综上所述,对于低运动、纹理简单的序列,运动矢量越界、主导帧内块的处理算法对结果基本没有影响;对于纹理复杂或者运动剧烈的视频序列,如果运动矢量越界、主导帧内块的处理算法不同,FDVS 的结果相差甚远,必须考虑运动矢量越界、主导帧内块对视频质量的影响;矢量越界的不同处理方法对

视频质量的影响程度远大于主导帧内块处理方法.

3 基于帧内刷新结构的 FDVS

3.1 基于帧内刷新结构的 FDVS

定义峰值信噪比增益 $\text{GPSNR}(\text{gain of PSNR})$

和码率增益 $GR(\text{gain of rate})$:

$$GPSNR(i) = PSNR(i) - PSNR(0), \quad (5)$$

$$GR(i) = Rate(0) - Rate(i), \quad (6)$$

其中, $i, i \in E^{25}$ 表示对应的处理方法. $GPSNR(i)$ 和 $GR(i)$ 分别表示相对于没有运动矢量合成时的 PSNR 增加、码率下降. 显然, 对于 i 处理方法, $GPSNR(i)$ 值越大, 则视频质量越好; $GR(i)$ 值越大, 则码率越低. 视频编码的目标是使用尽量少的码率达到最佳视频质量, 所以运动矢量合成算法应该尽量提高 $GPSNR(i)$ 和 $GR(i)$.

表 2 峰值信噪比和码率增益

(a) Foreman

Foreman		II	IFI	IOI
CI	GPSNR	0.74	0.74	0.74
	GR	6367.76	6370.42	6367.11
CFI	GPSNR	0.80	0.80	0.80
	GR	6270.06	6276.85	6271.68
COI	GPSNR	0.75	0.76	0.76
	GR	6311.52	6314.42	6311.11

(b) Table Tennis

Table Tennis		II	IFI	IOI
CI	GPSNR	0.44	0.46	0.42
	GR	3194.18	3198.87	3196.12
CFI	GPSNR	0.48	0.48	0.49
	GR	3165.74	3165.66	3160.48
COI	GPSNR	0.45	0.46	0.45
	GR	3193.05	3181.82	3186.91

越界方法采用 CI, CFI, COI 的处理结果显著优于 C, CO, 帧内主导块采用 II, IFI, IOI 的处理结果显著优于 I, IO. 表 2 给出上述实验的部分峰值信噪比增益 $GPSNR(i)$ 和码率增益 $GR(i)$. 对于 Foreman 序列, $GR(i)$ 变化范围在 6270.06 ~ 6370.42 bit, $GPSNR(i)$ 增加 0.74 ~ 0.80 dB; 对于 Table Tennis 序列, $GR(i)$ 变化范围在 3160.48 ~ 3198.87 bit, $GPSNR(i)$ 增加 0.42 ~ 0.46 dB. 显然, 矢量越界方法采用 CI, CFI, COI, 帧内主导块方法采用 II, IFI, IOI, $GPSNR(i)$ 和 $GR(i)$ 基本相同.

表 3 给出编码模式发生变化的矢量越界块和主导帧内块分别占全部矢量越界块数量和主导帧内块数量的比例, 分别表示为 *Clip* 和 *Intra*. 对于主导帧内块处理方法, 由于宏块模式变化的比例均大于 82.35%, 并且主导帧内块处理算法对视频质量的影

响较小, 所以对于主导帧内块无须采用原文献[7]中采用的重新计算编码模式的方法, 可以直接采用帧内编码模式, 即 IFI 方法, 可以降低算法的计算复杂度, 又不影响视频质量、码率.

对于 Foreman 序列, CI 的码率增益最大; 对于 Table Tennis 序列, CFI 码率最多比其他方法低 35.64 bit. 对于 CI 和 COI 矢量越界处理方法, 有一半左右的宏块编码模式没有发生变化, 视频质量和码率接近于 CFI, 虽然码率可能略低于 CFI, 但是 CFI 取得最佳的 PSNR 增益. 所以, 对于矢量越界问题, 当前块的编码模型应该为帧内编码, 即采用 CFI 方法.

表 3 宏块模式变化百分比

(a) Foreman

Foreman		II	IFI	IOI
CI	Intra	86.15	100.00	90.00
	Clip	49.01	49.27	49.01
CFI	Intra	86.15	100.00	90.00
	Clip	100.00	100.00	100.00
COI	Intra	85.39	100.00	90.00
	Clip	67.61	67.32	67.61

(b) Table Tennis

Table Tennis		II	IFI	IOI
CI	Intra	82.35	100.00	82.35
	Clip	45.95	45.95	46.37
CFI	Intra	82.35	100.00	82.35
	Clip	100.00	100.00	100.00
COI	Intra	82.35	100.00	82.35
	Clip	52.70	52.70	52.70

根据上述分析, 本文提出基于帧内刷新结构的前向主导运动矢量选择算法 IR-FDVS (intra-refresh FDVS):

(1) 跳帧时, 采用前向主导运动矢量选择转码算法进行运动矢量合成;

(2) 主导帧内块直接强制为帧内编码模式;

(3) 运动矢量越界块直接强制为帧内编码模式.

3.2 实验结果

将“Foreman”序列编码为 128 Kbps, 30 fps; 转码结果码流为 50 Kbps, 10 fps. 转码时, 矢量越界块和主导帧内块采用 IR-FDVS 算法, IR-FDVS 的峰值信噪比增益 $GPSNR(13)$ 为 2.32 dB. 图 4 是第 100 帧视频的主观视频质量, IR-FDVS 算法的峰值信噪比高出直接复用输入运动矢量(方法“0”)2.60 dB.



(a) 方法“0”, PSNR = 26.15 dB



(b) IR-FDVS, PSNR = 28.75 dB

图 4 Forman 第 100 帧视频质量比较

4 结论及将来的工作

由于终端设备、通信网络的异构性,视频服务需要为用户提供转码功能.在码率极低或者终端不支持高帧率时,必须采用跳帧技术.视频跳帧算法直接使用输入码流的运动矢量导致视频质量下降严重.在前向导运动矢量选择算法的基础上,本文分析了矢量越界和主导帧内块问题,以及对不同复杂度视频的影响,提出基于帧内刷新结构的前向导运动矢量选择算法,该算法可以有效阻断错误漂移,降低码率,提高视频质量.

本文讨论了运动矢量越界和主导帧内块问题,分析了实验结果,需要从理论上进一步分析跳帧中运动矢量合成、残差计算问题,这将是今后我们跳帧转码研究的方向.

参 考 文 献

- 1 A Vetro, C Christopoulos, S Huifang. Video transcoding architectures and techniques: An overview. *IEEE Signal Processing Magazine*, 2003, 20(2): 18~29
- 2 R Mohan, J R Smith, C S Li. Adapting multimedia Internet content for universal access. *IEEE Trans on Multimedia*, 1999, 1(1): 104~114
- 3 G Keesman, R Hellinghuizen, F Hoeksema *et al.* Transcoding of MPEG bitstreams. *Signal Processing: Image Communication*, 1996, 8(6): 481~500
- 4 B Shen, I K Sethi, B Vasudev. Adaptive motion-vector resampling for compressed video downscaling. *IEEE Trans on Circuit and Systems for Video Technology*, 1999, 9(6): 929~936
- 5 T Shanableh, M Ghanbari. Heterogeneous video transcoding to lower spatio-temporal resolution and different encoding formats. *IEEE Trans on Multimedia*, 2000, 2(2): 101~110
- 6 N Bjork, C Christopoulos. Transcoder architecture for video

coding. *IEEE Trans on Consumer Electronics*, 1998, 44(1): 88~98

- 7 J Youn, M T Sun, C W Lin. Motion vector refinement for high performance transcoding. *IEEE Trans on Multimedia*, 1999, 1(1): 30~40
- 8 ITU-T Rec H.263. Video Coding for Low Bit Rate Communication. 1998
- 9 J Youn, M T Sun, J Xin. Video transcoder architectures for bit rate scaling of H.263 bit streams. In: *Proc of the 7th ACM Int'l Conf on Multimedia*. New York: ACM Press, 1999. 243~250
- 10 ISO/IEC 14496-2. Coding of Audio-Visual Objects — Part 2: Visual. 2001
- 11 ITU-T/SG15. Video Codec Test Model, TMN8, 1997



曹 岗 男,1975 年生,博士研究生,主要研究方向为视频编码、视频转码、多媒体通信.



张勇东 男,1973 年生,博士,助理研究员,主要研究方向为多视角视频合成、视频检索、转换编码.



李锦涛 男,1962 年生,研究员,博士生导师,主要研究方向为数字化技术、视频内容处理、多媒体网络传输.



林守勋 男,1948 年生,研究员,博士生导师,主要研究方向为多媒体技术、数据库.