摘　要

随着信息技术的飞速发展，视频信息逐渐成为我们生活中必不可少的一部分。视频编码技术可以实现视频压缩，大大降低视频信息的数据量。因此，视频编码技术已经成为多媒体产业的核心技术之一。当前，新一代视频编码标准的研究正处于方兴未艾的阶段，是一个充满机遇与挑战的研究领域。在这一背景下，开展视频编码技术的研究具有非常重要的学术意义与产业意义。

高效预测算法的研究在视频编码领域中具有突出意义。预测技术是视频编码系统中最重要的核心技术之一，众多学者围绕预测编码提出了大量新技术、新方法。但现有的预测技术仍然存在一些问题：在帧内预测技术方面，没有充分考虑到预测与变换、熵编码等模块的结合，同时缺乏对预测模式本身的高效编码方式；在帧间预测技术方面，没有进一步挖掘帧间预测后残差域的相关性，同时各种高效的亚像素插值算法又带来了高复杂度的缺陷；在层间预测方面，缺乏对层间预测效率不高情况的判断与处理，导致在某些情况下可伸缩编码与单层编码相比压缩效率大为下降。

本文针对上述问题，围绕着预测技术这一主题，在帧内编码、帧间编码以及可伸缩编码等方面开展了一系列的深入研究，主要贡献如下：

1、在基于纹理特征的高效帧内预测编码方面，提出两项创新技术

第一项是基于上下文的自适应帧内预测模式判断技术。在现有的视频编码标准当中，如 H.264/AVC，使用了较多的帧内预测模式来对 Intra_4x4 做方向预测，这可以有效地提高帧内预测的效率，但也带来了两个问题。一方面，由于编码器需要遍历每个模式进行 RD 最优选择，较多的模式导致编码器的复杂度大大增加；另一方面，较多的帧内模式要求很大一部分码流用来传输模式信息，制约了帧内编码效率的进一步提高。针对上述问题，本文提出基于上下文的自适应帧内模式判断技术。这一技术首先建立帧内模式的 Markov 场模型，接着构造基于这个模型的表结构。此技术可以有两个应用方案。一是应用在快速算法中，不改变原有的 H.264/AVC 的语法；二是应用在提高模式信息编码效率上，改变语法。全 I 帧条件下的实验表明，本技术应用于快速算法时，可以在平均 PSNR 损失不超过 0.1dB 的前提下，节省大约一半的编码时间；应用于提高性能时，最高可以使 PSNR 提高约 0.3dB。

第二项是叠加块帧内预测技术与基于帧内预测的多模型 KL 变换技术。在新一代视频编码标准的研究热潮中，方向变换技术(MDDT)以其高效性与简洁性引起了人们的普遍关注，并被采纳进入 VCEG/KTA 参考软件。为了进一步提高帧内编码的效率，本文提出两种新的方法。首先，引入叠加块帧内预测的技术。在这一技术中，对某一位置的预测将利用三个可能的预测值做线性组合；这三个可能的预测值分别利用当前块、左侧块和上侧块的预测模式生成。线性组合的加权值则将利用经典的线性回归算法进行估计。其次，提出基于帧内预测的多模型 KL 变换技术(MMKLT)，以进一步加强 MDDT 的性能。研究发现某一残差块的最优变换并不只取决于其预测模式，同时还取决于残差的方差。因此，本文为每种帧内预测模式预先定义三类 KLT，这些 KLT 是根据不同的残差方差离线训练得到的。编码器根据 RDO 准则在宏块级选择最佳的变换类，并将所选择的变换类发送给解码器。实验表明，这两种方法结合使用，全 I 帧条件下平均可以比 MDDT 节省约 5%的码率，比 H.264/AVC 节省约 11%的码率。

2、在基于运动补偿的高效帧间预测编码方面，提出三项创新技术

第一项是基于运动对齐的残差域方向预测技术。基于块的运动补偿技术可以有效减少视频序列中的时间方向的冗余度，在现代视频编码标准中占据非常重要的地位。随着视频编码技术的进步，基于块的运动补偿越来越精细，但经过运动补偿之后，残差之间仍然会有一定程度的冗余存在。为了进一步减少残差的冗余性，提高运动补偿的能力，近年来学者们开始试图把 H.264/AVC 中的帧内方向预测技术与帧间预测相结合。尽管这些技术给了人们很大启发，但它们对编码效率的提升十分有限。针对上述问题，本文提出了基于运动对齐的残差域方向预测技术(DRP-MA)，创造性地在残差域上使用方向预测，使编码效率得到了显著提升。DRP-MA 的核心思想是采用当前块的运动矢量生成运动对齐的伪残差，再利用伪残差对当前块的真实残差进行方向预测。本文还给出了完全搜索和简化搜索两种针对 DRP-MA 的编码策略。实验表明，本技术在 IPPP 条件下最高可以节省约 20%的码率。

第二项是低复杂度八分之一精度亚像素插值技术。AVS 视频编码标准采用了四分之一像素精度的运动补偿技术，其插值的精度还有进一步提升的空间。本文提出了一种低复杂度八分之一像素插值算法。八分之一像素按位置被分为两类；分别使用两种 FIR 滤波器插值得到这两类八分之一像素；插值的过程只利用整像素与二分之一像素，而不依赖于四分之一像素。实验表明，采用本文提出的八分之一像素插值算法，AVS 视频编码器平均可以节约码率 6%以上，最高可达 11.5%；同时，编解码器的复杂度增加不大。本文所提出的这一技术已经被面向移动的 AVS 标准草案采纳。

第三项是单次编码的局部自适应插值滤波技术。最近，人们提出了一种新的高效编码工具，称为自适应插值滤波技术(AIF)。AIF 在亚像素插值过程中，引入了维纳滤波器，可以将帧间预测的误差降至其理论最小值，从而显著地提高了编码效率。但是从另一个角度来说，基于训练的维纳滤波机制也给 AIF 带来了固有的多次编码结构，这大大增加了编码器的计算复杂度和内存访问次数。针对这个问题，本文提出了一种基于单次编码的局部自适应插值滤波算法(SPL-AIF)。它既可以有效降低 AIF 的复杂度，又可以保留其良好的编码性能。研究发现相邻帧的最优插值滤波器有着很高的相关性，而同一帧的不同区域通常有着不同的统计特性。基于这些观察，本文设计了 SPL-ALF 算法。首先，为当前编码帧建立一个竞争滤波器集，它包含了前几帧的最优化插值滤波器和固定的 H.264/AVC 插值滤波器。接下来，使用 RDO 准则，在宏块级上选择最合适的插值滤波器。最后，为了降低额外信息，使用预测编码方法对每个 MB 的滤波器类型进行压缩。实验结果显示，本方法可以获得与多次编码 AIF 算法相当的编码效率，但编码复杂度却比 AIF 算法有大幅的降低。这一算法被 VCEG 采纳并进入了 VCEG/KTA 软件。

3、在基于最优化模型的高效层间预测编码方面，提出一项创新技术

主要提出了选择性层间预测技术。JVT 在 H.264/AVC 的基础上发展出了可伸缩视频编码标准(SVC)。为了更好地支持空间可伸缩性，SVC 引入了一种新的预测方式，即层间预测方式，基本思想是利用低分辨率码流对高分辨率码流进行预测，以提高高分辨率码流的编码效率。当需要播放多个分辨率的视频节目时，SVC 的空间可伸缩码流可以比联播(Simulcast)码流有较大幅度的性能提高。然而，SVC 有很多应用场景。SVC 的空间可伸缩编码方式不适用于终端设备只需要播放最高分辨率的视频节目的情况。这是由于层间预测给增强层所带来的益处并不能补偿基本层码流本身所带来的额外码率。针对上述问题，本文提出了选择性层间预测技术(SIP)。通过禁用某些选定帧上的层间预测，SIP 可以在空间可伸缩编码与联播编码之间取得最优的平衡。在理论上，这一选择问题可以用 0-1 背包问题进行建模，并用动态规划的方式求解。实验结果表明，在不需要基本层的情况下，SIP 技术可以在增强层上取得高达 1dB 的编码增益；而在需要基本层的情况下，SIP 在编码效率上的损失并不明显。SIP 技术在 JVT 中被采纳，进入 JSVM 参考软件。

关键词：视频编码；H.264/AVC；AVS；SVC；帧内预测；帧间预测；层间预测；方向预测；运动补偿；亚像素插值；

Study on Efficient Prediction Techniques in Video Coding

Kai Zhang (Computer Application Technology)

Supervised by Professor Wen Gao

With the rapid development of information technology, variety of video information has become an indispensable part of people's daily life. Video coding can compress video information and decrease the amount of video data dramatically. Therefore, the technology of video coding has become a central technology in multiple media industry. Nowadays, the study on the next generation video coding standards has just been unfolding, and it is a hot research topic with challenges and chances. In such a wave, the research on video coding becomes more and more important both in academic and industry.

The research on efficient prediction algorithms takes a central role in the area of video coding, because the prediction technology is one of the most important technologies in the video coding system. Although many researchers have proposed variety of techniques to improve prediction efficiency, there are still several shortcomings in existing methods. In intra-prediction, the combination of prediction with other modules such as transform and entropy coding has not been considered seriously. In addition, the coding method of the prediction mode is not efficient enough. In inter-prediction, the correlation among residues after inter-prediction has not been utilized sufficiently. Moreover, novel sub-pixel interpolation techniques which can obtain high performance also introduce high complexity as a disadvantage. In inter-layer prediction, spatial scalable video coding presents a significant loss compared with single-layer video coding in some cases, due to the lack of determination and treatment on the situation where the inter-layer prediction is inefficient.

Towards the above problems, we make several efforts on the study of intra-coding, inter-coding and scalable coding around the theme of prediction in this dissertation. The main contributions of this thesis are composed of:

1 Two innovations on efficient texture-based intra prediction coding

The first is a context-adaptive intra-prediction mode decision method. In state-of-art video coding standards such as H.264/AVC, quite a number of intra-prediction modes are

utilized to make directional intra-prediction for Intra_4x4 blocks. Despite of the high prediction efficiency, this brings forward two problems. On the one hand, more modes result in high encoding complexity due to the RD decision procedure. On the other hand, more modes require more overhead to convey the mode information, which restricts improvement of coding efficiency. Towards these problems, we propose a context-adaptive intra-prediction mode decision method. First, a Markov field model is built for intra-prediction modes. Then a table structure is constructed based on this model. This model can be utilized in two applications. On the one hand, it can be used in a fast encoding algorithm, which does not change the syntax structure of H.264/AVC. On the other hand, it can be used to improve the mode information coding, which should change the syntax structure. Experimental results show that the proposed fast algorithm can save about half the encoding time while keeping the PSNR loss less than 0.1dB in all I-frame coding. Moreover, the proposed high efficient coding method can improve the PSNR up to 0.3dB.

The second is overlapped-block intra prediction and multiple model KLT. Intra-coding plays an important role in video coding schemes and it has become a hot research topic of the next generation video coding standard. In this thesis, we propose two techniques to improve the efficiency of the prediction and the transform, respectively, for intra-coding. First, we introduce an overlapped block intra-prediction (OBIP) method, which makes use of the intra directional prediction modes of neighboring blocks as well as the mode of the current block. A position dependent weighted sum of several possible predictors will be treated as the final prediction. Second, we propose a multiple-model Karhunen–Loève transform (MMKLT) technique to further improve the mode dependent directional transform (MDDT) method. More than one KLT will be trained offline for each intra-prediction mode based on residual variances. Experimental results indicate that the proposed two methods together can achieve an average bit-rate reduction about 11% compared with H.264 or 5% compared with MDDT in all-intra coding.

2 Three innovations on efficient motion compensation-based inter prediction coding

The first is directional residue prediction with motion alignment for video coding. Block-based motion compensation (MC) can reduce the temporal redundancy significantly and plays a crucial role in predominant video coding standards. Although block-based MC becomes more precise with the development of video coding technology, there is still some

redundancy between residues. To overcome this problem, some researchers attempted to combine motion compensation with the intra-prediction adopted by H.264 recently. Although these ideas inspire us, they have not shown an apparent gain yet. We propose a new technique named directional residue prediction with motion alignment (DRP-MA), which also introduces the concept of directional intra-prediction into the residual domain. Different from existing methods, DRP-MA calculates motion-aligned neighboring residues, which are called pseudo residues, to predict residues in the current block with directions. Accordingly, both the optimized motion estimation (ME) strategy and the ad-hoc ME strategy are designed to take advantage of the potential capability within DRP-MA. Experimental results show that DRP-MA can achieve up to 20% bit-rate reduction in IPPP coding.

The second is a low complexity 1/8-pixel interpolation method. AVS-P2 utilizes a motion compensation technique with 1/4 sub-pixel precision, which can be improved furthermore. In this paper, we present a low complexity 1/8 sub-pixel interpolation method. 1/8 sub-pixels are classified into two kinds and can be generated by using two types of FIR interpolation filters respectively. The interpolation for 1/8 sub-pixels depends on integer pixels and 1/2 sub-pixels only, without depending on 1/4 sub-pixels. Experimental results indicate that the proposed method can save the bit-rate of AVS-P2 6% in average and up to 11.5%. Meanwhile, the complexity only increases slightly. The proposed method has been adopted into the draft of AVS for mobile applications.

The third is a single-pass based localized adaptive interpolation filter algorithm. Recently, a very powerful coding tool named adaptive interpolation filtering (AIF) has been proposed to hybrid video coding scheme. By introducing wiener filter into the fractional-pixel interpolation procedure, AIF can decrease the inter-prediction error to its theoretical minimum, so as to improve coding efficiency significantly. On the other hand, the training-based wiener filter mechanism brings AIF an inherent multi-pass encoding structure, which imposes big burdens on the encoder in terms of huge computational complexity and memory access. In this thesis, we propose a single-pass based localized adaptive interpolation filter (SPL-AIF) algorithm for video coding, which can reduce the complexity of AIF dramatically as well as preserve its outstanding coding performance. As a motivation, we have observed that there is a high correlation among optimal interpolation filters of consecutive frames, and different regions in a frame often possess different statistic characteristics. We design the SPL-ALF algorithm based to these phenomena. At first, a competitive filter set which includes optimal interpolation filters of several previous frames as well as fixed H.264/AVC interpolation

filters is built up for the currently coding frame. Then a rate-distortion optimization (RDO) criterion is used to select the most proper one at macro-block (MB) level. Finally, in order to reduce overhead, a predictive coding method is used to compress the filter type for each MB. Experimental results show that, by using the proposed algorithm, the encoding complexity can be reduced largely and an average coding gain can also be obtained compared with the multi-pass AIF. The proposed method has been adopted into the VCEG/KTA software.

3 One innovation on efficient optimization-based inter-layer prediction coding

We propose selective inter-layer prediction in scalable video coding. In the scalable video coding (SVC) standard, spatial scalable coding outperforms simulcast coding when programs with several display resolutions are needed. Nevertheless, it is not suitable for end devices which only need the high resolution, due to a serious performance loss on the high spatial layer compared with single layer coding. To tackle this dilemma, a selective inter-layer prediction (SIP) method is presented in this thesis. SIP attains an optimal trade-off by disabling inter-layer prediction on a set of selected frames. Theoretically, this selection can be modeled as a 0-1 knapsack problem which can be solved by dynamic programming. Experimental results show that the proposed method can achieve significant gains up to 1 dB on the high spatial layer when the content of the low spatial layer is not needed, and can keep the loss unapparent even when it is. The SIP method has been adopted into the SVC reference software JSVM.

Keywords: Video coding, H.264/AVC, AVS, SVC, intra-prediction, inter-prediction, inter-layer prediction, directional prediction, motion compensation, sub-pixel interpolation filter