摘要

视频应用在生产生活中的不断发展，造就了高质量的高清超高清视频编码需求。因此，国内外标准组织也前后制定了新一代的视频编码标准 AVS3 和 VVC，为编码质量提供了技术和规范上的保障。但新一代编码标准具有相对前代更高的复杂度，同时由于还需要满足更高编码吞吐率和质量等新需求，通用处理器已经难以满足视频编码的实时性要求。而专用的硬件编码器由于具有稳定和高效的优点，是实现的首选平台之一。运动估计作为最主要的帧间预测工具，无论性能还是复杂度在编码器中都具有举足轻重的作用。因此，其硬件算法和结构的研究是硬件编码器中的关键路径。而针对新一代编码标准，硬件运动估计存在以下几个关键问题：整像素运动估计中运动矢量预测的高依赖问题和块划分带来的额外复杂度问题，分像素和仿射运动估计中旧工具性能不足和新工具硬件花销大的问题，及整体运动估计中多种运动估计带来的大量资源消耗和带宽占用问题。本文针对上述的几个关键问题，对其进行深入分析，研究低依赖性且低复杂度的硬件友好整像素运动估计方法、低资源消耗的硬件高效分像素和仿射运动计算方法和多级复用的运动估计电路结构，以满足新一代硬件编码器的高效率高吞吐高质量的硬件运动估计需求。本文的主要创新点包括以下三个方面：

第一，提出了硬件友好的整像素运动估计方法。针对运动估计中运动矢量预测依赖性和块划分带来额外复杂度两个关键问题，分析编码树单元级依赖性条件和划分块之间空域上的相关性。对运动矢量预测的依赖性问题，提出了基于仿射运动场的运动矢量预测方法，通过使用仿射运动场和空时域上的运动信息对编码运动场进行建模，从而进行低依赖高精度运动矢量预测；接着，将可划分块分为 BTQT 和 TT 块两类；对 BTQT 块提出由粗到细的多分辨率搜索，在大范围以合理的复杂度和规则的数据流进行运动估计；对 TT 块提出基于误差平面的运动矢量推测方法，通过误差平面建立已估计和未估计块最优运动矢量之间的关系，从而以低复杂度直接推测其运动矢量。所提方法在低时延配置下仅造成 1.20% 的性能损失，但降低了整像素运动估计复杂度的 80%。

第二，提出了硬件高效的复合运动估计方法。针对分像素运动估计中传统误差平面方法性能不足和仿射运动估计中新算法硬件资源花销大的问题，分析了其性能不足和硬件花销大的原因。再通过以较低的额外复杂度在整像素搜索阶段收集额外子块信息，直接利用这些信息直接推断分像素和仿射运动矢量。针对分像素运动估计误差平面性能不足的问题，提出了基于混合误差平面的分像素运动矢量推测算法，通过子块的信息构造更加精准的误差平面，降低硬件花销和复杂度的同时提升了该系列方法的

性能；针对仿射运动估计现有算法硬件花销大的问题，提出了基于过定线性系统的仿射运动矢量推测算法，通过子块运动场构造过定线性系统，从而直接求解仿射运动矢量，降低了硬件花销和复杂度。所提方法在低时延配置下仅仅造成 0.86% 的性能损失，但降低了总体运动估计复杂度的 38.30% 且所需最小硬件资源降低了 75.35%。

第三，提出了多级复用的统一运动估计方法和架构。针对多种运动估计独立实现造成的高硬件花销和高带宽消耗问题，分析三种运动估计的硬件资源和带宽消耗情况。在算法级上，通过结合本文提出的整像素和复合运动估计方法，提出了统一运动估计方法，只进行一次整像素级运动估计后，将小块信息直接用于计算分像素和仿射运动矢量，实现信息复用。在架构级上，将整像素运动估计中和复合运动估计中相同原理的模块统一实现，通过交织调度，实现电路复用。在像素级上，提出新的一种高度并行搜索电路，充分利用已读取的参考像素，实现像素复用。上述多个层面上的设计共同构成了统一运动估计方法和架构，通过多级复用大幅降低硬件资源花销和带宽消耗。所提方法在低时延 P 配置下仅造成 1.88% 的性能损失，但降低了总体硬件运动估计复杂度的 87% 和 99% 的带宽消耗。所提出的电路结构经过 14nm ASIC 平台实现和验证，在 500MHz 工作频率下支持 4k@60fps 实时编码，并在新一代的 AVS3 硬件编码器芯片中得到应用。

综上，本文研究了针对新一代编码标准的硬件运动估计算法和结构，从新编码需求和新编码标准出发，分析了运动估计所面临的关键问题，并针对关键问题分别提出了硬件友好的整像素运动估计方法、硬件高效的复合运动估计方法和多级复用统一运动估计方法和架构。首先解决了运动矢量预测依赖性；再以合理的复杂度为大量待编码块计算了整像素运动矢量；又通过信息复用，利用整像素搜索的额外信息，直接计算分像素和仿射运动矢量；最后通过在电路和像素复用，设计并实现了支持高吞吐率、高编码性能和资源高效的运动估计硬件。本文提出的方法也在新一代的 AVS3 基准档硬件编码芯片中进行了应用，是其高效编码核心模块之一。

关键词：视频编码，硬件编解码，帧间预测，运动估计，硬件电路结构设计

Study on Hardware Motion Estimation Algorithm and Architecture for the Latest Generation Coding Standard

Xizhong Zhu (Technology for Computer Applications)
Directed by: Prof. Xie Xiaodong

ABSTRACT

With the continuous development of video applications in production and daily life, there has been a growing demand for high-quality HD and Ultra-HD video encoding. In response, domestic and international standards organizations have published a new-generation video encoding standards, AVS3 and VVC. They provide technical and regulatory guarantees for encoding quality. However, meeting these new encoding demands further requires the development of encoding products tailored to these new standards. Among them, hardware encoders have attracted considerable attention due to their stability and efficiency. Nevertheless, the new encoding standards possess higher complexity compared to their predecessors, and they must also satisfy new requirements such as higher encoding throughput and quality, posing significant challenges for research. Motion estimation, as the primary inter-frame prediction tool, plays a pivotal role in encoders, both in terms of performance and complexity. Therefore, the research on hardware algorithms and structures for motion estimation is crucial in hardware encoders. For the new generation of encoding standards, hardware motion estimation faces several key issues: the high dependency on motion vector prediction and the additional complexity caused by block partitioning in integer motion estimation; the excessive hardware consumption and insufficient performance of existing tools in fractional and affine motion estimation; and the significant resource consumption and bandwidth utilization caused by multiple motion estimation methods in overall motion estimation. This paper delves into these key issues, exploring methods for hardware-friendly integer motion estimation with low dependency and complexity, hardware-efficient fractional and affine motion computation with reduced resource consumption, and a motion estimation circuit structure with multi-level reuse. The aim is to meet the demands of high efficiency, high throughput, and high-quality hardware motion estimation in the new generation of hardware encoders. The main innovations of this paper lie in three aspects:

Firstly, a hardware-friendly integer motion estimation method is proposed. Addressing the issues of dependency in motion vector prediction and additional complexity from block partitioning, the method analyzes dependency conditions at the coding tree unit level and spatial correlations among partitioned blocks. For the dependency issue in motion vector prediction, a motion vector prediction method based on affine motion fields is proposed. By modeling the coding motion field using affine motion fields and spatio-temporal motion information, low-dependency and high-precision motion vector prediction is achieved. Furthermore, partitioned blocks are categorized, and integer motion vectors are obtained through different methods. For one category, a coarse-to-fine multi-resolution search is proposed, performing motion estimation with reasonable complexity and regular data flow over a wide range. For the other category, a motion vector inference method based on error planes is proposed, which establishes a relationship between the optimal motion vectors of estimated and unestimated blocks, allowing for the inference of motion vectors for unestimated blocks with low complexity. This method incurs only a 1.20% performance loss under low-latency configurations but reduces the complexity of integer motion estimation by 80%.

Secondly, a hardware-efficient composite motion estimation method is introduced. Addressing the issues of high hardware resource consumption and insufficient performance in fractional and affine motion estimation, the method analyzes the reasons for these shortcomings in existing methods. By collecting additional sub-block information during the integer search phase with relatively low complexity, this information is then utilized to directly infer fractional and affine motion vectors. For the issue of high hardware cost and insufficient performance in fractional motion estimation, a fractional motion vector inference algorithm based on a hybrid error plane is proposed. By constructing a more accurate error plane using sub-block information, the algorithm improves performance while reducing hardware cost and complexity. For affine motion estimation, which also suffers from high hardware cost, an affine motion vector inference algorithm based on an overdetermined linear system is proposed. By constructing an overdetermined linear system using sub-block motion fields, affine motion parameters and control point vectors can be directly solved, reducing hardware cost and complexity. This method incurs only a 0.87% performance loss under low-latency configurations but reduces the overall motion estimation complexity by 41.63%.

Thirdly, a unified motion estimation method and architecture with multi-level reuse is developed. Addressing the issues of high hardware cost and bandwidth consumption caused by the independent implementation of multiple motion estimation methods, this approach ana-

lyzes the hardware resource and bandwidth consumption of these methods. At the algorithmic level, a unified motion estimation method is proposed by combining the integer and composite motion estimation methods introduced in this paper. After performing integer-level motion estimation once, small block information is directly used to calculate fractional and affine motion vectors, enabling information reuse. At the architectural level, modules with similar principles in integer and composite motion estimation are unified and implemented, achieving circuit reuse through interleaved scheduling. At the computational level, a new highly parallel search circuit is proposed, fully utilizing the read reference pixels for pixel reuse. These design elements at various levels collectively constitute the unified motion estimation method and architecture, significantly reducing hardware resource cost and bandwidth consumption through multi-level reuse. This method incurs only a 1.88% performance loss under low-latency P configurations but reduces the overall motion estimation complexity by 87% and bandwidth consumption by 99%. The proposed circuit structure has been verified on a 14nm ASIC platform, supporting real-time encoding at 4k@60fps under a 500MHz operating frequency and has been implemented in the new-generation AVS3 hardware encoder chip.

In summary, this paper investigates hardware motion estimation algorithms and structures for the new generation of encoding standards. Starting from the new encoding demands and standards, it analyzes the key issues faced by motion estimation and proposes hardware-friendly integer motion estimation methods, hardware-efficient composite motion estimation methods, and a unified motion estimation method and architecture. Firstly, it addresses the dependency issue in motion vector prediction. Secondly, it calculates integer motion vectors for a large number of blocks to be encoded with reasonable complexity. Thirdly, it utilizes the additional information from integer search to directly calculate fractional and affine motion vectors through information reuse. Finally, through circuit and pixel reuse, it designs and implements a motion estimation hardware that supports high throughput, high encoding performance, and resource efficiency. The proposed methods have also been applied in the new-generation AVS3 benchmark hardware encoding chip, serving as one of its core modules for efficient encoding.

KEY WORDS: Video Coding,Hardware encoder,Inter-prediction,Motion estimation,Hardware Architecture Design