第 29 卷 第 12 期
2009 年 12 月

光 学 学 报
ACTA OPTICA SINICA

Vol. 29, No. 12
December, 2009

文章编号：0253-2239(2009)12-3351-06

三维电视系统中的视频-深度联合预测编码

杨海涛 常义林 霍俊彦 元 辉 刘晓仙

（西安电子科技大学 ISN 国家重点实验室，陕西 西安 710071）

摘要 深度图像能够有效表示三维场景几何信息，需要传输至三维电视终端用以辅助生成任意视点虚拟视图。为降低深度图像传输开销，需要对其进行压缩编码。深入分析了视频图像运动信息与深度图像运动信息的相似性，提出一种视频图像与深度图像联合预测编码方案，在编码深度图像过程中重用已编码视频图像的运动信息。该方案由视频－深度运动信息复制与视频-深度运动信息预测两部分组成。实验表明，提出的视频-深度联合预测编码能够高效利用已编码视频图像的运动信息，显著提高深度图像编码效率。

关键词 图像处理；三维电视；深度图像编码；视频图像编码；运动信息复制；运动信息预测

中图分类号 TN911.73 **文献标识码** A **doi**：10.3788/AOS20092912.3351

Joint Video-depth Predictive Coding for 3D. Television Systems

Yang Haitao Chang Yilin Huo Junyan Yuan Hui Liu Xiaoxian

(*State Key Laboratory of Integrated Services Networks*, *Xidian University*, *Xi'an*, *Shaanxi* 710071, *China*)

Abstract Scene geometry can be represented by depth images, which need to be transmitted to 3DTV terminals to assist in rendering of virtual view at an arbitrary view-point. Efficient compression is needed to reduce the expenditure for transmitting depth images. Similarity of motion information between video and depth images is analyzed, and a joint video-depth predictive coding scheme is proposed to reuse the motion information obtained in video coding to assist depth coding. The coding scheme comprises two parts, video-depth motion information copy and video-depth motion information prediction. Experimental results show that the proposed coding scheme can utilize the motion information of encoded video efficiently, and improve the coding efficiency of depth images significantly.

Key words image processing; three dimensional television (3DTV); depth image coding; video coding; motion information copy; motion information prediction

1 引 言

从 20 世纪 20 年代第一台电视机诞生至今，电视技术发展经历了从黑白到彩色、从模拟到数字的两次技术飞跃，如今电视已成为人们不可或缺的家庭娱乐设备。在现有普通二维电视系统中，用户无法直接从二维视频图像中获得自然的深度感觉，也就是通常所说的立体感，而是依据近大远小的透视原理以及关于物体大小的先验知识推断电视场景中各对象的远近深度关系。此外，用户观看电视中三维场景时所处的相对空间位置－视点，与所选取的观看角度-视角都由摄像机的三维空间位置和方向决定，而非用户自由选择。

为使用户能够自由选择观看的视点与视角，体验身临其境的观看效果，许多研究机构对各种三维电视（3DTV）系统及其相关技术进行了深入研究[1~6]。视频标准化组织运动图像专家组（MPEG）

收稿日期：2008-12-20；**收到修改稿日期**：2009-03-06

基金项目：国家自然科学基金（60772134）、陕西省自然科学基金（SJ08F03）、高等学校创新引智计划（BO8038）和西电-华为多媒体通信联合实验室合作专项基金资助课题。

作者简介：杨海涛（1983—），男，博士研究生，主要从事视频图像处理、编码、通信等方面的研究。

E-mail：htyang@mail.xidian.edu.cn

导师简介：常义林（1944—），男，博士生导师，主要从事多媒体通信和网络管理等方面的研究。

E-mail：ylchang@xidian.edu.cn

也从 2002 年起开始研究三维视频编解码技术[7]。

研究 3DTV 系统编码，首先需要确定三维场景信息格式。目前存在双目立体视频、多视点视频、单视点视频结合单视点深度、多视点视频结合多视点深度等多种三维场景信息格式。其中双目立体视频通过向人的左、右眼播放对应的视频信息从而提供最基本的立体视觉。多视点视频则进一步提供了有限的视点切换功能，用户可选择视点位置相邻的两路视频信号同时观看。然而，多视点视频数据量随视点数量线性增加，若要满足用户在任意视点以任意视角观看的需求，就必须使用密集排列的摄像机阵列对三维场景进行高密度采样，这显然无法实用。因此，欧洲三维电视系统技术(ATTEST)项目[2]采用单视点视频结合单视点深度表示三维场景信息。其中深度信息表示为具有像素精度的深度图像，记录对应视频图像中每一个像素点所表示三维场景对象的深度。基于场景深度信息，可在 3DTV 终端采用基于深度图像绘制(DIBR)技术[8]生成任意视点位置虚拟图像。但是该方案生成虚拟视图时可选择视点、视角范围较小，集中在所传输单路视频附近，且无法获取被遮挡的场景信息，合成得到的虚拟视图质量较差[8]。为了能够在较大视点与视角范围内的任意视点、视角生成高质量虚拟视图，可使用在多个视点位置摄像机采集的多视点视频和与之对应的多视点深度来表示三维场景信息。MPEG 中的三维视频(3DV)工作组已经开始探索多视点视频与多视点深度相结合的格式数据压缩编码方法[9]。可见，深度图像压缩编码已经成为 3DTV 系统关键技术之一。

基于传统视频编码标准，如 MPEG-X、H. 26X 的深度图像压缩编码方法[1,10~12]具有便于实现、高压缩比的优点，且能得到高质量恢复图像，因此被广泛采用。基于网格的视频编码方法也可用于深度图像序列压缩[13,14]。这类方法基于内容自适应非均匀采样原理去除单幅深度图像内空间冗余，并可进一步采用时间相邻图像网格节点运动估计与补偿的方法去除深度图像序列的时间冗余。此外，还可基于小波变换对深度图像进行压缩编码[15]。

本文研究基于传统视频编码标准的深度图像压缩方案。介绍了现有基于传统视频编码标准的深度图像压缩技术；分析了视频运动信息与深度运动信息相似性特征的基础上，提出一种基于运动信息重用的视频-深度联合预测编码方案；并实验验证提出编码方案压缩性能。

2 基于传统视频编码标准的深度图像压缩编码

图 1 为 3DTV 标准测试序列 Interview。其中(a)为视频图像，由普通摄像机采集得到；(b)为深度图像，由深度摄像机采集得到。深度图像是一种特殊的视频图像，可使用传统视频编码标准进行压缩编码。这种编码方法的优点是便于实现，具有良好的后向兼容性，并且具有较高的压缩效率。MPEG 中的三维自适应视频(3DAV)研究组针对该方案进行了初步探索[7,16]。

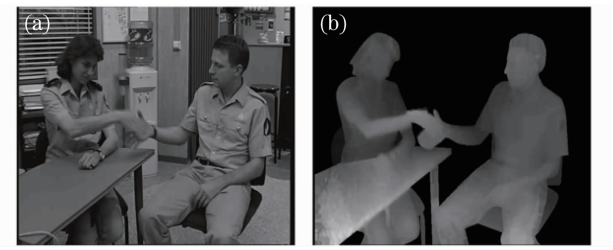

图 1 3DTV 标准测试序列 Interview 的视频与深度图像

Fig. 1 Video and depth of 3DTV standard test sequence Interview

文献[16]研究表明，使用 MPEG-X 与 H. 26X 编码标准压缩深度图像，仅需使用大的量化步长就可以得到高质量恢复图像。换言之，深度图像含有大量冗余信息，具有极高的压缩比。此外，在已有视频编码标准中，H. 264/AVC 编码效率最高。观察图 1 中深度图像可以看到，背景大面积区域深度值为 0，而 H. 264/AVC 中的连续跳过宏块(skipped macroblock)编码方法适用于此类视频图像。再者，位于同一对象区域的深度值呈平滑变化，甚至表现为大面积平坦区域，而 H. 264/AVC 引入的高效帧内预测编码模式在很大程度上能够去除这种空间冗余。此外，由于使用了更加灵活的可变块大小预测编码，H. 264/AVC 能够对 4×4 大小的图像块进行独立预测编码，因此能够更加准确地描述对象边缘的深度变化。可见，H. 264/AVC 中引进的编码技术能够适应深度图像统计特性，从而达到较高的压缩性能。

除使用已有视频编码标准独立编码深度图像，还可使用对应视频图像的编码信息辅助编码深度图像。视频图像与对应深度视频图像间具有极强的相关性，表现为对象边界的一致性和对象运动的一致性。利用此相关性进行视频-深度联合编码将进一步提高压缩编码效率和编码速度。考虑运动相似性，可在运动过程中使用与(1)式所示的传统判决条

件不同的判决条件，同时考虑视频图像与深度图像失真估计得到最优运动的向量。

$$mv = (u,v) = \min_{u,v}\{e_{\text{video}}(u,v)\}, \quad (1)$$

文献[17]使用相同的运动向量 mv 对视频图像与深度图像进行补偿编码。

$$mv = (u,v) = \min_{u,v}\{e_{\text{video}}(u,v) + e_{\text{depth}}(u,v)\}, \quad (2)$$

其中 $e(u,v)$ 为编码图像块与其在参考图像中偏移 (u,v) 后找到匹配图像块的平均绝对差值。结果表明在视频与深度间共享运动信息能够提高编码效率。文献[11]提出另一种在视频图像与深度图像间共享运动信息的编码策略。该方法首先编码视频图像，接着在编码深度图像时基于视频图像中运动信息通过分裂与组合操作得到 16×16，16×8，8×16，8×8 等各种帧间预测编码宏块模式下的运动信息直接用于运动补偿编码。由于在深度图像编码时省去运动估计，因此可在提高编码效率的同时减少编码时间。

3　视频图像与深度图像运动相似性分析

文献[11,17]研究表明，视频图像的运动信息与深度图像的运动信息在统计意义上存在相似性。下面通过实验进一步研究这种相似性的具体特征。选择 H.264/AVC 编码标准，使用 IPP 预测结构且仅使用 16×16 帧间预测编码宏块模式分别编码测试序列 Interview 与 Ballet 的视频图像与深度图像。序列 Interview 深度图像使用深度摄像机采集得到，而序列 Ballet 深度图像使用立体匹配算法[18]估计得到。

图 2 为 Interview 与 Ballet 测试序列中视频图像与对应深度图像的运动向量分布图。观察发现，尽管视频图像与深度图像运动向量存在极大的相似性，但具有不同的特性。具体地说，视频图像运动信息更加具有整体性与规则性，这是由于视频图像中包含丰富的纹理信息，从而存在更多的特征点，便于在运动估计中搜索得到更准确的匹配块。而深度图像整体非常平滑，特征点较少，因此不易搜索得到准确的匹配块，并会因错误匹配产生较大幅度的运动向量。此外，现有立体匹配算法本身就无法计算得到平坦区域准确深度值，对应的原始深度图像区域存在固有的时间抖动效应，例如 Ballet 深度图像序列中地板区域深度值表现出明显的时间抖动。这导致编码地板区域图像块时产生大幅度运动向量。

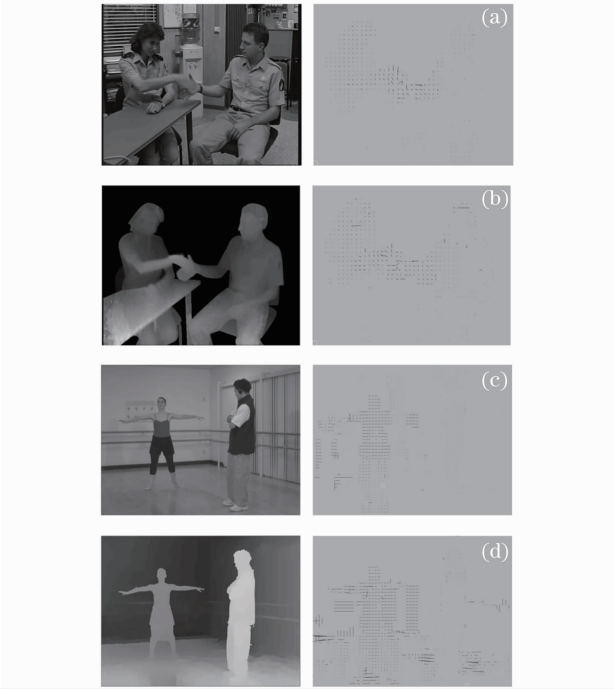

图 2　测试序列中视频图像与深度图像的运动向量分布图。(a) Interview 视频图像；(b) Interview 深度图像；(c) Ballet 视频图像；与(d) Ballet 深度图像

Fig. 2　Motion vector distribution of video and depth in test sequences. (a) Interview video; (b) Interview depth; (c) Ballet video; (d) Ballet depth

上述实验表明，视频图像运动向量与深度图像运动向量并非完全相同。因此按照文献[11]的方法直接使用编码视频图像时得到的运动向量对深度图像进行运动补偿是不准确的，无法得到最优率失真性能。同理，文献[17]中联合估计视频图像与深度图像运动向量的方法为视频图像与深度图像指定相同的运动向量，也无法达到最佳性能。

4　视频-深度联合预测编码

依据第 3 节实验结果，视频图像的运动信息与深度图像的运动信息既具有统计相关性，也存在局部差异性。编码视频图像得到的运动信息接近真实的运动信息，具有整体性与规则性的特征。因此设计了两种运动信息重用机制，即视频-深度运动信息复制与视频-深度运动信息预测，在编码深度图像时充分利用已编码视频图像的运动信息。

视频－深度运动信息复制指在编码深度图像块时无需运动估计，而是直接使用对应视频图像中同位置块的运动信息进行运动补偿编码，如图 3 所示。这是一种新的宏块编码模式，需要一个标志位

motion_copy_flag 指示当前宏块是否使用了该模式。如果对应视频图像中同位置宏块由于采用帧内预测编码模式而没有运动信息，则不衡量该模式，也无需传输 *motion_copy_flag* 标志位。显然，在视频图像运动信息与深度图像运动信息完全相同的情况下，该模式能够节约深度图像中宏块分区方式与运动信息的编码开销，从而提高深度图像编码效率。

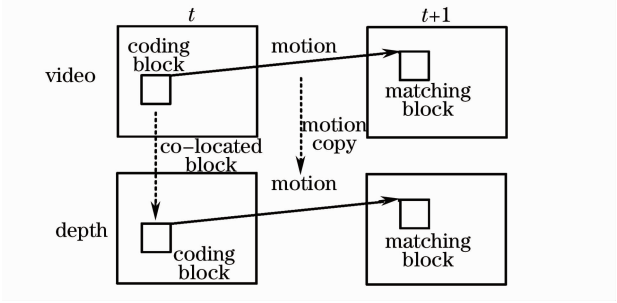

图 3 视频-深度运动信息复制

Fig. 3 Video-depth motion information copy

视频-深度运动信息预测指在编码深度图像中每一个宏块分区（macroblock partition），例如 16×16，16×8，8×16，8×8 等大小的宏块分区时，使用一个标志位 *motion_prediciton_flag* 指示该宏块分区是否使用对应视频图像中同位置、同大小图像块的运动信息作为自身运动信息的预测值，以及是否使用该预测运动信息对最终运动信息进行差分编码。如图 4 所示，若 *motion_prediciton_flag* 为 1，深度图像宏块分区则使用视频图像中对应块的参考图像索引 R_{video} 作为自身的参考图像索引 R_{depth}，同时使用视频图像中对应块的运动向量 mv_{video} 作为自身运动向量的预测值 p_{depth}。在这种情况下，深度图像中该宏块分区无需编码传输参考图像索引 R_{depth}，仅需要编码、传输估计得到运动向量 mv_{depth} 与 p_{depth} 的差值信号 d_{depth}。该过程可表示为

$$\left.\begin{aligned} R_{\text{depth}} &= R_{\text{video}} \\ p_{\text{depth}} &= mv_{\text{video}} \\ d_{\text{depth}} &= mv_{\text{depth}} - p_{\text{depth}} \end{aligned}\right\}, \tag{3}$$

若 *motion_prediciton_flag* 为 0，则按照普通运动估计算法确定深度图像宏块分区的最优参考图像索引与运动向量，而运动向量预测值也按照 H.264/AVC 规定的方法计算得到。当视频图像运动信息与深度图像运动信息相近但不完全相同时，该视频－深度运动信息预测机制能够在最大程度上利用视频图像的运动信息，从而提高深度图像压缩编码效率。

基于上面提出的两种运动信息重用机制，对深度图像编码过程进行率失真优化（Rate-Distortion Optimization，RDO）可用

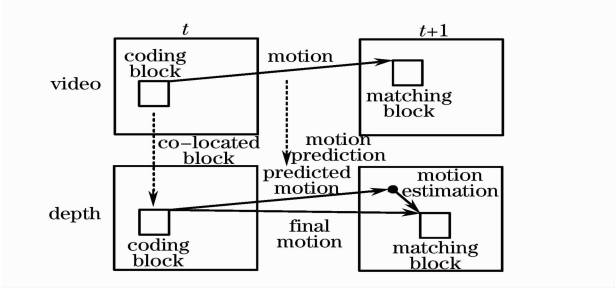

图 4 视频-深度运动信息预测

Fig. 4 Video-depth motion information prediction

$$\boldsymbol{I}^{*} = \underset{\boldsymbol{I}}{\operatorname{argmin}}\{J(\boldsymbol{S},\boldsymbol{I}\mid\lambda)\}, \tag{4}$$

式中 λ 为拉格朗日因子；$\boldsymbol{S}$ 为编码图像块数据；$\boldsymbol{I}$ 表示编码模式、运动信息等编码参数集合，其中包括标志位 *motion_copy_flag* 与 *motion_prediciton_flag*；$\boldsymbol{I}^{*}$ 为最小化拉格朗日代价函数 $J(\boldsymbol{S},\boldsymbol{I}\mid\lambda)$ 得到的最优编码参数集合。$J(\boldsymbol{S},\boldsymbol{I}\mid\lambda)$ 一般表示为

$$J(\boldsymbol{S},\boldsymbol{I}\mid\lambda) = D(\boldsymbol{S},\boldsymbol{I}) + \lambda R(\boldsymbol{S},\boldsymbol{I}), \tag{5}$$

式中 $D(\boldsymbol{S},\boldsymbol{I})$ 与 $R(\boldsymbol{S},\boldsymbol{I})$ 分别表示使用参数 $\boldsymbol{I}$ 编码 $\boldsymbol{S}$ 时的失真与速率。

5 实验结果与分析

考虑到与普通二维显示设备的兼容性以及压缩码流的可伸缩特性，提出的视频-深度联合预测编码方案采用与已有基于 H.264/AVC 的可伸缩视频编码（SVC）标准[19]相同的预测结构，将视频图像作为基本层独立编码，将深度图像作为增强层使用基本层运动信息进行预测编码。因此选择 SVC 参考软件 JSVM（Joint Scalable Video Model）作为实验平台，基于 JSVM 9.15 实现提出的编码方案。

文献[20]建议使用 ATTEST 序列 Interview 与 Orbi 衡量深度编码算法。这两个序列的深度图像由深度摄像机拍摄，并经空洞填充、错误数据校正、基于对象中值滤波等后处理操作得到。此外，还可以从普通摄像机拍摄得到的立体或多视点视频图像中，通过使用彩色分割匹配算法、图分割（Graph Cuts）算法、置信度传播（Belief Propagation）算法等各种先进的立体匹配算法计算得到深度数据。例如，微软研究中心交互式视频媒体研究组提供的 Breakdancers 与 Ballet 多视点视频序列及通过立体匹配计算得到的深度图像序列[18]。为全面衡量提出编码方案对这两种不同方法获得的深度图像的编码效率，实验对上述四个序列进行测试，序列参数在表 1 中给出。

表 1　测试序列参数

Table 1　Test sequence parameters

Sequence	Data format	Resolution	Frame rate /(frame /s)	Length
Interview	video+depth	720×576	25	100
Orbi	video+depth	720×576	25	100
Breakdancers	video+depth(view 3)	1024×768	15	100
Ballet	video+depth(view 3)	1024×768	15	100

首先编码视频图像序列得到运动信息。接着使用提出方案对深度图像序列进行预测编码，编码参数如表 2 所示。由于深度图像比较平滑，使用较大的量化参数就可以得到高质量的恢复图像[7,11]，因此实验选择大于 30 的量化参数进行编码。此外，MPEG 视频组正在研究深度图像质量对合成虚拟图像质量的影响[9]，其中编码深度图像的量化参数选择问题还有待进一步研究。因此这里选择 30～48 共 7 个量化参数进行测试，全面衡量提出编码方案性能。

表 2　编码参数

Table 2　Coding parameters

GOP size	12 (Interview, Orbi) 8 (Breakdancers, Ballet)
Inter-prediction structure	Hierarchical B pictures
Number of reference frames	1
Motion estimation search range	96 pixel
Entropy coding mode	CABAC
Rate-distortion optimization	High complexity mode
Quantization parameters	30, 33, 36, 39, 42, 45, 48

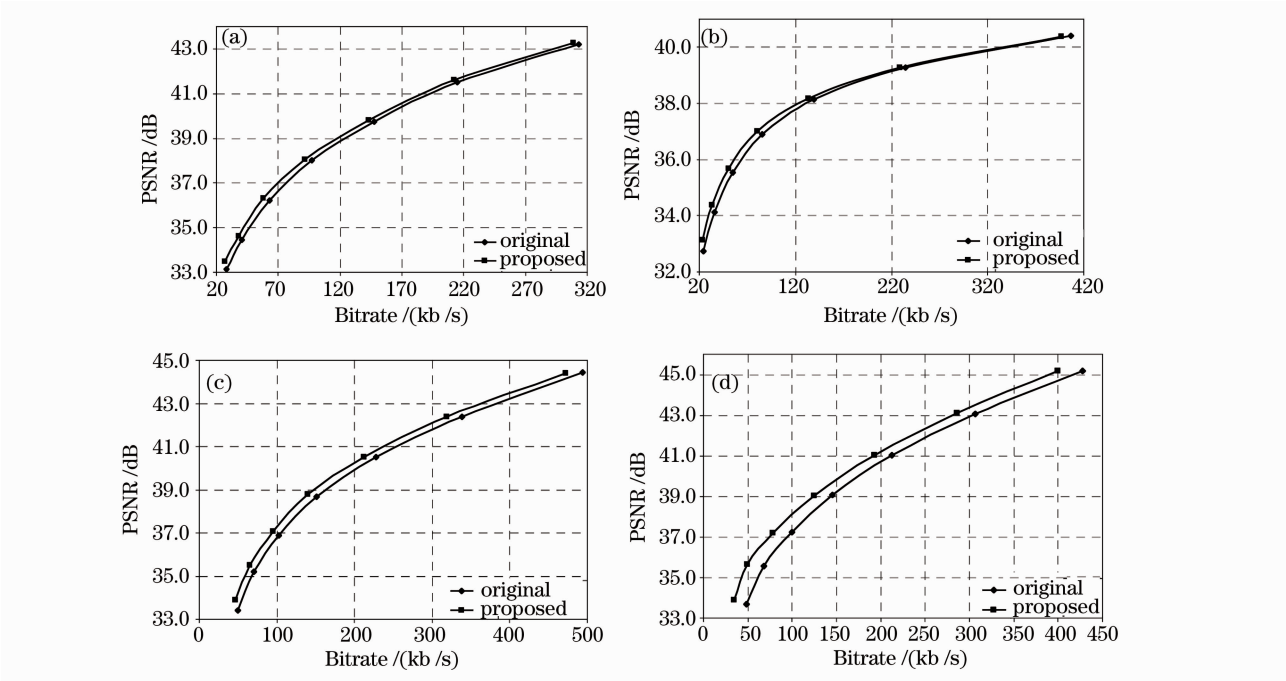

图 5　深度图像序列编码率失真性能曲线：(a) Interview；(b) Orbi；(c)Breakdancers；(d) Ballet

Fig. 5　Rate-distortion curves for depth image sequence of: (a) Interview; (b) Orbi; (c)Breakdancers; (d) Ballet

从图 5 中实验结果可以看出，与单独编码深度图像序列(Original)相比，提出的视频-深度联合预测编码方案(Proposed)能够提高编码效率。对于存在较大运动的序列，如 Breakdancers、Ballet，编码性能改善尤为明显。这是因为深度图像运动信息从视频图像预测得到，节省了用于编码大量运动信息的比特。这种编码增益在低码率时表现更加明显，例如 Ballet 序列在低码率时增益达到 2 dB。这是由于在低码率编码时运动信息在码流中所占比重增加，此时如果能降低运动信息编码开销，就可以提高整体率失真性能。此外，与文献[11]给出的仿真结果相比，提出算法不仅在低码率端获得显著增益，在高码率端也可以保持较高编码性能。

6　结　　论

重点研究基于 H.264/AVC 等视频编码标准的

深度图像压缩编码，提出的视频－深度联合预测编码方案包括视频－深度运动信息复制与视频-深度运动信息预测两种运动信息重用机制，可用于压缩编码单视点视频结合单视点深度、多视点视频结合多视点深度格式的三维场景数据。实验表明该方案能够高效利用已编码视频图像的运动信息，节约深度图像编码中的运动信息开销，显著提高深度图像编码效率。

参 考 文 献

1 A. Smolic, K. Mueller, N. Stefanoski *et al.*. Coding algorithms for 3DTV-a survey[J]. *IEEE Trans. Circuits and Systems for Video Technology*, 2007, **17**(11): 1606～1620

2 A. Redert, M. O. de Beeck, C. Fehn *et al.*. Advanced three-dimensional television system technologies[C]. *Proceedings of First International Symposium on 3D Data Processing Visualization and Transmission*, 2002. 313～319

3 M. Tanimoto. Overview of free viewpoint television[J]. *Signal Processing: Image Communication*, 2006, **21**(6): 454～461

4 Yang Qingguo, Liu Liren, Lang Haitao. Range estimation by optical differentiation[J]. *Acta Optica Sinica*, 2005, **25**(9): 1186～1190
阳庆国，刘立人，郎海涛. 图像深度估计的光学微分方法[J]. 光学学报，2005，**25**(9)：1186～1190

5 Ding Yabin, Peng Xiang, Tian Jindong *et al.*. Pose estimation of multiple viewpoints for three-dimensional digital imaging system [J]. *Acta Optica Sinica*, 2007, **27**(3): 451～456
丁雅斌，彭 翔，田劲东 等. 一种三维数字成像系统的多视点姿态估计方法[J]. 光学学报，2007，**27**(3)：451～456

6 Yang Haitao, Chang Yilin, Huo Junyan *et al.*. Depth characteristic-based image region partition and regional disparity estimation for multi-view video coding[J]. *Acta Optica Sinica*, 2008, **28**(6): 1073～1078
杨海涛，常义林，霍俊彦 等. 应用于多视点视频编码的基于深度特征的图像区域分割与区域视差估计[J]. 光学学报，2008，**28**(6)：1073～1078

7 K. Schüür, C. Fehn, P. Kauff *et al.*. About the impact of disparity coding on novel view synthesis [R]. ISO/IEC JTC1/SC29/WG11, Doc. M8676, Jul. 2002

8 C. Fehn. Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV[C]. *Proc SPIE*, 2004, **5291**: 93～104

9 ISO/IEC JTC1/SC29/WG11. Description of Exploration Experiments in 3D Video Coding[R]. ISO/IEC JTC1/SC29/WG11, Doc. N10173, Oct. 2008

10 P. Merkle, A. Smolic, K. Muller *et al.*. Multi-view video plus depth representation and coding [C]. *IEEE International Conference on Image Processing*, 2007, ***I***: 201～204

11 H. Oh, Y.-S. Ho. H. 264-based depth map sequence coding using motion information of corresponding texture video [J]. *Lecture Notes in Computer Science*, 2006, **4319**: 898～907

12 R. He, M. Yu, G. Jiang. A depth image coding method for 3DTV system based on edge enhancement [C]. 11*th IEEE International Conference on Communication Technology*, 2008. 665～668

13 B.-B. Chai, S. Sethuraman, H. S. Sawhney *et al.*. Depth map compression for real-time view-based rendering [J]. *Pattern Recognition Letters*, 2004, **25**(7): 755～766

14 S.-Y. Kim, Y.-S. Ho. Mesh-Based Depth Coding for 3D Video using Hierarchical Decomposition of Depth Maps [C]. *IEEE International Conference on Image Processing*, 2007, ***V***: 117～120

15 I. Daribo, C. Tillier, B. Pesquet-Popescu. Adaptive wavelet coding of the depth map for stereoscopic view synthesis [C]. *IEEE* 10*th Workshop on Multimedia Signal Processing*, 2008. 413～417

16 C. Fehn, K. Schüür, P. Kauff *et al.*. Coding results for EE4 in MPEG 3DAV[R]. ISO/IEC JTC1/SC29/WG11, Doc. M9561, Mar. 2003

17 Z. Chen, G. Li, Y. He. Experiment evaluation about motion compensation of MAC for stereoscopic video coding[R]. ISO/IEC JTC1/SC29/WG11, Doc. M9128, Dec. 2002

18 C. L. Zitnick, S. B. Kang, M. Uyttendaele *et al.*. High-quality video view interpolation using a layered representation[J]. *ACM Transactions on Graphics*, 2004, **23**(3): 600～608

19 H. Schwarz, D. Marpe, T. Wiegand. Overview of the scalable video coding extension of the H. 264/AVC standard[J]. *IEEE Trans. Circuits and Systems for Video Technology*, 2007, **17**(9): 1103～1120

20 C. Fehn, K. Schüür, I Feldmann *et al.*. Distribution of ATTEST test sequences for EE4 in MPEG 3DAV[R]. ISO/IEC JTC1/SC29/WG11, Doc. M9219, Dec. 2002