

2003年上海大学博士学位论文 25

立体视频帧估计与内插技术 及立体视频编解码系统研究

作者：骆艳

专业：通信与信息系统

导师：张兆扬



上海大学出版社

· 上海 ·

2003 年上海大学博士学位论文

立体视频帧估计与内插技术 及立体视频编解码系统研究

作 者： 骆 艳

专 业： 通信与信息系统

导 师： 张兆扬

上海大学出版社

· 上海 ·

7N911.2

Shanghai University Doctoral Dissertation (2003)

**Stereo Video Frame Estimation and
Interpolation and Stereo Video
Coding/Decoding System**

Candidate: Luo Yan

Major: Communication and Information System

Supervisor: Prof. Zhang Zhaoyang

Shanghai University Press

• Shanghai •

上海大学

本论文经答辩委员会全体委员审查，确认符合上海大学博士学位论文质量要求。

答辩委员会名单：

- | | | |
|---------------|-------------------|--------|
| 主任：戚飞虎 | 教授，上海交大计算机系 | 200030 |
| 委员：张立明 | 教授，上海交大电子工程系 | 200092 |
| 张一均 | 教授级高工，上海 SVA 集团公司 | 200052 |
| 余松煜 | 教授，上海交大图象通信所 | 200030 |
| 王朔中 | 教授，上海大学电子信息工程系 | 200072 |
| 导师：张兆扬 | 教授，上海大学通信与信息学院 | 200072 |

评阅人名单:

- | | | |
|------------|---------------|--------|
| 戚飞虎 | 教授, 上海交大计算机系 | 200030 |
| 陈咏恩 | 教授, 同济大学电气工程系 | 200072 |
| 王国中 | 教授, 上广电中央研究院 | 200062 |

评议人名单:

- | | | |
|------------|--------------------|--------|
| 余松煜 | 教授, 上海交大图象通信所 | 200030 |
| 方 勇 | 教授, 上海大学电子信息工程系 | 200072 |
| 王治钢 | 研究员, 上海航天局 809 研究所 | 200031 |
| 周源华 | 教授, 上海交大电子工程系 | 200030 |

答辩委员会对论文的评语

骆艳同学的博士学位论文《立体视频帧估计与内插技术及立体视频编解码系统研究》是国家自然科学基金(69972027)及(60202015)的资助项目中的重要内容之一。该文从提高立体视频帧估计与内插的图象质量、重建速度和编码效率的角度作了深入研究,取得了较好的研究成果。对促进立体视频的应用有一定作用。论文重要创新点如下:

提出了3种立体视频帧估计与内插算法,可以分别针对不同的应用场合。其中基于不规则四叉树运动估计和分割(IMRQT)的帧估计算法只需要少量的参考图象即可恢复图象;而CMDR方法以及改进的CMDR方法无需在解码端作运动和视差匹配搜索即可恢复图象。这3种算法在恢复图象的质量以及重建速度上均比现有的算法有着显著提高。

在上述基础上提出了一种立体视频编解码系统。其中对辅视频流只需对少量的参考帧进行基于视差和运动估计的预测编码,而大多数的帧不需编码传输,就可在解码端用帧估计与内插技术进行恢复。该系统可对立体视数据流进行高效压缩,在压缩性能上具有领先性。

提出了一种利用图象区域分割技术进行中间视生成的算法。该算法对视差场的精确性要求不高,计算量很低,生成的中间视图象具有可接受的视觉效果,在实时性要求高的三维视频会议等系统中可得到应用。

论文条理清晰、理论分析深入，提出的算法新颖，具有创造性，实验数据充分，表明作者具有坚实、宽广的专业理论基础，独立科研工作能力强。在答辩过程中，陈述清晰，回答问题正确。

答辩委员会表决结果

经答辩委员会表决，全票同意通过骆艳同学的博士学位论文答辩，建议授予工学博士学位。

答辩委员会主席：**戚飞虎**

2003年7月3日

摘 要

近年来,随着计算机和显示技术的进步,立体数字电视将会逐步取代 HDTV,成为下一代的数字视频主流和未来多媒体通信中的一个重要组成部分.但是传输或存储立体视频的数据量将是单通道视频的两倍以上.因此,在带宽有限的信道中,如何利用双目流之间的相关性以整体地提高两路视频信号的编码效率,就成为当前立体数字电视的重要研究方向.

由于增加了另一个流的图像帧,使得立体图像序列中增加了大量的冗余度.根据这样的事实以及人眼的立体视觉掩蔽特性,允许立体视频流中的许多帧无需进行编码传输而直接在解码端进行恢复.这样可提高立体视频的编码速度和压缩率,降低用于传送立体视频流的附加带宽.因此本文从提高立体视频帧估计与内插的图像质量以及重建速度,以及提高立体视频的编码效率的角度展开深入研究,相应地做了一些改进和创新的工作.

本文首先对影响帧估计和内插的各种因素加以分析,对当前的立体视频帧估计和内插算法——基于固定块匹配的 FSBM 算法中存在的各种问题进行了实验和讨论,指出了提高重建帧估计精度的关键在于解决重建帧中运动区域的运动匹配以及运动遮挡问题,而提高重建速度的关键在于提高运动估计的速度.

基于以上分析,提出了一种基于多分辨率不规则四叉树运动估计和分割方法(简称 IMRQT 法).在运动估计的同时进行分割,在分割时综合了运动和图像亮度信息,避免了块运动估计

中存在的块状可视干扰,保留了运动细节.在进行运动估计和分割同时采用多分辨率结构以降低计算负担.在获得 IMRQT 的运动估计和分割结果的基础上,又提出了新的立体视频帧估计方法,将被重建的图像分为运动匹配区域和运动失匹配区域,给出了对于这两类区域进行预测和填充的方法.基于 IMRQT 的立体视频帧估计和内插方法比 FSBM 方法帧重建的速度提高了约 20 倍,图像信噪比提高 1.27~3dB,并具有较好的估计精度和主观观感.由于仅需要 3 个参考图像即可恢复内插帧,方法简单易行.

由于以上两种方法均需要在解码端重做运动和视差估计,因此在提高解码速度上有着一定的局限性.为此本文利用立体透视投影几何模型,推导出立体视频中邻接帧之间视差矢量场与运动矢量场之间的对应关系,并将这种关系应用于立体视频的帧估计之中,提出基于运动/视差方程及最小代价方程的立体视频帧估计与内插方法(简称 MDR-LCF 方法):对于非视差遮挡区域中的像素点,利用视差和运动矢量之间的对应关系获得该内插点的映射位置和映射强度;对于视差遮挡区域中的像素点,可根据视差遮挡的特性及其约束条件,定义一个合适的置信度检验方法构造代价函数,利用其邻域中非视差遮挡区域像素点所对应的运动矢量来估计其对应的运动矢量并进行映射,以得到该像素点的位置和强度估计值,这样避免了解码端重作运动/视差搜索估计,同时可以获得较好的重建图像质量.

为了提高帧估计的鲁棒性,在实际进行立体视频帧估计和内插时,利用 MDR-LCF 方法分别对已获得的运动和视差矢量场以及前后两个参考图像帧组合进行帧估计,然后再对估计结果进行综合,本文称为基于混合 MDR-LCF(简称 CMDR-LCF)

的帧估计方法。

由于使用 CMDR-LCF 方法的一个前提条件是在视差矢量场中显式标记出视差遮挡区域和非遮挡区域并对它们分别进行处理,因此本文针对立体视频编码时用基于块的视差估计算法得到视差场后进行帧估计的情况,又提出了无遮挡检测的帧估计方法——改进 CMDR-LCF 方法。

实验结果表明,对于重建帧的 PSNR, CMDR-LCF 方法比基于 FSBM 方法提高了 2.5dB 左右,而改进 CMDR-LCF 方法则提高了将近 3dB 左右;由于直接使用解码端获得的运动和视差矢量,节省了大量的计算量,使得这两种方法的帧估计速度比 FSBM 方法提高了约 60 倍。

结合本文的前期研究结果,本文还提出了与单视编码序列标准相兼容的高效立体视频编解码系统。该系统对其中的辅视流,仅对少数的参考帧进行预测编码和传送,其余帧则完全“跳过”,而在解码端通过帧估计和内插进行恢复。实验结果表明,本文提出的立体数字视频系统的总比特流仅是通常单视数字视频传输比特流的 1.15~1.3 倍左右,且具有良好的 3D 立体视觉效果,在压缩性能上具有领先性。

由于立体图像与单视图像有质的区别,仅用 PSNR 测试重建立体图像的质量是不充分的,为此本文提出了一种测试重建立体图像对的视差相似性检测的评价方法。该方法通过比较采用重建立体图像对与采用原始图像对所获得的视差场之间的差异程度来反映帧估计算法的性能。可用来衡量各种立体视频帧估计算法的优劣。

作为立体视频帧估计的扩展,本文研究了图像分割技术在中间视生成和内插中的应用,提出一种与区域分割技术相结合

的视差场校正技术,对含有大量误匹配的视差场进行校正,不仅可以对整个视差场进行平滑,还可以在检测出具有较大误差匹配的视差矢量值的同时对错误视差矢量进行校正,并具有简单易行的特点.结合视差场校正技术,提出了一种快速合成中间视图像的方法.对视差场的精确性要求不高,计算量很低,生成的中间视图像具有可接受的视觉效果,在实时性要求高的三维视频会议等系统中可得到应用.

关键词 立体视频编码, 帧估计和内插, 四叉树分割, 中间视内插

Abstract

With the advance of computer and display technologies of today, stereo digital video will take the place of HDTV and become the main stream of next digital video generation and one important content in multimedia communication in soon future. However, the data needed to be transmitted and stored of stereo video is multi-fold more than that of monoscopic video. In order to transmit stereo video signal in limited bandwidth, taking advantages of co-relationship between the stereo streams to improve the coding efficiencies becomes the major stereo video research direction of today .

Much more redundancies are added since another video stream is added. With this fact and the stereo mask characters of human visual system, it is possible that many frames in the stereo video sequences are neither coded nor transmitted at the coder and are reconstructed at the decoder directly. This will greatly improve the coding speed and compression rate of stereo video, and decrease the additional bandwidth needed to transmit the auxiliary stream. In this dissertation, we study the frame estimation and interpolation algorithms in the aim of increase the image distortion performance and reconstruction speed, we also study how to improve coding effeciency of stereo video, some improvements and creativeness are made correspondly.

We firstly analyze various factors that affect the frame estimation and interpolation performance, experimental results and discussions given to the conventional algorithm—Fixed Size Block Matching based algorithm (FSBM) indicate that, the key of improving estimation

precision is to solve the motion matching and occlusion issues in the moving areas, the key of improving reconstruction speed is to improve the motion estimation speed. Accordingly, a computationally efficient irregular multi-resolution quadtree decomposition based motion estimation and segmentation scheme is developed (IMRQT), this technique do segmentation jointly with motion estimation, which avoids the block disturbing greatly, and the motion details are maintained as well. Multi-resolution struts are used in the processing to reduce the computation burden significantly. Based on the results of IMRQT, a novel stereo frame estimation and interpolation method that treat the reconstructed images as motion match areas and motion mismatch areas distinguishly is proposed, different prediction and filling methods are given to these two kinds of areas accordingly. Compared with FSBM algorithm, about 20 folds of reconstructing speeds are improved and the PSNR is about 1.27dB~3dB higher, t what's more, better estimation precision and finer visual perceptions are achieved using our novel algorithm.

Since the above two algorithms need do motion and disparity estimation at the decoder, the improvement of frame reconstruction speed is limited. Thus in this dissertation, according to the stereo perspective projection geometry, the mapping relationship between disparity vector fields and motion vector fields of neighbor frames in a stereo video streams is deduced. This relationship is used in the frame estimation techniques and a nervs algorithm which is called stereo frame estimation and interpolation based motion/disparity function and the least cost function (MDR_LCF) is proposed, in which the mapping location and intensity estimations of those pixels in the un-disparity-occluded areas can be acquired with the relationship of motion and disparity vectors; the others are seemed to be in the disparity occluded

fields, and a cost function is constructed with a suitable reliable detection method associate with the features and constrain conditions in the disparity occluded fields, the corresponding motion vectors of pixels in occluded areas are estimated with those vectors acquired in the neighbor unoccluded areas, after mapping with selected motion vectors, intensity estimation of disparity occluded pixels are achieved. This algorithm can avoid motion and disparity estimation at the decoder and can acquire good image quality.

In order to improved the robust of frame estimation, and make use of those coded and transmitted motion and disparity vector fields associated with reference frames at the decoder, another technique called combined MDR_LCF (CMDR_LCF) algorithm is proposed.

One precondition of utilizing CMDR_LCF is it is necessary to label the disparity occluded and unoccluded areas premeditatedly and processes them distinguishshly. So as to those disparity fields achieved with algorithms based block, we propose a stereo frame estimation algorithm without occlusion detection – advanced CMDR_LCF algorithm.

Experimental results show that, as to the PSNR results of reconstructed images, compared with the FSBM algorithm, 2.5dB is improved using the CMDR_LCF algorithm; 3dB is improved using the advanced CMDR_LCF algorithm. Moreover, since the computation of motion and disparity can be avoided taking use of motion and disparity vectors at the decoder, reconstruction speeds can be improved about 60 folds compared with FSBM algorithm.

Cooperating with the previous works of our research projection, a high efficiency stereo video coding/decoding system compatible with the monoscopic coding standard is constructed. In which one video stream is coded in a higher image quality and is called as main stream; the other

video stream is called auxiliary stream and only a few frames of it are coded and transmitted, the others are "skipped" at the coder, and reconstructed at the decoder with a stereo frame estimation and interpolation algorithms. Experimental results show that, the stream data needed to be transmitted in the stereo digital video is about 1.5~1.3 folds over that of monocular digital video and have favorable 3D visual perception..

The essential deference between the stereo and monocular images makes the evaluation of a stereo image pair using only PSNR be insufficient. Thus a disparity similarity detection evaluation between reconstructed stereo image pairs and original pairs is proposed in this dissertation. Performances of different stereo frame estimation and interpolation algorithms are reflected through the comparison of similarity degree between different disparity fields associate with reconstructed and original stereo pairs.

As an expanding of stereo frame estimation and interpolation, the mid-view image creation and interpolation combined with image segmentation techniques is studied in this dissertation. A disparity calibration jointed with area segmentation technique is also proposed, which can emend disparity fields containing a lot of error matching. This technique can not only smooth the total disparity fields, but also detect the disparity vectors with larger errors and do emendation with them. Jointed with this disparity calibration technique, a fast mid-view image creation and interpolation algorithm intruduced, this technique can synthesize mid-view images rapidly, and has low computation complexon and favorable interpolation visual efficiency.

Key words stereo video coding, frame estimation and interpolation, quatra-tree segmentation, mid-view interpolation

目 录

第一章 绪 论	1
1.1 问题背景	1
1.2 研究现状及存在的问题	4
1.3 课题研究的目的和内容	13
1.4 论文的主要贡献及结构	15
1.5 本章小结	19
第二章 3D 立体视频分析	20
2.1 三维与立体	20
2.2 立体成像几何原理及视差	25
2.3 视差特性分析	31
2.4 立体图像的视差估计问题	36
2.5 数字图像压缩与立体图像/视频压缩编码	38
2.6 立体图像/视频编码	43
2.7 性能测试及评价	49
2.8 本章小结	50
第三章 基于不规则二叉树运动估计和分割的立体 视频帧估计和内插	51
3.1 单视序列的帧估计和内插	51
3.2 立体视频帧估计和内插的方法及研究现状	57
3.3 基于多分辨率不规则二叉树分割的运动 估计与分割	66
3.4 基于 IMRQT 的立体视频帧估计和内插算法	77
3.5 实验与讨论	81

3.6 本章小结	84
第四章 基于立体视频运动/视差场关系的	
帧估计和内插算法	89
4.1 立体视频中视差矢量与运动矢量的关系	90
4.2 利用运动和视差场的关系进行帧估计和内插	97
4.3 混合 MDR-LCF 方法及对右 B 帧进行估计和内插	108
4.4 无遮挡检测的帧估计方法——改进的 CMDR-LCF 方法	113
4.5 实验结果及分析	117
4.6 本章小结	126
第五章 基于帧估计和内插的立体视频编解码系统	128
5.1 系统特点说明	128
5.2 基于帧估计和内插的立体视频编解码系统总框架	130
5.3 实验结果及分析	136
5.5 本章小结	140
第六章 中间视图像生成与内插	142
6.1 图像分割技术——“分-合”区域增长	144
6.2 视差场误差检测与校正技术	148
6.3 中间视生成技术	151
6.4 实验结果及分析	155
6.5 本章小结	158
第七章 总结和展望	159
7.1 本文工作总结	159
7.2 进一步的工作	161
参考文献	163
附 录	175
致 谢	178