

# 一种新的使用两种比特分辨率图象的块匹配 运动估计算法及其低功耗 VLSI 结构

张武健<sup>1</sup>, 邱晓海<sup>1</sup>, 周润德<sup>1</sup>, 陈弘毅<sup>1</sup>, Kondo Toshio<sup>2</sup>, Nakashima Takayoshi<sup>2</sup>, Ishitani Tsunehachi<sup>2</sup>  
(1. 清华大学微电子学研究所, 北京 100084; 2. NTT 电子公司(NEL 株式会社), 海老名市, 日本)

**摘要:** 本文给出了一种新的块匹配运动估计算法, 它根据视频图像内容的复杂程度自适应地选择常规的或者低比特分辨率的图像来进行块匹配, 并且采用了一种混合使用两种比特分辨率图像的新望远镜搜索算法. 模拟结果表明, 新算法具有较低的计算复杂性, 并且能够保证较好的视频质量. 基于该算法, 我们设计了一种新的脉动阵列结构的搜索引擎. 该引擎具有可分割的数据通道, 从而在使用低比特分辨率图像进行块匹配时能够通过加强处理的并行性来提高吞吐率. 新的运动估计器可工作在较低的时钟频率和电源电压之下, 具有低的功耗消耗.

**关键词:** 运动估计; 低比特分辨率; 望远镜搜索; VLSI 结构; 并行性; 低功耗

**中图分类号:** TN912      **文献标识码:** A      **文章编号:** 0372 2112 (2001) 06 0860 05

## New Block-Matching Motion Estimation Algorithm Using Dual Bit-Resolution Image and Its Low-Power VLSI Architecture

ZHANG Wujian<sup>1</sup>, QIU Xiaohai<sup>1</sup>, ZHOU Runde<sup>1</sup>, CHEN Hongyi<sup>1</sup>,  
Kondo Toshio<sup>2</sup>, Nakashima Takayoshi<sup>2</sup>, Ishitani Tsunehachi<sup>2</sup>

(1. Institute of Microelectronics, Tsinghua University, Beijing 100084, China; 2. NTT Electronics Corporation, Japan)

**Abstract:** A new motion estimation algorithm is proposed in this paper. It adaptively selects the full or reduced bit resolution image for block matching based on the texture complexity of the matching block. In addition, a new telescopic search method using dual bit resolution image is adopted. Simulation results show that the proposed algorithm has low computational complexity while achieves good image quality. We have designed a novel systolic architecture based search engine that has split data paths. Enhancing the operating parallelism in the low bit resolution case heightens the throughput. The new motion estimator works at a low clock frequency and a low supply voltage, therefore has lower power consumption.

**Key words:** motion estimation; low bit resolution; telescopic search; VLSI architecture; parallelism; low power

### 1 引言

视频压缩是实现数字电视广播, 数字存储媒体等应用的必要技术. 由于高度的计算复杂性, 其实时实现必须要求硬件的支持, 因此迫切需要低成本和低功耗的视频编码器.

运动估计是视频压缩的关键. 块匹配是最常用的运动估计方法, 最小绝对差值和(SAD)准则是最有效的也是最适合于VLSI实现的块匹配准则. 最直接的全局搜索方法要求很大的运算量, 其硬件实现代价太大.

目前有许多降低块匹配运动估计计算复杂性和硬件实现代价的方法<sup>[1,2]</sup>. 其中有一类方法的基本思想是使用低比特分辨率的图像进行块匹配. 通常都使用以8比特表示的像素亮度数据, 而当使用低于8比特表示的像素数据时, 实现匹配操作的硬件量就能够减小. 文[1]给出了一种通过量化来降低像素值比特分辨率的方法. 由于该方法中量化阈值是随待编

码宏块搜索窗中的图像自适应变化的, 因此每对一个新的待编码宏块作匹配比较之前, 都要求对此宏块及其搜索窗中的像素数据重新量化. 大量像素数据的量化使用了额外的专门硬件. 此外因为确定量化阈值时需要使用常规比特分辨率的搜索窗像素数据, 所以该方法并不能减小用于缓存图像数据的片上存储器的面积和带宽. 一种简单的降低比特分辨率的方法是直接截去像素值的低几位比特<sup>[2]</sup>. 它不需要任何额外的硬件代价, 不仅能够减小运算部件的硬件量, 而且能够减小存储器的面积和带宽. 但是该方法的缺点是没有考虑视频图像的特征. 当图像很平滑, 内容相对简单时, 其像素值都会比较接近. 此时减小像素值的比特位数可能会使多个候选宏块都具有相同的最小块匹配误差, 于是运动矢量的检测具有随机性. 当搜索频繁地陷入局部极小点时, 算法的性能就会严重衰减. 搜索范围越大, 随机性会越大, 算法性能也会越不稳定.

本文给出了一种使用两种比特分辨率图像来进行块匹配的新运动估计算法。我们根据视频图像内容的复杂程度自适应地选择常规的或者使用直接截方法得到的低比特分辨率图像,并且采用了一种混合使用两种比特分辨率图像的新望远镜搜索方法。新算法因为使用了低比特分辨率而具有较低的计算复杂性,而又因为必要时仍然使用了常规比特分辨率所以保持了较好的视频质量。此外我们还使用了运动跟踪和自适应搜索窗技术来进一步降低运动估计的运算量。在进行 VLSI 实现时,本文的目标是低功耗。为此,我们设计了一种新结构的脉动阵列,它具有可分割的数据通道。在使用低比特分辨率时,该阵列的处理并行性较常规比特分辨率时大大加强,从而能获得更高的吞吐率。基于新阵列结构的运动估计器可工作在较低的时钟频率和电源电压之下,所以有低的功耗消耗。

## 2 算法

本文即将给出的新算法以多分辨率望远镜搜索(MRTIcS)算法<sup>[3]</sup>为基础,这是因为它在性能和计算复杂性两方面已经有比较好的折衷,而且被用于了实用的 MPEG2 视频编码器中。

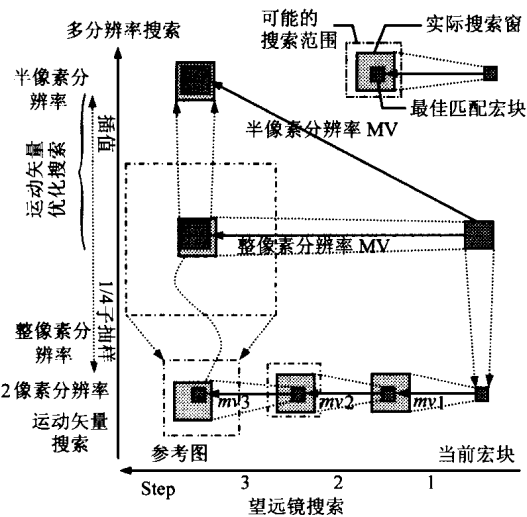


图 1 多分辨率望远镜搜索

如图 1,对原始图像分别进行水平和垂直方向的 2 1 子采样,运动矢量的搜索过程包括 2 像素分辨率搜索,  $\pm 1 \times \pm 1$  像素范围内的 1 像素分辨率搜索和  $\pm 0.5 \times \pm 0.5$  像素范围内的半像素分辨率搜索。在进行 2 像素分辨率的运动矢量搜索时使用望远镜方法<sup>[4]</sup>,借助于当前待编码图和参考图之间的中间图通过逐步跟踪找到最终的运动矢量。设中间图数目为  $D$ ,每一步实际搜索窗的大小均为  $SX \times SY$ ,则最终可能的搜索范围可扩大到  $(D+1)SX \times (D+1)SY$ 。MRTIcS 算法的运算量仅为全局搜索算法的  $1/(16(D+1))$ 。下文的改进都是针对 2 像素分辨率搜索即望远镜搜索过程的。

### 2.1 使用两种比特分辨率图像

通过直接截去像素值低比特位的方法来降低比特分辨率非常简单,不过在视频图像的内容比较简单时会因为“同匹配误差”问题导致算法性能严重衰减。对于有着复杂内容的图

像,其像素值波动变化很大。即使截去像素值的低比特位,像素差别仍然能够被区分出来,从而图像内容的基本特征也能够被有效地保留下来。此时使用直接截得到的低比特分辨率图像来进行块匹配是能够找到最佳匹配宏块的。因此针对当前待编码宏块所处四周图像内容的不同复杂程度,可使用不同比特分辨率的图像来进行块匹配。

因为每一步望远镜搜索的范围并不大,当前宏块图像内容的复杂程度也能够粗略地反映该宏块所处四周图像内容的复杂程度,所以本文只分析当前宏块图像。为此需要计算当前宏块的像素分散值。设当前宏块大小为  $M \times N$  像素,则其像素平均值为:

$$AVG = \frac{1}{M \cdot N} \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} T(i, j) \quad (1)$$

其中  $T(i, j)$  表示当前宏块在  $(i, j)$  位置处的像素亮度值。当前宏块的像素分散值为:

$$VAR = \sum_{j=0}^{N-1} \sum_{i=0}^{M-1} T(i, j) - AVG^2 \quad (2)$$

像素分散值反映了宏块中所有像素值的相对波动变化程度,因而能够比较好地表示图像内容的复杂性。像素分散值越大,意味着图像内容越复杂。当像素分散值  $VAR > T_{RB}$  时,我们使用截去低 4 比特位后得到的 4 比特分辨率图像进行块匹配,相反则仍然使用常规的 8 比特分辨率图像。这里  $T_{RB}$  是经过优化后预先确定的常阈值。应当指出,在视频编码过程中为确定编码类型(帧内/帧间编码)等原本就需要计算当前宏块的像素分散值,因而此处并不需要引入额外的计算和硬件。

望远镜搜索虽然进行多步跟踪,但最终目的是找到参考图中与当前宏块最佳匹配的宏块。由于中间步的搜索只不过是辅助得到运动的轨迹,因而允许存在小偏差。而另一方面因为每一步搜索都是在有一定大小的范围内进行的,所以即使中间某步搜索没能找到最优解,最终搜索的准确性一般也不会受到影响。据此,本文在进行望远镜搜索时,除开相对参考图的搜索(最终搜索步)使用上面给出的自适应方法来选择图像的比特分辨率外,其余步搜索(中间搜索步)都使用 4 比特分辨率图像(直接截去像素值的低 4 比特位,下同)。

于是在本文的新算法中,块匹配比较使用了 8 比特和 4 比特两种分辨率的图像。究竟使用哪一种比特分辨率一方面根据当前宏块像素分散值的大小来自适应选择,另一方面由望远镜的搜索步序来决定。

### 2.2 低复杂性算法

传统的望远镜搜索需要进行多步局部范围内的“全搜索”,并且每一步搜索窗的大小是固定不变的。为了进一步减小运动估计的计算复杂性,本文还采用了运动跟踪和自适应搜索窗技术。对隔行帧图序列在 MPEG2 标准下进行场编码,下面给出针对 2 像素分辨率运动矢量搜索的完整的新算法过程。

选择相邻  $P$  图之间存在两幅  $B$  图的图组格式。如图 2,当对  $P$  图宏块的顶场预测时:

(1)如果当前宏块处于首位位置,则转向(2);否则进行如下的运动跟踪搜索:(a)根据宏块像素分散值大小选择使用 8

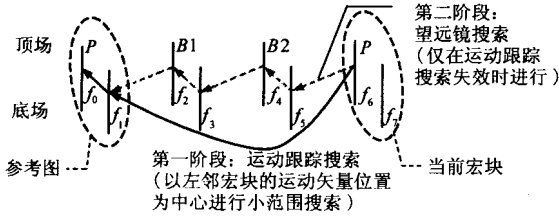


图2 2像素分辨率运动矢量搜索过程

比特或4比特分辨率; (b) 使用较小的搜索窗  $SSW$ . (c) 直接相对参考图的底场, 以相邻左侧宏块顶场相对参考图底场的运动矢量作为当前搜索窗的中心位移进行搜索, 得到最小块匹配误差  $MSAD$ . 如果  $MSAD < T_{ending}$ , 则以该步搜索得到的运动矢量作为相对参考图底场的运动矢量, 转向(3); 否则继续执行(2).

(2) 望远镜搜索. 初始化: 运动矢量置0, 使用常规大小的搜索窗  $RSW$ . (a) 如果当前步是相对参考图底场的搜索, 则转向(c); 否则进行: 跟踪上一步搜索的运动矢量, 使用4比特分辨率搜索, 得到该步的相对运动矢量  $(mux, myy)$ . 如果  $|mux| < T_{sw}$ , 则下一步搜索使用小窗  $SSW$ , 否则使用常规窗  $RSW$ . (b) 执行(a). (c) 使用根据宏块像素分散值选择的比特分辨率搜索, 得到相对参考图底场的运动矢量.

(3) 使用常规窗  $RSW$  和自适应选择的比特分辨率. 跟踪前面搜索得到的相对参考图底场的运动矢量, 相对于参考图的顶场进行搜索, 得到相对参考图顶场的运动矢量.

上述过程中  $T_{ending}$  和  $T_{sw}$  都为经过优化预先设定的常数值, 它们分别用于判断搜索是否可以提前中断和下一步搜索窗是否可以缩小. 对  $P$  图宏块的底场预测, 以及对  $B$  图宏块的预测都有与上述相似的过程.

### 2.3 模拟结果

本文对5种具有不同运动特点的视频序列进行了测试, 它们是“bus”, “cheer girl”, “flower garden”, “football”, 和“mobile calendar”, 所有这些序列均为隔行帧图, 30帧/秒, 图像尺寸为720像素×480行, 亮色度格式为4:2:0. 采用MPEG2标准, 基于MP@ML和场预测模式进行编码, 编码速率为5M比特/秒.

表1 本文算法与MRTIcS算法的性能和计算复杂性比较

算法		M1	M2	M3	M4	M5
Bus	PSNR	31.31	30.95	31.20	31.18	31.24
	Load	1.	0.50	0.74	0.64	0.35
Cheer girl	PSNR	29.67	29.56	29.65	29.64	29.62
	Load	1.	0.50	0.74	0.61	0.34
Flower garden	PSNR	29.91	29.72	29.86	29.84	29.89
	Load	1.	0.50	0.74	0.62	0.32
Football	PSNR	36.88	36.53	36.80	36.80	36.93
	Load	1.	0.50	0.74	0.73	0.43
Calendar	PSNR	27.11	26.98	27.09	27.07	27.13
	Load	1.	0.50	0.74	0.59	0.30
Average	PSNR	30.98	30.75	30.92	30.91	30.96
	Load	1.	0.50	0.74	0.64	0.35

M1: MRTIcS 算法; M2: 基于 MRTIcS 算法, 望远镜的所有搜索步使用4比特分辨率图像; M3: 基于 MRTIcS 算法, 仅望远镜的中间搜索步使用4比特分辨率图像; M4: 基于 M3 算法, 并且望远镜的最终搜索步使用

自适应比特分辨率图像; M5(本文算法): 基于 M4 算法, 并且使用运动跟踪和自适应搜索窗策略

我们对各序列的两个图组共28帧图像进行模拟后的统计结果如表1. 为了考察本文提出的各种策略的效果, 我们对5个算法进行了模拟, 其中M5为本文2.2小节给出的算法. 所有算法使用了相同的可能搜索范围. 规定常规搜索窗  $RSW$  的大小为  $[-16, 15]$  ( $[-8, 8]$ ). 若以4:1子采样后的场图格式表示,  $RSW$  的大小为  $[-8, 7]$  ( $[-2, 2]$ ), 本文选择小搜索窗  $SSW$  的大小为  $[-4, 3] \times [-2, 2]$ . 表1中算法性能是以解码图的峰-峰信噪比(PSNR~dB)来衡量的, 而运算量以基本操作的次数来衡量并且参照MRTIcS算法. 设定MRTIcS算法的运算量为1(全局搜索算法的运算量为123). 为简单计, 我们认为8比特操作按两步实现, 其中每一步为4比特操作, 因而8比特操作的运算量为4比特操作的2倍.

对照表1, 尽管采用简单截取方法(M2)能够将平均运算量减少为MRTIcS算法的50%, 但是平均的PSNR衰减却达到0.23dB. 本文提出的各种策略能够在相似的PSNR下减少运算量. 首先, 仅对望远镜的中间搜索步使用4比特分辨率图像(M3)能够节省26%(1-0.74)的运算量, 平均的PSNR衰减仅为0.06dB; 然后, 对望远镜的最终搜索步使用自适应比特分辨率图像(M4)能够继续节省10%(0.74-0.64)的运算量, 平均的PSNR衰减仅为0.01dB; 最后, 使用运动跟踪和自适应搜索窗策略(M5)能够再节省29%(0.64-0.35)的运算量. 最终本文算法是优于MRTIcS算法的, 平均运算量仅为MRTIcS算法的35%, 并且获得了很好的视频质量, 平均的PSNR衰减仅0.02dB. 从表1还可以发现对某些图像序列运动跟踪和自适应搜索窗策略还能够改善图像质量, 从而本文算法(M5)获得了比MRTIcS算法稍好的性能. 这是因为(1)如果运动跟踪搜索成功, 就减少了使用4比特分辨率图像的机会; (2)自适应策略使得每一步搜索窗更精确, 从而减少了因为在望远镜中间步使用4比特分辨率图像引起的搜索偏差; (3)运动跟踪改善了运动矢量场的均匀性, 从而减少了采用望远镜搜索时有时不能找到正确运动矢量的可能性.

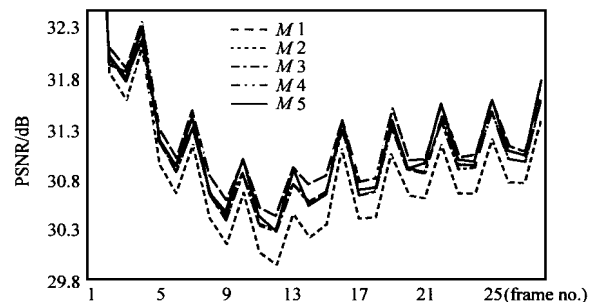


图3 “bus”序列解码图PSNR值曲线

图3给出了“bus”序列解码图的PSNR值曲线, 该序列是本文新算法对应的最差情形. 很显然采用新算法的视频质量是明显优于直接使用4比特分辨率方法的.

### 3 低功耗 VLSI 结构

无论是传统的MRTIcS算法还是本文算法实质上都是一

个多步的局部范围内的“全搜索”过程,因而 VLSI 实现的关键就是设计针对“全搜索”的引擎。

本文设计的低功耗搜索引擎结构如图 4 所示,它是面向宏块像素的一维脉动阵列,由  $M \times N$  个处理单元(PE)组成,对应当前宏块的  $M \times N$  个像素。这些像素数据  $T$  通过一条全局广播的总线在脉动阵列开始计算的同时串行写入到对应 PE 单元的锁存器中。而搜索窗数据  $R$  是通过梳状总线供给各个 PE 单元的。处理单元 PE 的功能是计算  $T$  与对应  $R$  的差值绝对值,然后累加前面处理单元传递过来的部分绝对差值和  $PS$ ,其结构如图 4(b)所示。求差值绝对值运算被分解为求差值和求绝对值两次操作,并让它们分布在两个周期中完成。求绝对值操作被合并并在累加运算中。

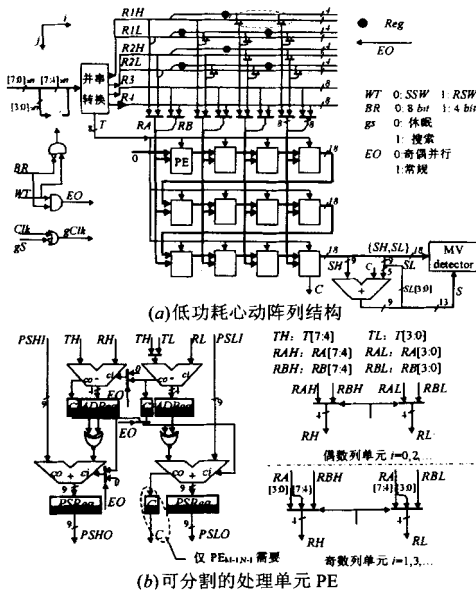


图 4 低功耗的脉动阵列搜索引擎

本文算法使用了 8 比特和 4 比特两种比特分辨率的图像,而针对 4 比特像素的数据通道宽度可较 8 比特时减半。与传统脉动阵列结构引擎<sup>[3]</sup>不同的是,本文给出引擎的 PE 单元数据通道被设计成可分割的,这样 PE 单元在一个周期中就能够处理一个 8 比特或者两个 4 比特的搜索窗像素。在使用 8 比特分辨率的图像时,图 4 所示的脉动阵列与传统引擎一样按行扫描顺序依次相隔一个时钟周期开始对候选宏块计算 SAD 值。但是在使用 4 比特分辨率的图像时,对处于奇偶列位置上相邻两个候选宏块的 SAD 计算是同时开始的,此时 PE 单元中左侧和右侧两套相同的数据通道相互独立,分别用于处理针对偶数列和奇数列候选宏块的相关计算。运动矢量检测单元需要按偶数列和奇数列位置来分别比较 SAD 值,搜索完成之后再行奇偶列最小 SAD 的比较。在 8 比特分辨率时,两个 4 比特减法器合并为一个 8 比特的减法器,此时一个周期中 PE 只能处理针对一个候选宏块的相关计算了。为了保证短的路径延迟,两个累加器并没有合并,不过只要对阵列中最后一个 PE 两个累加器的结果做加法就能够得到一个候选宏块完整的 SAD 值。

通过设计可分割的数据通道,提高了 4 比特分辨率情形下的处理并行性,而没有增加太多的硬件量,此时完成一次  $SX \times SY$  范围内的全搜索需要  $M \cdot N + (SX/2) \cdot SY$  个时钟周期,这比使用传统结构引擎少了  $(SX/2) \cdot SY$  时钟周期。由于处理并行性的加强,本文给出的脉动阵列结构具有很高的吞吐率,因而可工作在较低的时钟频率下。较低的时钟频率意味着允许更长的关键路径延迟,从而可以通过降低工作电压的方法来降低功耗。由于功耗与电压成平方关系,因而这种方法降低功耗的效果会是很显著的。

使用日本 NEL(NTT Electronics)公司提供的 0.54 $\mu\text{m}$ , CMOS 工艺库进行综合得到新结构电路的面积仅为传统结构<sup>[3]</sup>的 1.2 倍。采用本文算法,基于新结构的搜索引擎可工作在 54MHz 时钟频率和 2.5V 电压下,而传统脉动阵列结构的搜索引擎要求工作在 81MHz 时钟频率和 3.3V 电压下<sup>[3]</sup>。由于功耗  $P \propto C_{\text{eff}} \cdot VDD^2 \cdot f$ ,其中  $C_{\text{eff}}$  为等效的电容负载,认为正比于电路面积,因此:

$$\frac{P}{P_0} = 1.2 \cdot \left( \frac{2.5}{3.3} \right)^2 \cdot \frac{54}{81} = 0.46 \quad (3)$$

其中  $P$  和  $P_0$  分别表示新结构电路和传统结构电路的功耗。可见采用降低电压的办法可以将新结构电路在最坏情形下的功耗降低到传统结构电路的 46%。

在新结构的搜索引擎中还使用了门控时钟  $gClk$ ,在搜索提前完成时,控制信号  $gS$  将时钟屏蔽,从而能够节省来自阵列以及时钟树的功耗消耗。所以运动估计器的功耗还能够更进一步降低。

片上存储器用于缓存当前宏块及其搜索窗数据。为对下一个宏块预测作准备还需要缓存下一个宏块及其搜索窗数据。在望远镜搜索中,搜索窗数据包括参考图和中间图的像素数据。在新算法的望远镜搜索中,除开相对于参考图的搜索可能使用 8 比特分辨率图像外,其余搜索步均使用 4 比特分辨率的中间图像,因此对于中间图来说要求的存储器容量可节省一半,相应的存储器带宽也减小一半。总体上,本文算法要求的存储器容量和带宽分别下降到 MRTlc 算法的 78.1% 和 73.0%。在存储器容量和带宽减小后,不仅硬件代价减小了,而且与存储器相关的功耗和整个运动估计器的功耗也随之降低。这是因为小容量存储器对应的一次数据操作消耗的功耗减小,而带宽的降低意味着存储器操作次数减少了。

## 4 结论

本文对一种被实用 MPEG2 视频编码器采用的基于多分辨率望远镜搜索的运动估计器进行了算法和结构上的改进,使其具有低的功耗消耗。在算法上我们提出了使用两种比特分辨率图像、运动跟踪,以及自适应搜索窗三种降低计算复杂性的策略。相应地在结构上通过加强处理的并行性,使用门控时钟的方法能够大大降低运动估计器的功耗消耗。新的运动估计器与传统的运动估计器相比,其功耗降低到不到 46%,而仍然保持着相近的视频质量;同时存储器容量和带宽也分别减小到原来的 78% 和 73%。

## 参考文献:

- [ 1 ] S Lee, et al. New motion estimation algorithm using adaptively quantized low bit resolution image and its VLSI architecture for MPEG2 video encoding [ J ]. IEEE Trans. Circuits & Systems for Video Tech. , 1998, 8( 6 ): 734- 744.
- [ 2 ] Y Baek, et al. An efficient block matching criterion for motion estimation and its VLSI implementation [ J ]. IEEE Trans. Consumer Electronics, 1996, 42(4): 885- 892.
- [ 3 ] K Sugui, et al. A real-time motion estimation and compensation LSI with wide search range for MPEG2 video encoding [ J ]. IEEE Trans. Solid State Circuits, 1996, 31( 11 ): 1733- 1741.

- [ 4 ] K Matsuda, et al. A new motion compensation coding scheme for video conferences [ A ]. IEEE International Communications Conference [ C ], 1984: 234- 237.

## 作者简介:



张武健 男. 1974 年生于湖北浠水. 1996 年毕业于清华大学电子工程系, 获学士学位. 1996 年至今在清华大学微电子学研究所攻读博士学位. 主要从事数字视频和图像处理等方面的研究.