

视频码流转换编码器结构及其分析

郭 映 李桂苓 徐 岩

(天津大学电信学院, 天津 300072)

摘 要 随着数字音视频及多媒体技术的发展, 迫切需要压缩巨大的视频信号数据, 适用于不同领域的 MPEG-1、MPEG-2、MPEG-4、H. 261、H. 263、H. 264 等国际编码标准应运而生。为了适应用户终端的多样性以及网络自身的传输特性, 常需将已压缩的视频码流进行转换编码, 以实现网络间资源共享。主要介绍了对已有的压缩视频流根据目标的要求如何进行相应的变换, 如通过改变量化参数获得码率的变换; 通过下变换获得缩小的图像, 实现空间分辨率的变换, 以及跳帧获得时间分辨率的变换。对不同变换器结构形式进行了分析, 以获得结构复杂性和图像质量的均衡, 并重点剖析了转换编码器核心技术——运动矢量的确定。

关键词 转换编码 漂移误差 运动矢量

中图分类号: TN911 文献标识码: A 文章编号: 1006-8961(2004)10-1153-07

Architectures and Analysis about Video Transcoder

GUO Ying, LI Gui-ling, XU Yan

(Electronic and Information Engineering School, Tianjin University, Tianjin 300072)

Abstract Along with digital AV and multimedia technology developing, large numbers of video datum need to be compressed. At present, the International Coding Standards, such as MPEG-1, MPEG-2, MPEG-4, H. 261, H. 263, H. 264 have been established and applied to different application areas. In order to adapt to heterogeneous client terminals and transmission characteristic of networks, pre-encoded video should be transcoded into another coding format. So, the resources of networks can be shared. The content of this paper is about how to convert a previously compressed video bitstreams into a lower bit-rate video bitstreams according to the requirements of heterogeneous client terminals, such as bitrate transform by changing quantizer parameter, spatial resolution transform by down-sampled, temporal resolution transform by skipping frames. In this paper the architectures and mechanism of transcoder are analyzed and the best tradeoff between quality and complexity are offered. Also the core technique; how to obtain the motion vector is anatomied.

Keywords transcoding, drift error, motion vector

1 引言

多媒体技术实用化的关键技术之一是解决数字视频、音频数据量大与数字存储媒体承载能力和通信网络带宽的矛盾, 其解决途径是压缩。为此, 国际上先后制定了视频、音频编码标准: MPEG-1 和 MPEG-2。这两个标准的制定使得基于 CD-ROM 的数字视频和数字电视成为可能。而为了满足视频会议、可视电话等低比特率传输的需要, 又出现了

H. 261、H. 263、MPEG-4 等编码标准, 其中 MPEG-4 标准支持更大范围的目标视频流码率, 为个人移动通信、卫星网络系统和跨网络环境应用提供了解决方案。

目前, 新一代的视频编码标准 H. 264 已经制订完成。该标准的编码框架与现有的 MPEG-1、MPEG-2 视频标准一致, 都是基于块的混合编码。但是它采用了成熟的技术, 如: 不同大小和形状的宏块分割、多参考图像的运动预测、帧内预测、去块效应滤波器, 在追求更高的编码效率和简洁表达形

基金项目: 天津自然科学基金(033600511); 天津大学教学振兴计划

收稿日期: 2003-08-05; 改回日期: 2004-06-04

式的同时,也提供了良好的视频质量,是当前最高效但也更加复杂的视频压缩方法。

这些应用于不同领域的编码标准,在语法、压缩效率、输出码率、主要编码技术上存在着差异,但它们具有相似的压缩算法,如:帧间运动补偿(MC)、帧内离散余弦变换(DCT)、量化(Q),或在局部层面上,互成特例。这些相似性为支持不同系统、实现网络间资源共享的转换编码创造了条件。

编码流的转换可以是同类码流的转换,如 MPEG-2 到 MPEG-2, H. 263 到 H. 263, 由于是同类转换,一般不用考虑语法的问题,只进行结构上的相应变换;也可以是异类码流间的转换,如 MPEG-2 到 H. 263, MPEG-2 到 MPEG-4。这类转换由于涉及到不同的编码标准,不仅要考虑语法和结构问题,同时也要考虑两种标准间是否有兼容的部分,如涉及 MPEG-4 的异类码流转换编码限于简单类、矩形 VO(视觉对象)情况,在这种情况下 MPEG-4 与传统 MPEG-1、MPEG-2、H. 263 标准兼容。如果涉及 H. 264,虽然该标准也是基于块的混合编码,但采用了更先进、更复杂的技术,所以兼容性问题也需要考虑。目前转换编码主要有码率变换、空间分辨率变换和时间分辨率变换 3 类。

2 码率变换转换编码器

码率变换靠改变量化系数来降低编码码流的码率,其有多种结构。结构 1 为解码-编码级联型如图 1 所示,即先解码视频流,再根据目标码率重新编码(图中, F 为帧存储器,下标 1 表示输入,下标 2 表示输出)。该结构灵活,无图像质量漂移,但由于运动矢量(MV)和宏块编码模式需重新确定,所以计算量大,不利于实时处理。为简化计算、确保图像质量,重新编码时,可使用输入码流中的运动矢量和宏块编码模式。但将高码率码流中的信息直接用于低码率编码会损伤图像,需进行运动矢量优化处理。结构 2

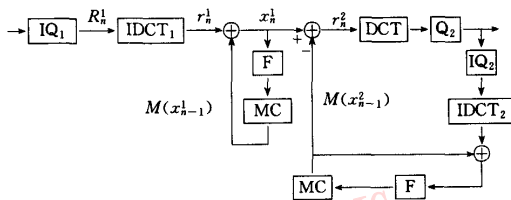


图 1 解码-编码级联型码率变换器

为图 2 所示的开环型码率变换器。变字长解码(VLD)后的 DCT 系数,经反量化(IQ)处理,再按目标码率重新量化,最后进行变字长编码(VLC)获得输出码流。另外有一种直接截除高频 DCT 系数的开环结构,运算量更小。二者虽形式简单、运算量小,但因丢失高频信息,解码图像质量损伤在图组(GOP)内累加,从而加重了漂移现象。

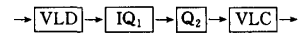


图 2 开环型码率变换器

对比结构 1、结构 2,前者复杂,后者图像质量差。由结构 1 得

$$r_n^2 = x_n^1 - M(x_{n-1}^1) = r_n^1 + M(x_{n-1}^1) - M(x_{n-1}^2) = r_n^1 - M(e_{n-1}^2) \quad (1)$$

其中, R, r 为剩余信息(即预测误差),大写字母代表 DCT 域,小写字母代表像素域, e 为误差, $M()$ 为运动补偿,上标 1 表示输入信号,上标 2 表示输出信号。

由 DCT/IDCT(反 DCT)的线性得

$$R_n^2 = R_n^1 - \text{DCT}[M(e_{n-1}^2)] \quad (2)$$

据此可得结构 3,即简化的像素域转换编码器 SPDT 如图 3 所示。这种方法在重新编码时直接使用输入码流中的运动矢量和宏块编码模式,即假设运动矢量、宏块编码模式在编码器与解码器中保持一致,并且假设运动补偿是线性的: $M(x_{n-1}^1) - M(x_{n-1}^2) = M(x_{n-1}^1 - x_{n-1}^2)$, DCT/IDCT 是线性的。

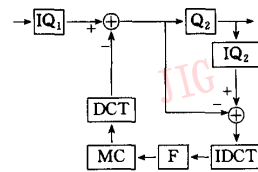


图 3 简化的像素域转换编码器

若在 DCT 域做运动补偿(DCT-MC),可进一步简化,得图 4 所示结构。

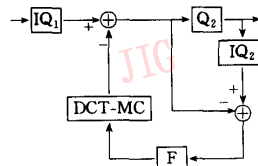


图 4 DCT 域转换编码器

3 空间分辨率变换编码器

空间分辨率变换指按需要缩小图像尺寸。结构 1 为闭环级联型空间分辨率转换器如图 5 所示。它类似码率变换器结构 1, 只多一下变换处理单元, 同样表现出结构复杂、计算量大、无漂移的特点, 一般作为参考结构。

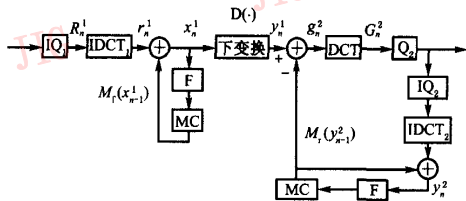


图 5 闭环级联型空间分辨率转换器

结构 2 是开环型空间分辨率转换器如图 6 所示, (图中 MV_i 为全分辨率运动矢量, MV_r 为下变换后运动矢量)。与图 2 相比, 因变换空间分辨率, 故运动矢量等参数不能直接使用, 需映射处理。以 1/2 下变换为例, 4 个宏块对应下变换后的一个宏块, 因此, 4 个运动矢量要映射为一个运动矢量; 若目标编码标准不支持一个宏块存在混合编码模式, 而 4 个宏块编码模式又不同时, 则需重新确定一个编码模式作为下变换后的宏块使用。

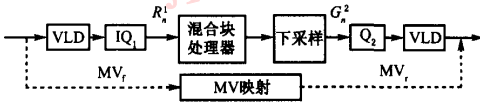


图 6 开环型空间分辨率转换器

由结构 1 得

$$g_n^2 = D(x_n^1) - M_r(y_{n-1}^2) = D(r_n^1) + D(M_t(x_{n-1}^1)) - M_r(y_{n-1}^2) \quad (3)$$

式中, $D(\cdot)$ 表下采样, $M_t(\cdot)$ 表全分辨率运动补偿, $M_r(\cdot)$ 表下变换后运动补偿, R, r 为全分辨率预测误差, G, g 为下变换后预测误差。大写字母表示 DCT 域, 小写字母表示像素域。

由结构 2 得

$$g_n^2 = D(r_n^1) \quad (4)$$

与结构 1 比较, 误差

$$d = D(M_t(x_{n-1}^1)) - M_r(y_{n-1}^2)$$

分解

$$d = D(M_t(x_{n-1}^1)) - M_r(y_{n-1}^2) + M_r(y_{n-1}^2) - M_r(y_{n-1}^2) = d_r + d_q$$

其中,

$$\begin{aligned} d_r &= D(M_t(x_{n-1}^1)) - M_r(y_{n-1}^2) \\ &= D(M_t(x_{n-1}^1)) - M_r(D(x_{n-1}^1)) \\ d_q &= M_r(y_{n-1}^2) - M_r(y_{n-1}^2) \end{aligned}$$

误差分为两部分: 下变换引起的误差 d_r ; 量化引起的误差 d_q 。

上述分析表明, 级联型结构复杂, 恢复的图像质量好; 开环型结构简单, 漂移大。根据输入序列的不同, 可采取不同补偿方式, 以折衷结构复杂度和图像质量^[1]。

补偿方式 1

$$\begin{aligned} \text{设 } D(M_t(x_{n-1}^1)) &\approx M_r(D(x_{n-1}^1)) = M_r(y_{n-1}^2), \\ \text{则 } g_n^2 &= D(r_n^1) + M_r(y_{n-1}^2) - M_r(y_{n-1}^2) \\ &= D(r_n^1) - M_r(e_{n-1}^2) \end{aligned} \quad (5)$$

由式(5)得结构 3, 漂移补偿型空间分辨率下变换器如图 7 所示。与图 6 相比, 量化引起的误差 d_q 得到了补偿, 采用下变换后补偿, 漂移误差只源于下变换 d_r 。

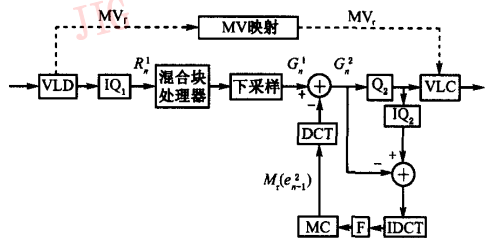


图 7 漂移补偿型空间分辨率下变换器

补偿方式 2

$$\begin{aligned} \text{若 } M_r(y_{n-1}^2) &= M_r(D(x_{n-1}^1)) \approx D(M_t(x_{n-1}^1)) = \\ &D(M_t(U(y_{n-1}^2))), \text{ 则} \\ g_n^2 &= D(r_n^1) + D(M_t(x_{n-1}^1)) - D(M_t(x_{n-1}^1)) \\ &= D[r_n^1 + M_t(x_{n-1}^1) - x_{n-1}^1] \end{aligned} \quad (6)$$

由式(6)得空间分辨率下变换前的漂移补偿型结构 4 如图 8 所示。

结构 4 和结构 3 一样, 都是对由量化引起的误差 d_q 进行补偿, 所以, 漂移误差也只源于 d_r , 不同点是采用下变换前进行补偿。因为补偿是在下变换前进行, 对下变换后量化系数需经上采样回到原分辨率, 所以存在上采样单元, 处理起来比前一种更复杂。

方法 3 加权平均(weighted mean)

$$MV = \frac{1}{2} \frac{\sum_{i=1}^4 mv_i A_i}{\sum_{i=1}^4 A_i} \quad (10)$$

式中, mv_i 为输入的运动矢量, A_i 为下变换前, 预测误差宏块的空间活动性度量, A_i 越小说明预测越准确(一般采用非零 AC 系数的个数做空间活动性度量)。这是一种用预测误差宏块的活动性度量对相应运动矢量进行加权处理, 从而获得下变换后运动矢量的方法。利用这种方法获得的新矢量向有较差预测效果的运动矢量偏移。

方法 4 DC 最大(DC max)

$$MV = \frac{1}{2} mv_i \Big|_{\max DC_i} \quad DC_i = \sum_{j=1}^4 dc_j \quad (11)$$

式中, $i=1, 2, 3, 4$ 代表相邻 4 个宏块, dc_j 为 8×8 块对应的直流系数。一个 16×16 宏块有 4 个 8×8 块, 所以有 4 个 DCT 直流系数, 将 4 个 DCT 直流系数求和, 选择和最大宏块对应的运动矢量作为下变换后的运动矢量。这实际是一种用 DC 值取代非零 AC 系数个数来表征块活动性的方法, 由于不需要乘法运算, 所以更简便。

5 时间分辨率变换转换编码器

时间分辨率下变换最简单的方法是丢掉 B 帧。B 帧不做参考帧, 处理较简单。但因 B 帧采用双向预测, 压缩率较高, 故丢 B 帧对进一步压缩码率起不到明显效果。若丢 P 帧, 则该 P 帧下一帧所指的帧由于 P 帧被跳过而变成无效, 所以对跳帧情况需采用适当处理。

时间分辨率变换器也可采用级联形式, 如图 11 所示结构 1 ((U_i, V_i) 为运动矢量的水平和垂直分量, 上标 S 表跳帧后的运动矢量)。它与码率变换结构 1 极相似, 不同处是存在跳帧控制开关(SW)。对

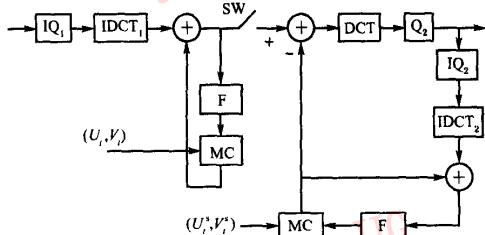


图 11 像素域跳帧转换编码器

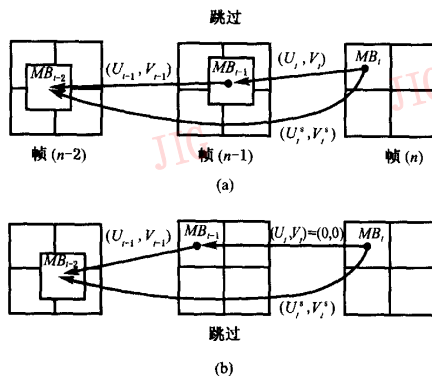


图 12 跳帧后预测误差再计算

跳帧的说明如图 12 所示, 假设 MB_{i-1} 是 MB_i 的最佳匹配宏块, MB_{i-2} 是 MB_{i-1} 的最佳匹配宏块。因为 $(n-1)$ 帧被跳过, 所以对于宏块 MB_i 需在 $(n-2)$ 帧中找到与其最佳的匹配宏块, 实际上就是 MB_{i-2} 。这时码流中需要确定的就是 MB_i 对 MB_{i-2} 的运动矢量和预测误差。

MB_i 对 MB_{i-2} 的预测误差 r_i^s 由 $MB_i = MB_{i-1} + r_i$, $MB_{i-1} = MB_{i-2} + r_{i-1}$ 得 $r_i^s = MB_i - MB_{i-2} = r_i + r_{i-1}$ 。有两种情况: 当 $(U_i, V_i) = 0$, 即 MB_i 不需运动估计时(图 12(b)), $Q[\text{DCT}(r_i)]$, $Q[\text{DCT}(r_{i-1})]$ 可直接由输入码流获得; 当 $(U_i, V_i) \neq 0$ 时(图 12(a)), $Q[\text{DCT}(r_{i-1})]$ 需重新计算。因 $r_{i-1} = MB_{i-1} - MB_{i-2}$, 故需重构 MB_{i-1} 和 MB_{i-2} (确定相应 MB 的位置)。

由以上分析可得结构 2, 即跳帧转换编码器如图 13 所示^[6]。开关 B_1 、 B_2 对非运动补偿块处理, 出现图 12(b)所示情况时, 二者闭合, 直接从输入码流

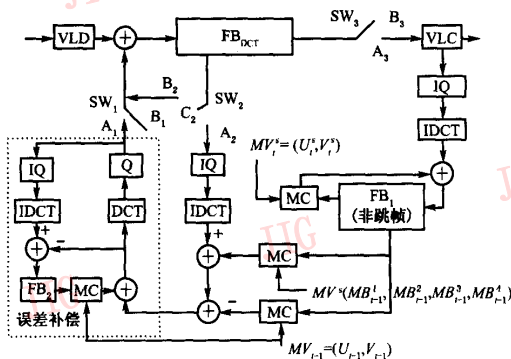


图 13 跳帧转换编码器

获得 $Q[DCT(r_t)]$ 、 $Q[DCT(r_{t-1})]$ ，进行求和运算。开关 A_1 、 A_2 对运动补偿块处理，出现图 12(a) 所示情况时，它们闭合，通过存储器 (FB_{DCT}) 中 $(n-1)$ 帧的信息，重构 MB_{t-1} 、 MB_{t-2} ，获得 r_{t-1} ，然后 DCT、量化处理 r_{t-1} ，最后与输入流中的 $Q[DCT(r_t)]$ 求和。由于重构引入了误差，所以在结构中加入误差补偿环以控制误差的积累。

6 时间分辨率变换器运动矢量的确定

以上两种结构形式时间分辨率变换器中，结构 2 的复杂度和质量好于结构 1，但都存在一个关键问题，即 MB_{t-1} 运动矢量 $MV_{t-1} = (U_{t-1}, V_{t-1})$ 的确定。跳帧后，当前帧对前一非跳帧的运动矢量 $MV_t^* = (U_t^*, V_t^*)$ 需重新确定。这类似于预测误差的确定，即都采用求和方式： $(U_t^*, V_t^*) = (U_t + U_{t-1}, V_t + V_{t-1})$ 。对图 12(b)， $(U_t, V_t) = 0$ ，因而 $MV_t^* = (U_{t-1}, V_{t-1})$ ；对图 12(a)， $(U_t, V_t) \neq 0$ ，但可直接从输入码流中获得。因此，核心问题是如何确定 $MV_{t-1} = (U_{t-1}, V_{t-1})$ 。目前有以下几种方法^[7~10]：

方法 1 双线性插入矢量选择 BiVS 该方法按重叠面积加权获得所需运动矢量

$$MV_{t-1} = (1-\alpha)(1-\beta)mv_{t-1}^1 + (\alpha)(1-\beta)mv_{t-1}^2 + (1-\alpha)(\beta)mv_{t-1}^3 + (\alpha)(\beta)mv_{t-1}^4 \quad (12)$$

其中， mv_{t-1}^1 、 mv_{t-1}^2 、 mv_{t-1}^3 、 mv_{t-1}^4 是 MV_{t-1} 所覆盖的 4 个宏块的运动矢量（当前宏块所指向的宏块最多可覆盖相邻跳帧中的 4 个宏块）， α 、 β 分别代表 MV_{t-1} 左上角对于 mv_{t-1}^1 左上角的水平和垂直距离。这种方法在连续跳帧情况下采用的是后向处理顺序，所以需要存储所有跳帧的运动矢量，会占用较多存储空间。

方法 2 前向主控矢量选择 FDVS 这种方法选择覆盖面积最大宏块的运动矢量做 MV_{t-1} ，在连续跳帧情况下采用前向处理顺序，避免多存储器。但若两个或两个以上宏块有相近重叠面积时，这种方法便失去了意义。

方法 3 活动性主控矢量选择 ADVS 这是一种用宏块活动性来决定运动矢量的方法，即选择 4 个宏块中活动性最大的一个，以它的运动矢量作为 MV_{t-1} （宏块的活动性可用非零 DCT 系数的个数判定）。此方法的计算量略大于 FDVS，但在高速运动情况下，ADVS 的效果优于 FDVS。

以上方法确定的 MV_{t-1} 都只是一种近似估计，根据需要还要进行运动矢量的优化 (MVR)。为此需

在一新的小搜索窗内进行全搜索计算，这就加重了计算负担。于是，出现了许多快速处理方法，通过减少搜索点以减少运算量。

方法 1 水平垂直搜索 HAVS 这是一种通过在搜索窗内先水平、后垂直搜索，来确定最佳运动矢量的优化算法。即以上述 3 种方法获得的基本运动矢量为起点，先在水平方向搜索。当基准点左边点的 SAD 值（绝对差值之和）大于基准点时，再进行基准点右边点的 SAD 计算。垂直方向采用相似步骤。由于有基本运动矢量，所以搜索范围可控制在 ± 2 像素，最好情况只需检测 5 个点，最坏情况也只需检测 7 个点。

方法 2 变步长搜索 VSS 该方法是根据基本运动矢量的大小确定搜索步长，运动矢量大，搜索范围大，步长长；反之，范围小，步长短。这是一种根据实际情况确定搜索范围的方法，可提高搜索精度。

7 结论

本文对多种转换编码器结构进行了分类、分析和对比。级联型是最一般的转换编码器结构，即先将压缩的视频流解码至像素域，然后根据不同要求，改变量化参数、进行空间分辨率或时间分辨率的变换等处理，再重新编码。这种结构灵活、无漂移，但较复杂、计算量较大。开环结构是最简单的结构形式，但存在严重的漂移，对图像质量影响较大。

评价转换编码器的两个主要指标是实现的复杂度和转化后解码重建图像的质量。通过分析两种结构的特点，抓住主要矛盾，对开环结构进行适当的补偿处理，以改善其图像的质量；对级联结构进行一些不会造成太大影响的假设处理，并充分利用源码流中的有用数据，简化结构，最终实现结构复杂性和图像质量的均衡。

目前视频转换编码的研究基本上是基于运动补偿和 DCT 变换模式的编码标准，如 MPEG-1、MPEG-2、H. 263 等标准的内部转码或它们之间的转码。MPEG-4 是一种注重多媒体系统的交互性和灵活性的新的编码标准，这种基于视频对象的编码标准和基于 MC+DCT 模式的编码标准之间差别很大，有大量的技术问题需要解决。另外，采用更先进、更复杂技术的 H. 264 标准的出现也给现有的转码技术提出了新的问题。因此，根据需要从 MPEG-1、MPEG-2 等到 MPEG-4 或 H. 264 的转换

编码将成为今后视频转码研究领域的热点。

参 考 文 献

- 1 Peng Yin, Vetro A, Bede Liu, *et al.* Drift compensation for reduced spatial resolution transcoding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12 (11): 1009~1020.
- 2 Vetro A, Peng Yin, Bede Liu, *et al.* Reduced spatio-temporal transcoding using an intra refresh technique[A]. In: Proceedings of IEEE International Symposium on Circuits and Systems(ISCAS)[C], Phoenix, AZ, USA, 2002;723~726.
- 3 Chen Mei-Juan, Chu Ming-Chung, Shen Yi-Lo. Motion vector composition algorithm for spatial scalability in compressed video [J]. IEEE Transactions on Consumer Electronics, 2001, 47(3): 319~325.
- 4 Tamer Shanableh, Mohammed Ghanbari. Heterogeneous video transcoding to lower spatio-temporal resolutions and different encoding formats[J]. IEEE Transactions on Multimedia, 2000, 2(2):101~110.
- 5 Lei Zhi-jun, Georganas N D. H. 263 video transcoding of spatial resolution downscaling[A]. In: Proceedings of the International Conference on Information Technology: Coding and Computing [C], Massachusetts Ave., NW Washington, DC USA, 2002: 425~430.
- 6 Fung Kai-Tat, Chan Yui-Lam, Siu Wan-Chi. New architecture for dynamic frame-skipping transcoder[J]. IEEE Transactions on Image Processing, 2002, 11(8): 886~900.
- 7 Hwang Jenq-Neng, Wu Tzong-Der, Lin Chia-Wen. Dynamic frame-skipping in video transcoding [A]. In: IEEE Second Workshop on Multimedia Signal Processing [C], Redondo Beach, CA USA, 1998:616~621.
- 8 Chen Mei-Juan, Chu Ming-Chung, Pan Chih-Wei. Efficient motion-estimation algorithm for reduced frame-rate video transcoder[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(4):269~275.
- 9 Jeongnam Youn, Sun Ming-Ting, Lin Chia-Wen. Motion vector refinement for high-performance transcoding [J]. IEEE Transactions on Multimedia, 1999, 1(1):30~40.
- 10 Jeongnam Youn, Sun Ming-Ting, Lin Chia-Wen. Motion estimation for high performance transcoding [J]. IEEE Transactions on Consumer Electronic, 1998, 44(3):649~658.



郭 映 1972 年生。博士研究生。主要研究方向为数字视频、多媒体技术。
E-mail:gying_tj@yahoo.com.cn



李桂苓 1938 年生。教授、博士生导师。1962 年于天津大学无线电系毕业。从事声音和图像信息的处理与传输、数字视频与多媒体技术的教学、科研工作。



徐 岩 1977 年生。博士研究生,主要研究方向为数字视频、多媒体技术。