



视频处理 与编码中的 运动估计技术

端木春江 著

视频处理 与编码中的 运动估计技术

端木春江 著

图书在版编目(CIP)数据

视频处理与编码中的运动估计技术 / 端木春江
著. —南京:南京大学出版社, 2011. 8
ISBN 978 - 7 - 305 - 08524 - 6

I. ①视… II. ①端… III. ①视频编码
IV. ①TN762

中国版本图书馆 CIP 数据核字(2011)第 134042 号

出版发行 南京大学出版社
社 址 南京市汉口路 22 号 邮编 210093
网 址 <http://www.NjupCo.com>
出版人 左 健

书 名 视频处理与编码中的运动估计技术
著 者 端木春江
责任编辑 王秉华 编辑热线 025-83595720

照 排 南京玄武湖印刷实业有限公司
印 刷 江苏南大印刷厂
开 本 787×960 1/16 印张 8.5 字数 144 千
版 次 2011 年 8 月第 1 版 2011 年 8 月第 1 次印刷
ISBN 978 - 7 - 305 - 08524 - 6
定 价 25.00 元

发行热线 025-83594756 83686452
电子邮箱 Press@NjupCo.com
Sales@NjupCo.com(市场部)

* 版权所有,侵权必究
* 凡购买南大版图书,如有印装质量问题,请与所购
图书销售部门联系调换

本书作者承担的与本书内容相关的科研项目：

(1) 视频信号压缩中快速运动估计技术的研究,2006年浙江省科学技术厅钱江人才科研项目(项目编号:2006R10010)。

(2) 适于 MPEG-4 和 H.264 的快速运动估计算法的开发与研究,2006年浙江省自然科学基金项目(项目编号:Y106657)。

(3) 第四代移动通信中的视频信息的优化传输技术的研究,2011年浙江省自然科学基金项目(项目编号:Y1110510)。

前 言

通信、数字信号处理和计算机图像学的发展、融合带来了多媒体技术的广泛应用。在当今的多媒体技术中,视频信号的处理担负着重要的角色。为了解决视频信号占用大量数据带宽的问题,视频信号的压缩成了多媒体系统中必备的手段与方法。在现今的国际视频信号的压缩标准中,块运动估计是一个关键的组成部分,用以去除视频信号时间域上的相关性。然而,这些标准并没有具体定义采用哪种块运动估计的算法,所以采用什么样的运动估计算法的工作留给了这些标准的具体实现者。因此,开发一种快速且高效的运动估计算法成了近年来的研究热点。本书基于作者在加拿大康克迪亚(Concordia)大学读博士期间和在浙江师范大学做科研项目期间的科研成果,以及在国际知名会议和期刊上所发表的论文编写,紧紧围绕块运动估计算法这一研究热点而展开。

本书首先回顾了当前国际上流行的运动估计算法,在此基础上介绍了作者在国际会议及期刊上提出的几种运动估计的算法。

现有的运动估计的算法可以分为两大类:多步搜索算法类和穷尽搜索算法类。二维对数搜索算法(2-D logarithmic search algorithm)、正交搜索算法(orthogonal search algorithm)、逐步搜索算法(one at a time search algorithm)、共轭方向搜索算法(conjugate direction search algorithm)、三步搜索算法(three-step search algorithm)、四步搜索算法(four-step search algorithm)、无约束的基于中心的星形搜索算法(unrestricted diamond search algorithm)、预测区域搜索算法(predictive-area search algorithm)和基于块的梯度下降搜索算法(block-based gradient-descent search algorithm)属于多步搜索算法类。在此类算法中,运动估计进程被分为多步的搜索过程。在每一步的搜索过程中,选择了一些搜索窗中的搜索点。在每个搜索点上计算当前宏块与参考块的平均绝对差值(mean absolute difference)。由于平均绝对差值在搜索窗中存在很多局部最小值,多步搜索算法类的结果常常局限于某一局部最小值而得不到全局最小值,因此,多步搜索算法损失了运动估计的精度。全搜索算法(full-search algorithm)和选择性淘

汰算法(selective elimination algorithm)属于穷尽搜索算法类。由于在此类算法中依次检查了搜索窗中的每一个搜索点上的平均绝对差值,全搜索法能够得到全局最优解,因此,全搜索算法保持了运动估计的精度。对现有的运动估计算法的仿真结果表明多步搜索法中的每一个算法的复杂度明显低于全搜索算法。然而,这个复杂度的降低是在算法精度的损失的代价上而获得的。仿真结果同时表明,虽然选择性淘汰算法的复杂度低于全搜索算法,此算法的复杂度依然很高。因此,选择性淘汰算法和全搜索算法一样不适于实时应用的场合。

本书介绍了一种新的判别视频信号帧间静止宏块的算法。这种方法利用了视频信号帧间的统计特性以检测静止宏块。运用这种方法可以加速块运动估计的进程。当某一宏块被检测为静止宏块时,块运动估计中对此宏块的搜索过程可以跳过。试验结果表明所提出的方法在显著地降低运动估计过程的复杂度的同时保持了其精度。描述此算法的论文被 2002 年 IEEE (国际电子与电气工程协会)于克罗地亚召开的电子大会所采纳与录用。

本书接着提出了一种基于矢量的快速运动估计算法。此算法非常有利于单指令多数据(single instruction multiple data)的框架结构。此算法在大量地降低全搜索运动估计算法的复杂度的同时保持了其精度。实验结果显示通过采用单指令多数据的框架结构,所提出算法的运行速度还能进一步地降低大约 74%。描述此算法的论文被 2002 年 IEEE 于美国召开的电路与系统大会所采纳与录用。

在本书中,提出了 8 比特的部分和的新概念。这些部分和能够利用单指令多数据(SIMD)的框架结构中的并行字节数据的特性加快运动估计的执行速度。由于它们仅为 8 比特,一个 64 比特的 SIMD 寄存器能够同时处理 8 个这样的部分和。然后,提出了一个利用这些部分和加速运动估计运行速度的方法。仿真实验结果表明所提出的方法能够通过减少计算复杂度而给各种运动估计算法带来大量的运行速度的提高。实验结果同时显示,运用 SIMD 框架结构能进一步加快所提出方法的运行速度达 93%以上。描述此方法的论文发表在国际权威刊物 IEEE 的会刊《视频技术的电路与系统》(IEEE Transactions on Circuits and Systems for Video Technology) 2007 年第 8 期中。

在本书中,提出了一种优于 UMHexagonS 算法(被 H. 264 国际压缩标准推荐的算法)的快速运动估计算法。所提出的算法在如下三个方面对 UMHexagonS 算法进行了改进。首先,在运动矢量的初始预测中,增加了

一个预测矢量,以避免陷入局部最小。然后,UMHexagonS 算法中的 5×5 的全搜索被替换为一个八边形搜索和两个菱形搜索以减少计算复杂度。最后,基于多六边形网格的搜索被基于多八边形网格的搜索所取代,以在更好地发现最优运动矢量的方向和位置的同时减少计算复杂度。仿真结果显示,所提出的算法能在不损失精度的情况下,减少 UMHexagonS 算法的计算复杂度达 5%~10%。描述此算法的内容发表在 2008 年 IEEE 的关于多媒体的国际会议(2008 IEEE International Conference on Multimedia and Expo)上。

针对当前 H. 264/AVC 编码器中运动估计比较耗时的情况,本书提出了一种新的基于多模板搜索的快速运动估计算法。该算法首先将当前块划分为准静止块、小运动块和大运动块,然后根据判别的运动块类型相应地选择搜索模板和搜索策略。实验表明,在图像质量基本保持不变的情况下,该算法提高了编码速度达 20% 以上。描述这方面的部分内容已发表在 2008 年 IEEE 的关于多媒体的国际会议(2008 IEEE International Conference on Multimedia and Expo)上。

本书提出了 H. 264 标准中快速运动估计模式的确定算法。这个算法减少了全模式搜索算法中的一些不必要的计算量以快速高效且较优地确定运动估计的模式。实验结果表示所提出算法能在高于 30% 加快运动估计模式选择过程的速度的同时保持全模式搜索算法的精度。描述这方面的部分内容已发表在 2009 年无线通信和信号处理的国际会议(2009 International Conference on Wireless Communications and Signal Processing)上。

本书的撰写和出版得到了很多人的帮助,在此表示感谢。首先,要感谢作者在读博期间的导师 Dr. Omair Ahmad 和 Dr. M. N. S. Swamy,他们是 IEEE 的会员,在科研上给了我极大的指导和鼓励。对于作者在读博期间所发表的每一篇论文,他们都做了精心的修改。同时,要感谢朱卫平教授、Shatnawi 教授、王养利教授、熊惠霖教授和杜克林博士,与他们有意义的讨论使作者的科研工作更趋完善。另外,要感谢作者所指导的两名研究生:陈彤和张瑜。最后,要感谢我的父亲端木义万教授和夫人沈虹女士,没有他们的支持,本书的作者不可能取得这些科研成果。

本书适于对多媒体处理及其编码感兴趣的大学高年级学生与研究者的阅读,也可作为开发多媒体应用者所必备的参考书。

目 录

第一章 引言

- 1.1 视频压缩系统的结构 1
- 1.2 本书的结构和所涉及的领域 3

第二章 块运动估计算法

- 2.1 块运动估计过程和全搜索算法 5
- 2.2 快速运动估计算法 7
 - 2.2.1 多步搜索算法 7
 - 2.2.2 穷尽搜索算法 12
- 2.3 性能比较 12
 - 2.3.1 计算复杂度 12
 - 2.3.2 码率-失真度性能 14
- 2.4 小结 16

第三章 跳过静止宏块中的搜索过程的快速运动估计方法

- 3.1 所提出的方法 17
- 3.2 检测静止宏块的准则 20
- 3.3 自适应阈值 21
- 3.4 仿真结果 22
- 3.5 小结 35

第四章 基于矢量的快速运动估计算法

- 4.1 基于矢量的快速运动估计算法 36
 - 4.1.1 部分和的形成 37
 - 4.1.2 MAD 的下界 39
 - 4.1.3 计算部分和的快速算法 41
 - 4.1.4 算法 44
- 4.2 所提出算法在 SIMD 结构下的实现 46

4.3	所提出算法的计算复杂度	47
4.3.1	对一个 $l(0 \leq l \leq 5)$ 级别 $MAD^l(\vec{V})$ 的计算 复杂度	47
4.3.2	计算帧部分和的计算复杂度	48
4.3.3	所提出算法的计算复杂度	49
4.3.4	理论上的加速	49
4.3.5	实际的加速	50
4.4	仿真结果	50
4.5	小结	53
第五章 SIMD 框架下的基于 8 比特部分和的快速运动估计算法		
5.1	运动估计的原理和 8 比特部分和的引入	54
5.2	基于 16 个亮度值的 8 比特部分和	56
5.3	多层次 8 比特的部分和	60
5.3.1	多层次 8 比特部分和	60
5.3.2	$SAD^l(\vec{V})$ 的上界	62
5.3.3	利用多层次 8 比特部分和的方法	64
5.3.4	利用多层次 8 比特部分和的最优方法	65
5.3.5	SIMD 下计算 $SAD^l(\vec{V})$ 的实现	67
5.4	计算复杂度的分析	67
5.4.1	任一运动估计算法的计算复杂度	67
5.4.2	采用所提出方法的任一运动估计算法的计算 复杂度	68
5.4.3	理论上采用方法 5.2 所带来的运动估计过程的 加速	70
5.4.4	实际上采用方法 5.2 所带来的运动估计过程的 加速	70
5.5	仿真结果	71
5.6	小结	75
第六章 基于多八边形搜索的快速运动估计算法		
6.1	问题的背景	76

6.2	UMHexagonS 算法回顾	77
6.2.1	初始预测	77
6.2.2	非对称十字形搜索	78
6.2.3	5×5 区域的全搜索	78
6.2.4	多六边形网格搜索	78
6.2.5	扩展的六边形搜索	78
6.3	本书所提出的算法	79
6.3.1	初始预测的改进	79
6.3.2	对 5×5 全搜索的改进	80
6.3.3	多八边形网格搜索	81
6.4	仿真结果	82
6.5	小结	84
第七章 基于多模板的快速运动估计		
7.1	问题的背景	85
7.2	所提出的基于多模板搜索的快速运动估计	86
7.2.1	准静止块判定准则	86
7.2.2	搜索起点的预测	86
7.2.3	多模板自适应搜索	87
7.2.4	算法描述	88
7.3	仿真结果	90
7.4	小结	92
第八章 在 H. 264 标准中的快速运动估计模式确定算法		
8.1	提出算法的过程	93
8.1.1	运动估计模式的预测	93
8.1.2	运动估计模式搜索过程的终止判别	94
8.1.3	SKIP 模式的选择	94
8.1.4	快速运动估计模式选择算法	94
8.2	仿真结果	95
8.3	小结	97
第九章 结论和研究展望		
9.1	结论	98

9.2 研究展望	100
与本书相关所发表的论文	101
参考文献	102
附录 A 块运动估计算法	
A.1 全搜索算法(FSA)	112
A.2 二维对数搜索算法(2DLSA)	113
A.3 正交搜索算法(OSA)	113
A.4 一次一方向搜索算法(OATSA)	114
A.5 共轭梯度方向搜索算法(CDSA)	115
A.6 三步搜索算法(3SSA)	115
A.7 四步搜索算法(4SSA)	116
A.8 无约束中心偏置钻石搜索算法(UDSA)	117
A.9 预测搜索区域算法(PSAA)	117
A.10 基于块的梯度下降搜索算法(BGDSA)	118
A.11 选择性淘汰算法(SEA)	118
附录 B 块运动估计算法的计算复杂度	
B.1 全搜索算法和多步搜索算法的计算复杂度	120
B.2 选择性淘汰算法的计算复杂度	122

第一章 引言

视频信号通常每秒包括 15 帧以上的静态图像。视频信号的处理通常需要存储、传输并操作大量数据。例如：对于采用 QCIF(四分之一通用中间格式)的低分辨率视频序列，如果不采用数据压缩的话，需要 6 Mbps 的网络带宽。为在只有 64 kbps 的公众服务电话网(PSTN)上传输这个 QCIF 格式的视频序列，需要用压缩倍数大于 100 的方法来压缩此视频信号。即使对于带宽较高的网络来说，视频压缩能使该网络提供更多的服务，或是为更多的用户服务。

为满足在多媒体通信中视频压缩的需要，人们提出了一些视频压缩方面的国际标准，以使多媒体产品能广泛应用。在这些标准中，块运动估计是其中的必需和重要的一部分。然而，这些标准中并没有规定具体采用哪一个运动估计算法，所以这部分的工作留给了标准的具体实现者。因此，具有高压缩性能的快速块运动估计算法的开发是近年来研究的热点，同时也会一直吸引人们从事此方面的研究。

1.1 视频压缩系统的结构

图 1.1 给出了一个典型的视频编码器的结构。现今国际视频压缩标准 MPEG-1 [56, 57], MPEG-2 [31, 58—60, 74], MPEG-4 [2, 3, 32, 61—63, 70, 88, 90], H.261 [25, 33, 53], H.263 [8, 25, 34, 93, 98], H.263++ [8] 和 H.264 [35] 采用了这个结构。在重建的视频序列的失真度小于规定值的情况下，视频编码器的主要目的是用最少的输出比特数来表示原始的视频序列。为达到这一目标，块运动估计被用来去除视频序列中时间域上的冗余度，而 DCT(离散余弦变换)或 DWT(离散小波变换)被用来去除空间域上的冗余度。

我们通常利用数字视频中各帧之间的相关性来压缩视频信号。例如，在当前帧中有大量的块可用它们在前一重建帧中相应的区域(参考块)来表示，这个重建的视频帧被称为参考帧。运动估计和运动补偿的基本思想是用参考块来表示当前块。在编码器端，为当前块寻找一个好的参考块的过

程被称为运动估计。参考块从参考帧到当前帧之间的位移用一个运动矢量来表示,这个运动矢量的信息被编码并传输到解码端,这样,在接收端的解码器可提取参考块。这个提取参考块的过程被称为运动补偿。由于用来表示运动矢量的比特数在通常情况下大大少于直接表示当前帧的比特数,因此利用块运动估计和运动补偿的技术可以节约大量表示当前帧的比特数。

在大多数情况下,当前块和参考块之间的误差,即预测误差较大,同时预测误差存在空间域上的冗余度。因此,在编码器中利用 DCT 或 DWT 来对预测误差做变换,以去除预测误差在空间域中的相关性。DCT 或 DWT 的系数经量化后将被编码和传输给接收端,这样,接收端的解码器可用编码和表示预测误差的比特来提高重建视频帧的质量。

在图 1.1 中,存在一个模式确定的模块来决定编码器应采用的最佳模式和参数。在当前块为 Inter 模式下,先进行运动估计,然后编码运动矢量信息,接着对预测误差做变换、量化和编码。在当前块为 Intra 模式下,不对当前块做运动估计和对运动矢量进行编码,当前块将直接被变换、量化和编码。Inter 模式适于视频序列中除场景变化外的大多数情况。在场景变化时,由于运动估计不能很好地预测当前块,此情况下适合采用 Intra 模式。除运动矢量和预测误差的信息外,这些模式和参数信息也被编码和传输到解码器,这样解码器才能正确地对收到的数据进行解码。

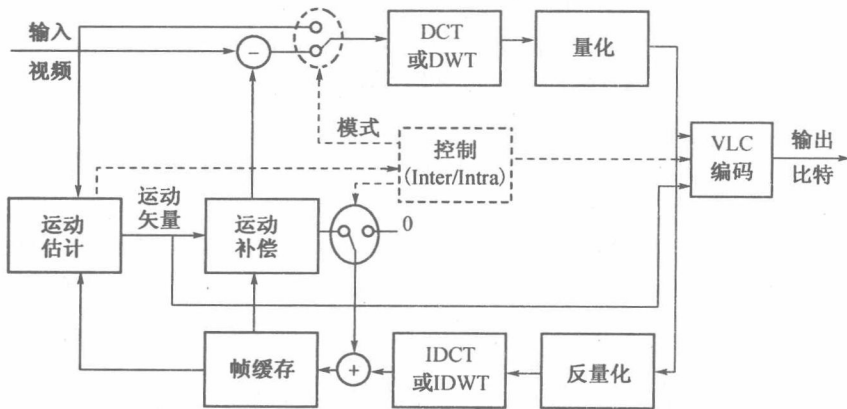


图 1.1 视频编码器的结构

1.2 本书的结构和所涉及的领域

如上所述,在国际视频标准中并没有规定所采用的运动估计算法,这个工作被留给了实现这些标准的开发者。这样,不同厂商的视频编解码产品具有不同的性能,相互间可进行竞争。因此,开发具有良好精度的快速块运动估计算法就十分必要。

本书的目的在于:(1)在不损失其精度的情况下,开发出一种降低某一运动估计算法计算复杂度的方法。(2)利用单指令多数据(SIMD)技术来加速块运动估计的过程。(3)开发新的适于 SIMD 结构应用的快速运动估计算法。本书的结构如下。

在第二章中,我们先简单地介绍块运动估计的过程和全搜索算法,通过已有文献回顾一些典型的快速运动估计算法。接着,为衡量一个运动估计算法的压缩性能,我们将讨论码率-失真度曲线的概念,并用它来评价一个视频编码器的压缩性能。然后,我们对这些已有的运动估计算法进行仿真,并比较它们的计算复杂度和压缩性能。

基于我们的观察,对于一个存在大量静止块的视频序列,这些块在视频信号的相邻帧之间是无运动的。这样,在运动估计中对这些块所进行的搜索过程是不必要的。在第三章中,我们提出了检测这些静止块的方法,并在运动估计的过程中跳过这些块的搜索过程。所提出的方法被运用到了一些已有的运动估计算法中,以减少这些算法的计算复杂度。

SIMD 技术利用数据并行的特点提供了一个可以用来加速某些进程执行的方法,于是在第四章中我们提出了一个适于 SIMD 实现的基于矢量的快速运动估计算法。我们证明了此算法具有和全搜索算法一样的精度和压缩性能。

为了很好地利用已有的 SIMD 技术中字节数据并行的特点,我们在第五章中引入了 8 比特部分和的概念。由于这些部分和仅为 8 比特,8 个这样的部分和能被一个 64 比特的 SIMD 寄存器同时处理。接着,我们描述了一个利用这些部分和加速一给定的运动估计算法的方法。然后,这个 8 比特部分和的概念被推广到了具有 4 个级别的部分和的情况,同时展示了有 15 种利用这些多级 8 比特部分和在不损失精度的情况下加速一运动估计算法的方法。我们用全搜索算法来确定这 15 种方法的性能,并找到具有最小计算复杂度的方法。最后采用这个具有最小复杂度的方法来加速各种块

运动估计算法,并用其对一些典型的视频测试序列进行仿真。

非对称六边形搜索(UMHexagonS)算法由于其具有较少的计算复杂度和较高的精度而被选为 H. 264 标准中实现运动估计的一种算法。在第六章中提出了一种优于 UMHexagonS 算法的快速运动估计算法,对 UMHexagonS 算法进行了改进。

针对当前 H. 264/AVC 编码器中运动估计比较耗时的情况,本书的第七章提出了一种新的基于多模板搜索的快速运动估计算法。该算法首先将当前块划分为准静止块、小运动块和大运动块,然后根据判别块的运动类型相应地选择搜索模板和搜索策略。

在国际视频压缩标准 H. 264 中,存在很多可供选择的运动估计的模式。如何用较少的计算量确定最优的运动估计模式是近年来的研究热点。本书的第八章提出了一个快速确定 H. 264 中运动估计模式的算法。

在点明本书所做的贡献和对未来研究建议的基础上,第九章对全书作了总结,并描述了可继续开展研究的相关领域。

第二章 块运动估计算法

块运动估计算法的目的是从参考帧(先前重建的帧)的搜索窗中寻找和当前块相匹配的块(参考块)。参考块和当前块的相似度通常是用平均绝对差值(Mean Absolute Difference, MAD)来表示的。因为对于搜索窗中的每一个候选运动矢量都需要计算一次它的平均绝对差值,所以块运动估计中的全搜索算法是最费时但最精确的。

由于块运动估计的全搜索算法具有很高的计算复杂度,因此在已有的文献中,人们提出了许多快速运动估计算法。这些算法可大致归为多步搜索算法和穷尽搜索算法这两类。对于多步搜索算法,只选择搜索窗中所有候选运动矢量的一个子集,这样可以减少所需计算 MAD 的次数。对于穷尽搜索算法,先计算 MAD 的一个下界,然后尽可能地利用此下界来避免计算 MAD。

这一章的 2.1 节描述了块运动估计的过程以及全搜索算法,然后在 2.2 节回顾一下具有代表性的一些快速运动估计算法,最后在 2.3 节将给出各种块运动估计算法的计算复杂度和码率-失真度性能的仿真结果。

2.1 块运动估计过程和全搜索算法

在块运动估计算法中,当前帧被分割为固定大小 $b \times b$ (b 为 2 的整数次幂)的块。当前块 B_c 的运动矢量将根据已定义的匹配准则去寻找参考帧的搜索窗中最匹配的块。在现有的视频压缩标准中^[7,8,27,30,56,65], b 被设置为 16 并采用 MAD 作为匹配准则。我们用 x 和 y 分别表示一个像素在一视频帧中行和列的位置序号,同时用 $I_c(x, y)$ 和 $I_r(x, y)$ 分别表示当前帧和参考帧中在 (x, y) 处像素的亮度值。当前宏块的左上角在一帧中的位置用 (x_0, y_0) 来表示。这样,和候选运动矢量 $\vec{V} = (V_x, V_y)$ 相对应的 MAD 值可表示为

$$\text{MAD}(\vec{V}) = \sum_{n=0}^{15} \sum_{m=0}^{15} |B_c(m, n) - B_r(m, n, \vec{V})| \quad (2.1)$$

这里,

$$\begin{cases} B_r(m, n, \vec{V}) = I_r(x_0 + V_x + m, y_0 + V_y + n) \\ B_c(m, n) = I_c(x_0 + m, y_0 + n) \end{cases} \quad (2.2)$$

在式(2.1)中,为不失一般性,我们忽略了除以 256 这个除法。

一般情况下,大小为 $S \times S = (2w + 1) \times (2w + 1)$ 的搜索窗被用来寻找最佳匹配块,并限制了在参考帧中搜索最佳匹配块的范围。这样,运动矢量的值被限制于 $-w \leq V_x \leq w, -w \leq V_y \leq w$ 中。同时,在 $S \times S$ 的搜索窗中的候选运动矢量构成了一个集合

$$P = \{(V_x, V_y) \mid -w \leq V_x \leq w, -w \leq V_y \leq w\} \quad (2.3)$$

通常在视频压缩标准中 w 的取值为 15,这导致了搜索窗的大小为 31×31 。

图 2.1 显示了在运动估计过程中的当前宏块、参考块、搜索窗和运动矢量的相互关系。从图 2.1 中可以看出,从当前宏块左上角的位置到参考帧中参考块左上角的位置的位移就是当前宏块相应的运动矢量。

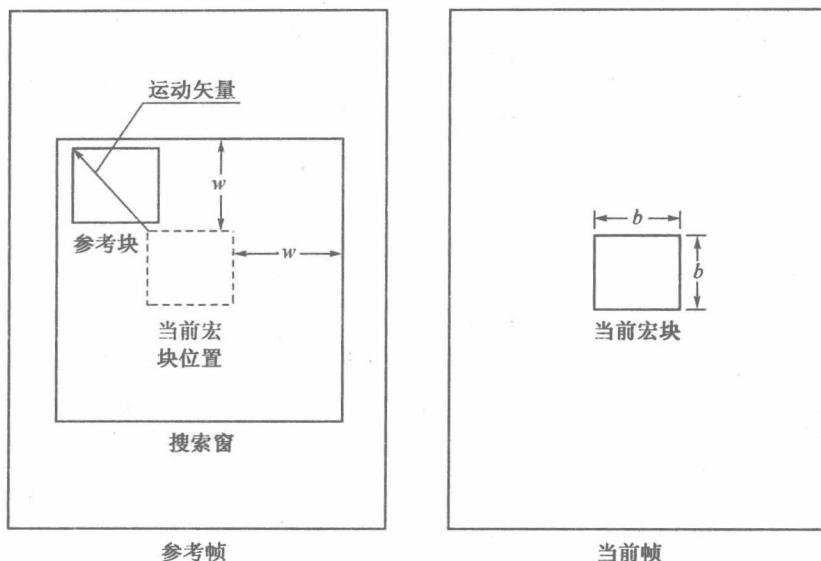


图 2.1 块运动估计

在块运动估计的全搜索算法(FSA)中,对每一个 $\vec{V} \in P$ 的候选运动矢