

# 摘要

视频信息生动直观，内容丰富，是传递信息的理想载体。但未经压缩处理的视频数据，其数据量极其庞大，给信息的存储和传播带来了极大的困难，所以压缩效率一直是视频编码领域关注的焦点。随着 Internet 的迅猛发展，3G 移动通信和无线宽带技术的成熟，网络逐渐成为人们日常生活的一部分，数字视频压缩的重心也开始从面向存储转为面向传输，而网络传输的新特点及各种具有不同分辨率和处理能力的多媒体终端的不断涌现，为视频编码应用提供了更广阔的舞台，同时也提出了更多新的挑战。可伸缩视频编码技术以其高度的灵活性，可以适应各种网络环境、各种用户终端及较高的编码效率而得到了学术界的青睐。本文首先针对帧间亮度变化剧烈时运动补偿预测效率会大大降低这一情况，讨论了适应于帧间亮度变化的运动补偿预测方法。随后将主要就可伸缩编码中的空域可伸缩功能的实现进行研究探讨，提出了基于联合运动估计的空域可伸缩视频编码框架及利用 DCT 系数切分实现空域可伸缩功能的方法。本文的主要研究成果如下：

1. 提出了亮度变化自适应的运动补偿预测方法，可以减弱视频帧间的各种亮度变化带来的影响，大大提高运动补偿预测的效率。

运动补偿预测是视频编码的关键技术，大多视频编码系统都采用了基于 SAD 块匹配的方法进行运动补偿预测。而 SAD 块匹配准则在视频帧间有明显的亮度变化时会严重影响运动估计的准确性，增大残差系数，降低运动补偿预测的效率；而且预测效率的突然降低还会引起帧码率的突变，增大码率控制的难度，不利于网络传输。这一问题得到了学术界的高度重视，很多处理方法被提了出来。但已有方法大都针对镜头切换时的淡入淡出特效情况提出来的，通常在帧一级或条带 (Slice) 一级作处理，这对全局性均匀渐变的亮度变化有较好的效果，但对全局非均匀或局部的亮度变化效果就较差，而后两种亮度变化也是普遍存在的。本文提出的亮度变化自适应的运动补偿预测方法，在宏块一级做调控，运用新的变换及参数计算编码方法，可以很好的处理全局和局部的亮度变化，大大提高运动估计的准确性，使视频帧码率更为平滑，极大提高视频压缩效率。

2. 提出基于联合运动估计的空域可伸缩视频编码框架，在少量降低基本层编码效率的情况下，大大提高增强层的编码效率。

视频码流主要包含残差，运动向量和模式选择三项数据。而空域可伸缩视频编码中低分辨率的视频通常由高分辨率的视频下采样得来，它们之间在这三个方面都存在着大量的冗余。MPEG-2 中提出的空域可伸缩视频编码框架以

其简单实用而得到了广泛的应用，很多改进算法都是基于这一框架提出来的。但该框架只部分消除了残差方面的冗余，编码效率不高是它的最大缺点。本文提出的基于联合运动估计的空域可伸缩视频编码框架，通过联合基本层和增强层作运动估计，只生成一套运动向量和模式选择，这样可以极大的减少运动向量和模式选择方面的冗余；同时使用了将基本层重构图像作为增强层预测图像的选择之一的方法，可以很好的减少残差方面的冗余。这样就可以减少上述三个方面的冗余，大大提高增强层的编码效率。该框架为了保证增强层的运动补偿预测效率，设定运动向量精度是基于增强层的，这个增强层精度的运动向量将编入基本层中，必然会增加基本层的负担。但通过在基本层追加插值及调整运动估计参数等方法，可以保证基本层性能的下降在可接受范围之内。相较于增强层压缩效率的提高，这个基本层性能的下降是完全值得的。

3. 提出增强层与基本层在频域内结合生成增强层预测图像的方法。

在 MPEG-2 空域可伸缩编码框架中，为了减少基本层和增强层之间的冗余，增强层的预测图像有三个。一是增强层自身作帧内或帧间预测产生，二是基本层重构图像上采样，三是前两者的均值结合。其中第三个预测图像与前两个有重复之感，效果不够好。考虑到使用 DCT 变换下采样时，基本层图像是增强层图像去掉高频信息后的产物。因此在生成增强层预测图像时，可以走一个相反的过程，即以基本层为低频部分，结合增强层的高频部分，构成一个具有完整频域信息的图像，然后反变换即可得到一个新的增强层预测图像。以这个频域内结合生成的预测图像代替第三个预测图像，可以收到很好的效果。

4. 提出通过 DCT 域中系数切分实现视频编码空域可伸缩功能的方法。

标准 DCT 变换有一个重要的特性，即对一个  $N \times N$  的图像块做  $N \times N$  的变换后，取其左上角  $M \times M$  的系数，乘上  $M/N$  的放缩因子，再做一个  $M \times M$  的 DCT 反变换，可以得到一个  $M \times M$  的图像块。这个  $M \times M$  图像块是原来  $N \times N$  图像块的缩小版。DCT 的这个特性常被用来作为图像下采样的一种方法，在视频转码中有很多研究应用。我们也可以利用这一特性来实现视频的空域可伸缩。在对视频预测的残差系数做 DCT 变换之后，左上角的低频信息自动地对应于低分辨率的视频图像。因此可以考虑将左上角低频信息单独编码生成基本层，其余高频信息编入增强层，结合分块的残差系数扫描方式，可以简单高效的实现我们需要的空域可伸缩功能。该方法的关键在于作运动估计时的参考图像选择，如果选择基本层则效率太低，而如果选择增强层，则在解码基本层时会有漂移误差 (drifting)。我们设定阈值来控制漂移误差的累积程度，通过恢复机制以达到提高编码效率和控制漂移误差之间的平衡。

**关键词：**视频编码；运动补偿；可伸缩视频编码；亮度变化自适应；H.264/AVC

## **Research on motion compensated prediction and spatial scalable video coding techniques**

Debing Liu (Computer Application Technology)

Directed By Prof. Wen Gao

Video is vivid and it contains much information, which makes it to be an ideal media for information distribution. Since original video data is too large to save or transmit, video coding techniques become very important and valuable. With the rapid development of Internet, WLANs and 3G mobile telecommunication systems, enjoying video online become a big opportunity for media business. It's also a big challenge to make people to share the video online as many as possible. The idea of scalable video coding has gotten much attention for its high coding efficiency and adaptability to bandwidth change. In this paper, we will firstly discuss a new motion compensated prediction scheme. It can improve the coding efficiency greatly for the sequences that have abruptly luminance intensity change inter frames. After that, we will focus on the spatial scalable video coding. A new coding framework based on combined motion estimation is proposed and a new method using DCT coefficients division to realize spatial scalability is introduced. The main contributions of this paper are listed as follows:

1. It proposes new motion compensated prediction scheme adaptive for inter frame luminance intensity change.

Motion compensated prediction (MCP) is one of the most important techniques of video coding. The main international video coding standards such as MPEG and H.26x have their own MCP component. Most of them are realized based on SAD block match. When inter frame luminance intensity changes abruptly, the SAD block match won't work well. It will influence the Motion estimation and then result in the coding efficiency dropping. Many researchers have investigated to solve the problem. But most of them aim at the luminance change of scene transition, so the methods they proposed are effective for global uniform luminance change. We propose a new method which is applied on macro block level. Either uniform or non-uniform luminance change can be well processed. It can smooth the frame rate and improve the coding efficiency greatly for the sequences that have abrupt luminance intensity change inter frames.

2. It proposes a new framework for spatial scalable video coding based on combined motion estimation.

Video stream mainly includes three kinds of data: motion vectors, mode selections and residual coefficients. While realizing a spatial scalable video coding scheme, there is much redundancy between the base and enhancement layer in those three aspects. The goal of spatial scalable video coding is to decrease the redundancy as much as possible. MPEG-2 spatial scalable video coding framework is one of the most popular frameworks in video coding field. It's simple and effective which makes it to be used widely. Many methods have derived from it. But it only partly decreases the residual coefficients redundancy, so its coding efficiency is not high. We propose a new spatial scalable video coding framework which combines the base and enhancement layer to do motion estimation. It only generates one set of motion vector and mode selection. Moreover, the base layer reconstructed picture is used for one of the predictions of the enhancement layer. Thus it can decrease all the three kinds of redundancies and then improve the coding efficiency greatly.

3. It proposes new method to generate one of the predictions of enhancement layer based on combining base and enhancement layer in frequency field.

In MPEG-2 spatial scalable video coding framework, there are three predictions for enhancement layer. The first is generated from the intra or inter prediction of the enhancement layer itself; the second comes from the up-sampled base layer reconstructed picture and the last is the average combine of the former two predictions in pixel domain. It is the second and the last predictions that make the redundancy between the base and enhancement layer decreased. It should be noted that the last prediction is a little reduplicate with the former two, so its prediction efficiency is not so good. Because the coded picture of base layer is generated from the enhancement layer with some high frequency information being erased, we can keep the base layer reconstructed picture as low frequency part and combine the high frequency part of enhancement layer. Thus we can get a picture which has all the frequency information. The coding efficiency is improved when use the new generated picture to replace the last prediction.

4. It proposes new spatial scalable video coding scheme based on DCT coefficients division.

There is an important feature for standard DCT (Discrete Cosine Transform). That is after an  $N \times N$  DCT on an  $N \times N$  block of picture, take its up-left  $M \times M$  coefficients, multiply a scale factor  $M/N$  and apply an  $M \times M$  reverse DCT on it, we can get an  $M \times M$  picture which is the down-sampled picture of the original

NxN picture. This feature is often used to down sample picture and is widely utilized in video transcoding field. We can use this feature to realize the video spatial scalability. After a DCT on the residual coefficients, encoding its up-left MxM coefficients in base layer and the rest in enhancement layer, thus the spatial scalability is realized. It's simple and effective. The Key problem is how to select the reference picture. If the base layer reconstructed picture is up-sampled for reference, the enhancement layer coding efficiency will drop greatly. While if the enhancement layer reconstructed picture is used for reference, the terminals which only get the base layer stream will mismatch in decoder (drifting error). We set a threshold to control the drifting error. When the drifting error is below the threshold, it uses the enhancement layer as the reference to improve the coding efficiency; if the drifting error is above the threshold, it uses the base layer as the reference to recover.

**Keywords:** Video coding, motion estimation, motion compensation, scalable video coding, luminance intensity change