

基于运动估计的深度信息生成技术研究

姜太平, 王 帅, 占 涛

(安徽工业大学 计算机学院, 安徽 马鞍山 243002)

摘 要:国内外的多家电视媒体和拥有立体显示技术的企业都进行过开播数字立体频道的尝试,立体电视(3DTV)是未来的一个趋势。而立体视频片源制作费用高,周期长,是目前亟需解决的问题。描述了一种3DTV解决方案,提出了一个平面视频转换立体视频的方法。并主要针对方法中的深度信息生成问题,结合运动估计算法,解决了小范围内非关键帧深度信息自动生成问题,简化了转化过程中的人为工作量,提高了效率,降低了制作成本。证明了单目平面视频转换为基于深度信息的立体视频的可行性,得到较为满意的结果。

关键词:立体电视;立体视频;深度信息;运动估计

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2010)10-0094-04

Generation Technique Research of Depth Information Based on Motion Estimation

JIANG Tai-ping, WANG Shuai, ZHAN Tao

(School of Computer Science, Anhui University of Technology, Maanshan 243002, China)

Abstract: Many TV media and companies having a three-dimensional display technology have been trying to launch digital stereo channels, 3DTV is a trend of future. However, the high production costs and cycle length of the source of 3D video, is a urgent need to resolve. Described a solution for 3DTV, proposed a method of converting plane video to 3D video. For the generation of depth information in the method, solved a problem of automatically generating the depth information of non-key frame combined with motion estimation algorithm. Simplified the human workload in conversion process, improved efficiency and reduced production costs. Proved the feasibility of conversion from 2D to 3D video and received a more satisfactory results.

Key words: 3DTV; 3D video; depth information; motion estimation

0 引 言

黑白电视机到彩色电视机是电视机的第一次飞跃,标清电视到高清电视是电视机的第二次飞跃,那么高清晰电视机到立体电视(3DTV)将是电视机的第三次飞跃^[1]。目前,3DTV已越来越受到人们重视,图1为一套3DTV系统方案,包括3D内容制作、3D编码、传输、3D解码和终端显示五个部分。在现有的数字广播网进行实施完成后,未来人们在家里就可利用裸眼立体显示器或者经改造后的家用电视(比如佩戴专业眼镜设备等)等多种模式观看到立体电视节目,而且节目源可以在压缩之后经由电视机顶盒接收,或者制作成专门的DVD格式播放,可解决现有节目源匮乏、不

稳定的问题。

图1中的3D内容制作方法,主要有三种:

一种是通过立体拍摄方法直接获取3D内容,即通过双(多)镜头对场景进行拍摄,并通过后期处理获得立体视频内容。

一种是通过动画制作模型,结合动画脚本,实现立体场景的多通道渲染,最终获得立体视频内容。

一种是将平面视频转换为立体视频,通过对平面视频进行识别和分析,获得一定的立体信息。

由于前两种方法获取立体视频内容的成本高,尤其是现有片源偏少,因此如何有效利用现有的大量平面视频,并通过一定的处理转换成立体视频的技术越来越受重视。不同的立体显示设备有不同的立体视频格式,由于“二维视频+深度图”这种立体视频格式十分有利于传输和储存,又能最大限度兼容现在的平面视频,在未来有可能成为3DTV的一种标准。该格式最核心的就是深度信息的生成,如何快速获得有效的生成深度信息成为一个技术瓶颈。文中将通过运动估

收稿日期:2010-02-03;修回日期:2010-05-16

基金项目:国家863项目(2009AA01Z328);校青年教师基金项目(QZ200819)

作者简介:姜太平(1974-),男,博士,副教授,研究方向为立体显示、立体图像处理。

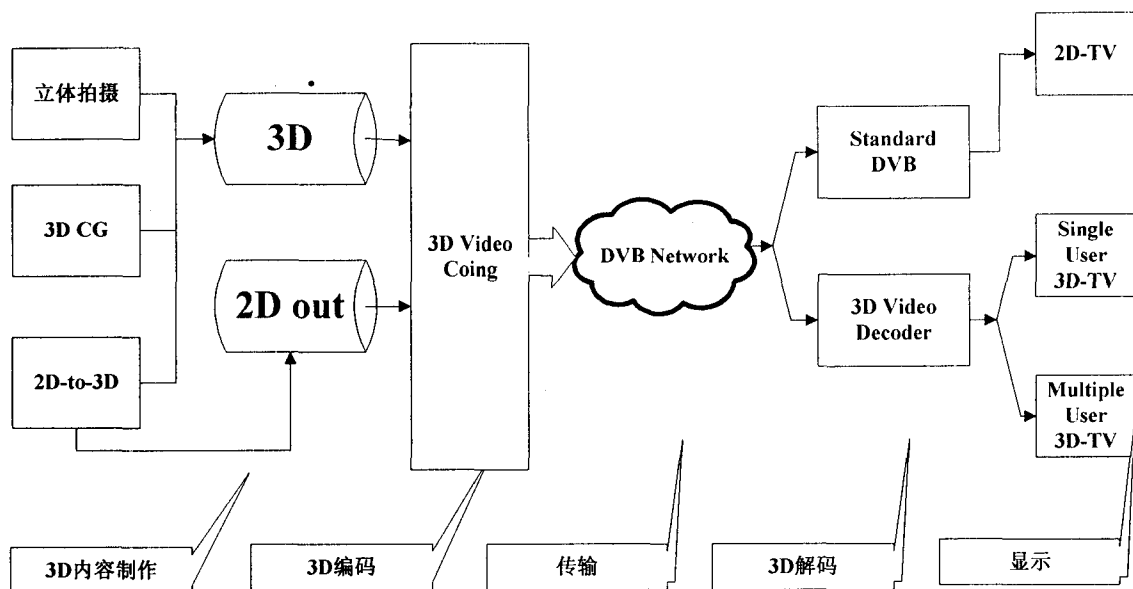


图 1 3DTV 系统方案图

设计的算法来实现平面视频的深度信息生成,降低技术难度,加快转换过程。

1 平面视频转换立体视频方案

1.1 深度图

人眼能看到立体效果是因为看到的左视图和右视图有视差,Liang Zhang 和 Wa James Tam 提出了一种通过平面图像和视差算出左右视图的简单方法^[2]。如图 2 所示: P 是空间中任一点, Z 代表 P 点相对摄像机的深度, f 是焦距, C_l 和 C_r 分别是左摄像机和右摄像机。 t_x 是摄像机之间的距离。 P 投影到成像平面上的点分别为 (x_l, y) , (x_c, y) 和 (x_r, y) 。由图 2 中几何关系,可以得到以下公式:

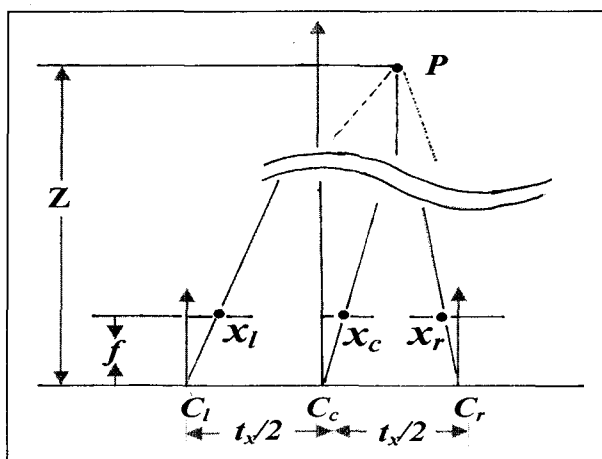


图 2 立体图像对生成

$$x_l = x_c + (t_x/2) \times (f/Z) \quad (1)$$

$$x_r = x_c - (t_x/2) \times (f/Z) \quad (2)$$

由(1)和(2)得:

$$d = x_l - x_r = t_x \times (f/Z) \quad (3)$$

其中 d 是视差, f 是焦距, 根据自己需要调整(3)中 t_x 的值, 可以在已知左视图或右视图的情况下, 通过相对深度信息, 计算像素偏移从而得另外一张视图。

深度图就是基于视差与二维视频可以生成双目视觉的概念上提出的。深度图采用 8 位灰度编码(256 级), 分辨率和对应的二维图像相同, 像素与二维图像一一对应, 它包含了对应二维图像的视差信息, 其中 0 对应屏幕后面视差最大的物体, 255 对应最靠近观察者的景物^[3]。图 4(a) Hamburger 右半边为左半边的深度图, 中间汉堡部分显示时要有突出效果, 最靠近观察者, 深度赋值为 255, 看起来最亮, 薯条和其他景物离观察者较远, 深度值递减, 看起来比较暗。

1.2 转换方案

平面视频转立体视频就是设法获得原平面视频每帧的深度信息从而得到深度序列, 可见深度信息的生成是平面视频转立体视频的关键。

目前深度信息的提取方法有两种^[4,5]: 一种是基于计算机算法自动提取, 如四元 SOBEL 滤波、四元 GABOR 小波运动估计或线性透视方法^[6]。另一种是利用视频或图像编辑软件, 对平面视频每一帧进行前景物体分割, 根据视频序列内容(前后景物位置和运动关系), 进行深度赋值, 获得平面视频对应的深度序列。前者虽然高效, 不过目前没有成熟产品问世, 而后者在实践过程中, 逐帧提取效率太低, 针对这一问题, 提出利用运动估计算法自动生成非关键帧深度图。具体方案流程如下:

1) 人工按照平面视频内容划分内容相关子序列。

2) 计算机自动将子序列视频解码为一帧一帧的图像。人工确定关键帧和非关键帧, 前景物体分割, 生成关键帧的深度图。

3) 计算机利用运动估计算法自动生成相关非关键帧的深度图。

4) 重复 2 和 3, 直到生成该子序列中所有非关键帧的深度图。

5) 重复 1~4, 直到生成原平面视频所有帧对应的深度图。

6) 计算机自动根据特定立体显示设备要求, 将原有帧和深度图合成特定立体视频帧。

7) 计算机自动将立体视频帧成立体视频序列。

8) 在特定立体显示设备上实现播放。

1.3 深度信息生成

运动估计可以根据前一帧数据产生当前帧。关键帧的深度图是根据二维图像人工生成的, 结合对二维图像序列运动估计得到的运动矢量, 可以自动生成非关键帧的深度图。

运动估计算法常用的有全搜索法(FS)、二维对数搜索法(TDL)、三步搜索法(TSS)、钻石搜索法(DFS)^[7-9]。常用的匹配准则有最小绝对误差 MAD(如公式(4)), 最小均方误差 MSE(如公式(5))和归一化互相关函数 NCCF(如公式(6))^[10,11], 由于 MAD 没有乘除操作, 不需要做乘法运算, 故实现起来方便, 使用较多。实践中通常使用求和绝对误差 SAD(如公式(7))代替 MAD。其中 M, N 指宏块大小为 $M \times N$ 像素。

$$MAD(i, j) = \frac{1}{MN} \sum_{m=0}^M \sum_{n=0}^N |f_k(m, n) - f_{k-1}(m + i, n + j)| \quad (4)$$

$$MSE(i, j) = \frac{1}{MN} \sum_{m=0}^M \sum_{n=0}^N [f_k(m, n) - f_{k-1}(m + i, n + j)]^2 \quad (5)$$

$$NCCF(i, j) = \frac{\sum_{m=0}^M \sum_{n=0}^N f_k(m, n) f_{k-1}(m + i, n + j)}{[\sum_{m=0}^M \sum_{n=0}^N f_k^2(m, n)]^{\frac{1}{2}} [\sum_{m=0}^M \sum_{n=0}^N f_{k-1}^2(m, n)]^{\frac{1}{2}}} \quad (6)$$

$$SAD(i, j) = \sum_{m=0}^M \sum_{n=0}^N |f_k(m, n) - f_{k-1}(m + i, n + j)| \quad (7)$$

四种运动估计算法在 Pentium IV 计算机, TMN8 模型, VC++ 6.0 平台下, 对 QCIF 格式 Claire 序列前 100 帧比较结果^[12]如表 1 所示。

考虑速度和质量, 在人工生成关键帧深度图后, 采取 DFS 生成非关键帧深度图, 匹配准则为 SAD。

具体算法步骤如下:

表 1 四种搜索的平均搜索次数和 PSNR 比较

算法	平均搜索次数	平均 PSNR
FS	289	35.80
TDL	19	35.72
TSS	25	35.73
DFS	18	35.72

步骤 1: 读取关键帧的深度图和其对应的二维图像序列。

步骤 2: 获取当前帧(N)和下一帧($N+1$)二维图像的亮度分量 Y_n 和 Y_{n+1} , 把 Y_{n+1} 分成 16×16 像素的宏块。跳至步骤 3。

步骤 3: 拿 Y_{n+1} 中未获取运动矢量的宏块, 在 Y_n 中按大菱形搜索模式(LDSP, 见图 3(a)), 在一定的搜索范围内搜索, 将 LDSP 中心放在搜索窗口的原点, 比较 LDSP 九个点的 SAD, 判断 LDSP 中心点的 SAD 是否最小, 是则跳至步骤 5, 否则跳至步骤 4。

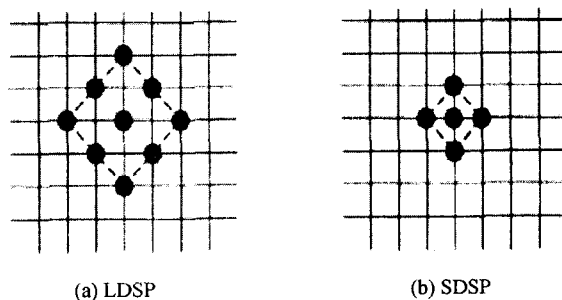


图 3 两种菱形搜索模式

步骤 4: 以步骤 3 中 SAD 最小点为中心形成新的 LDSP, 每次仅再计算三或五个点的 SAD。判断 LDSP 中心的 SAD 是否最小, 是则跳至步骤 5, 否则重复执行步骤 4。

步骤 5: 把 LDSP 换为小菱形搜索模式(SDSP, 见图 3(b)), 五个点中 SAD 最小的那个点即为匹配块起点。计算并保存当前宏块与匹配宏块的运动矢量。判断获取 Y_{n+1} 中宏块的运动矢量是否结束。是则跳至步骤 6, 否则跳至步骤 3。

步骤 6: 根据运动矢量和 N 所对应的深度图, 生成下一帧深度图的每个宏块, 继而生成下一帧深度图。判断是否继续生成下一帧深度图, 是则跳至步骤 2, 否则跳至步骤 7。

步骤 7: 结束。

当深度图生成后, 需要和二维图像重构为特定的立体视频帧格式, 即“二维视频 + 深度图”(见图 4(a)), 所有立体视频帧组成立体视频序列。在特定的立体显示器上播放时, 二维视频作为左视图, 左视图和深度图合成得到右视图, 使人左右眼分别看到左右视

图,左右视图具有视差,则产生立体效果。

2 实验与结果分析

在 Pentium Dual - Core 计算机下生成深度图,通过创图视维公司的 40 寸立体显示器观察显示结果。选择两种视频,深度图轮廓细节多、运动强度大的 Hamburger 和轮廓细节少、运动强度小的 Flower。人工生成第一张深度图,利用运动估计算法自动生成非关键帧深度图。

实验结果如图 4 和图 5 所示。

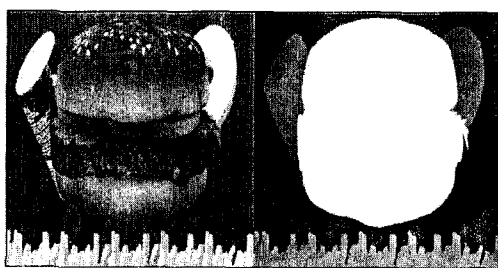
通过实验可以看到,运动估计算法在深度图轮廓细节少、运动强度小的情况下能大量生成深度图,在轮廓细节多、运动强度大的情况下,只能保证 10~20 帧深度图的精确生成。不过应用于实践,在人机交互的

平面视频转换立体视频系统中,已能降低 3~5 倍的人为工作量,提高了效率,降低了制作成本。

3 结束语

文中描述了一种 3DTV 方案,提出了一个平面视频转换立体视频的方法,证明了基于深度图的平面视频转换立体视频的可行性。针对逐帧提取深度图制作成本高、效率低的问题,结合运动估计算法,在一定程度上解决了非关键帧深度图的自动生成问题,具有实用价值,降低了制作成本。

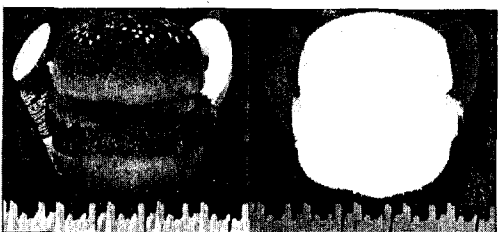
目前,在基于深度图的转换过程中,还有以下问题值得研究:如何自动基于内容的视频场景划分,如何精确的图像边缘分割生成深度图,图像遮挡问题也是进一步研究的重点。



(a) Hamburger 第 1 帧



(b) Hamburger 第 10 帧非关键帧深度图



(c) Hamburger 第 30 帧非关键帧深度图

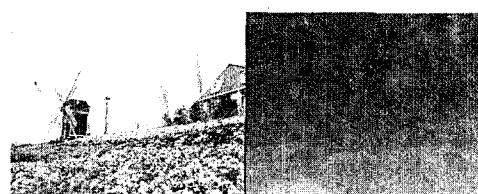


(d) Hamburger 第 90 帧非关键帧深度图

图 4 Hamburger 序列二维图像和对应的生成深度图



(a) Flower 第 1 帧



(b) Flower 第 10 帧非关键帧深度图



(c) Flower 第 30 帧非关键帧深度图



(d) Flower 第 90 帧非关键帧深度图

图 5 Flower 序列二维图像和对应的生成深度图

Q1 是简单路径查询,查询路径中不含属性值,对于这种查询,RTL-Index 直接从相应的 B+ 树中检索结果,而序列索引 ViST 则需要不同的 B+ 树种逐一进行匹配。Q2 是含有属性值的简单线性查询请求,RTL-Index 在相应的 B+ 树中根据属性值进行检索,这个过程一般只需一次磁盘交换。Q3、Q6 是两个简单分支查询请求,RTL-Index 索引结构具有很好的查询效率。Q5 是带有通配符的查询请求,多次的结构匹配操作影响了 RTL-Index 的效率,但是仍然可以接受。

RTL-Index 索引在响应线性查询和简单分支查询请求时具有非常好的效果,在处理复杂分支查询请求(带“*”、“\”)时由于多次的结构匹配和磁盘页面交换操作影响了执行效率,但仍与序列索引结构 ViST 相当。

5 结束语

RTL-Index 索引实际上是直接对 XML 文档树中的叶子节点建立索引,因为叶子节点具有最大的选择度。文中通过 Trie 树存储叶子节点的前缀路径,实现前缀路径共享,有效地提高了索引的空间效率。文中提出的结构匹配操作解决了 XML 文档分支路径查找的问题,比起传统索引结构中的连接操作有了很大改进。但是在最坏情况下多次的结构匹配操作会增加 I/O 操作,在一定程度上影响查询效率,同时在处理通配符和祖先轴的查询路径时,效率提高不够明显,这都是将要进一步研究的内容。

参考文献:

- [1] 孟小峰,王 宇,王小峰. XML 查询优化研究[J]. 软件学

(上接第 97 页)

参考文献:

- [1] 姜太平. 立体显示技术成熟,短期有望进入家庭市场[J]. 实用影音技术,2009(5):40-42.
- [2] Liang Zhang, Tam W J. Stereoscopic Image Generation Based on Depth Images for 3D TV[J]. IEEE Transactions on Broadcasting, 2005, 51(2):191-199.
- [3] 王惠明,董文辉. “二维+深度信息”的立体电视应用[J]. 广播与电视技术,2009(10):62-65.
- [4] 戴琼海,李 涛. 基于人机交互的平面视频转立体视频的方法[P]. 中国专利: ZL200810102033.1, 2008-09-03.
- [5] 戴琼海,尤志翔,刘继明. 基于实时人机对话的平面视频转立体视频方法[P]. 中国专利: ZL200810111774.6, 2008-10-15.

报,2006,17(10):2069-2086.

- [2] 王国仁,于 戈,杨晓春,等. XML 数据管理技术[M]. 北京:电子工业出版社,2007.
- [3] 王晓峰,于 江. XML 文档索引研究[J]. 科技信息(学术研究),2008(28):351-375.
- [4] 白治国,徐 慧,张霞萍,等. 基于素数编码 Schema 的 XML 索引结构的研究[J]. 计算机工程与设计,2009,30(8):1043-1046.
- [5] Li Q Z, Moon B. Indexing and querying XML data for regular path expressions[C]//In: Apers P M G, Atzeni P, Ceri S, Paraboschi S, Ramamohanarao K, Snodgrass RT, eds. Proc. of the 27th Int'l Conf. on Very Large Data Bases(VLDB). [s. l.]: Morgan Kaufmann, 2001:361-370.
- [6] Goldman R, Widom J. DataGuides: Enable query formulation and optimization in semistructured databases[C]//In VLDB. Massachusetts, USA: Morgan Kaufmann, 1997:436-445.
- [7] Min Jun-Ki, Chung Chin-Wan, Shim K. An adaptive path index for XML data using the query workload[M]//Information Systems. Oxford, UK: Elsevier Science Ltd, 2005:467-487.
- [8] Wang Haixun, Meng Xiaofeng. On the Sequencing of Tree Structures for XML Indexing[C]//ICDE. Washington, DC, USA: IEEE Computer Society, 2005:372-383.
- [9] Rao P, Moon B. PRIX: Indexing and Querying XML Using Pruffer Sequence[C]//ICDE. Washington, DC, USA: [s. n.] 2004:288-300.
- [10] Sleepcat Software. The Berkeley database(berkeleydb)[EB/OL]. 2006-02. <http://www.sleepyeat.com>.
- [11] DBLP xmlrecords[EB/OL]. 2008-09. <http://dblp.unitrier.de/xml/>.
- [12] XMARK: The XML-benchmark project[EB/OL]. 2002-10. <http://monetdb.cwi.nl/xml/2002>.

- [6] 孙 阳. 二维视频转换为三维视频的关键技术研究[D]. 上海:上海交通大学,2008.
- [7] 禹 晶,苏开娜. 块运动估计的研究进展[J]. 中国图象图形学报,2007,12(12):31-40.
- [8] ISO/IEC 13818-2, MPEG-2. “Generic Coding of Moving Pictures and Associated Audio”, Video[S]. 1994.
- [9] JVT-Q042, Revised H. 264 /MPEG-4 AVC Reference Software Manual[S]. 1995.
- [10] 李 淳,苏开娜. 基于菱形搜索的改进的运动估计算法研究[J]. 计算机技术与发展,2008,18(11):117-122.
- [11] 马力妮,郑志辉,潘 峰. H. 264/AV 视频编码技术研究[J]. 计算机技术与发展,2008,18(7):163-166.
- [12] 邢 恺. 运动估计算法研究[J]. 哈尔滨师范大学自然科学学报,2008,24(4):55-57.