

文章编号:1004-9037(2012)06-0670-07

基于OMAP3530的2D到3D视频自动转换系统

黄转娣 农镇铭 赖 豪 刘佩雯 王卫卫 朱跃生

(北京大学深圳研究生院信息工程学院,深圳,518055)

摘要:研究和实现了一个基于OMAP3530的2D到3D视频自动转换系统,重点研究深度图获取和深度信息渲染等主要核心技术及其实现。该系统利用OMAP3530其特有的双核结构,进行系统优化:由其ARM处理器完成系统界面的设置及控制、显示等功能,其DSP处理器实现图像分割和渲染等核心算法,实现了1个高效的2D到3D视频的视频转换系统。2D转3D技术可把现有的大量传统2D视频转换成3D视频,简化和节省了用摄像机直接捕捉3D场景的复杂性和高成本,可以极大解决3D片源不足的问题。

关键词:深度图获取;深度信息渲染;平面视频转换立体视频

中图分类号:TP391.41

文献标识码:A

System of 2D-to-3D Video Auto-Conversion Based on OMAP3530

Huang Zhuandi, Nong Zhenming, Lai Hao, Liu Peiwen, Wang Weiwei, Zhu Yuesheng

(School of Computer and Information Engineering, Shenzhen Graduate School,

Peking University, Shenzhen, 518055, China)

Abstract: The project studies and implements a system for converting 2D to 3D videos automatically based on OMAP3530. It focuses on the algorithms of depth-map extraction and depth-image-based rendering (DIBR). Using the dual-core structure of OMAP3530's to optimize the system, the functions like setup, control and display are implemented by ARM processor, and the algorithms of image segmentation as well as DIBR are implemented by DSP. Therefore, a system converting 2D videos to 3D videos with high efficiency is implemented. The conversion of lots of traditional 2D videos to 3D videos is one way to alleviate the predicted lack of 3D content in the early stages of 3D-TV rollout. The project has great development prospects since it reduce the complexity and money cost compared with the tradition 3D video production using video cameras directly.

Key words: extraction of depth-map; depth-image-based rendering; 2D videos to 3D videos

引 言

3DTV的商业化及产业化,不仅在于3D显示和标准化技术的进步,而且取决于是否能够提供大量广泛的3D视频内容^[1]。虽然3D电影正在制作,但3D视频内容仍不丰富,不足以满足未来3D视频市场需求;而视频内容提供商拥有大量的2D视频内容,因此将2D视频高质量地转换为3D视频,使现有的电影和纪录片可以在3D显示器上播放是缓解3DTV节目素材不足的重要途径。另外,2D普通视频到3D立体视频的转换技术,可以在不增加传

输带宽的情况下,在终端完成视频转换,实现成本也远低于直接的3D拍摄。既解决了立体视频内容不足的燃眉之急,使电影业界可从旧的电影内容获得新的收益,节省3D获取成本,同时也能缓解传输带宽问题。

2D转3D技术可以细分成两类:基于模型的方法(Model-based conversion, MBC)^[2-4]和基于图像的方法(Image-based conversion, IBC)。基于模型的3D重建方法的优点在于可以对单幅图像进行处理,但是它需要事先对大量训练数据进行学习,而且需要3D建模,算法复杂度高,计算量大,不适合实时应用的场合。2D转3D技术的另一种实现方法

是基于图像的方法,这一方法是将 2D 图像分解为一系列对象,为每一个对象指定相应的深度,然后填入封闭区域。转换过程最终生成逐像素的深度图像,用于完成后续虚拟视点的生成过程。基于图像的方法直接对视频或者图像处理,不需要先验知识和 3D 建模,转换的速度与场景的复杂度无关,对计算机性能要求不高,转换速度较快,能够实现实时处理。鉴于以上优点,基于图像的转换方法已经成为当前 2D 转 3D 技术的主流研究方向。

基于图像的 2D 到 3D 转换关键技术是提取准确的深度图。研究者在分析人类生理和心理视觉的基础上,提出了多种用于提取深度信息的深度线索原理^[5-6]。

为了能获得更真实、更舒适的立体效果,本文采用了多种深度信息融合的方法,并在 OMAP3530 上实现了 2D 到 3D 的自动转换系统。深度图提取算法提取了基于运动估计、线条透视和颜色分量 3 种深度图,并做有效融合。其中,针对基于块匹配的运动估计算法存在的问题提出两点改进:使用两种不同大小的匹配块和采用颜色。实验表明,由此深度图提取算法可获得更准确的深度图。

1 算法框架

2D 转 3D 的流程可描述如下。利用各种深度索引从原始的 2D 视频中获取深度信息,生成初步的深度图;然后对深度图进行滤波处理,改善深度图的质量;基于提取的深度图,可将原始的 2D 图像的像素左右平移得到左右眼视图;最后将修复后的左右眼两路视图合成一路 3D 视频,输入到 3D 显示设备中显示。

可见,2D 转 3D 系统中的核心算法主要由深度图获取算法模块和基于深度图的渲染算法 (Depth image based rendering, DIBR) 模块^[7]组成。

1.1 深度图提取算法

本文提出的深度图提取算法如图 1 所示,算法提取了基于运动估计、线条透视和色度分量 C_r 3 种深度图,并加权融合,最后经过联合双边滤波器平滑过滤得到最终的深度图。

1.1.1 基于运动估计的深度图

在同一场景的两个物体,即使运动速度相同,对观察者来说,会感觉较近物体运动速度快,较远物体运动速度慢,这就是人眼视觉系统的特性。

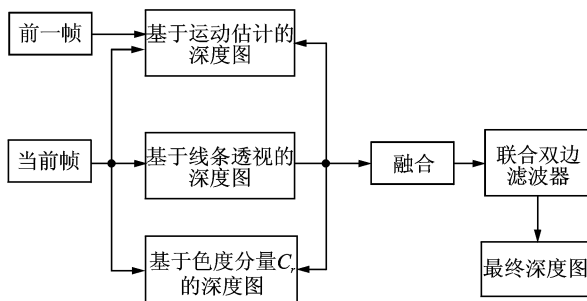


图 1 深度图提取算法

根据人眼视觉系统的特性,使用运动矢量作为一个深度索引提取深度图。通常采用基于块匹配的运动估计算法,然而块的大小和图像中亮度值分布等因素影响块匹配的准确度。当块匹配错误时,将导致错误的运动矢量,深度值也将不符合实际情况。针对以上所述的块匹配的不足,本文在文献[8]工作上,提出了基于改进的块匹配的运动估计算法。

(1)使用两种大小的匹配块。在块匹配算法中,块的大小和形状不仅会影响算法运行速度,还会影响算法的准确度。为了解决这个问题,本文使用两种不同大小的块,8×8 的大块和 4×4 的小块(图 2)。在匹配时,先使用大块进行匹配,当最小误差匹配点在中心时,再使用小块在大块的范围内进行匹配,最后小块的最小误差匹配点就是最优匹配点,小块中的像素拥有相同的深度值。

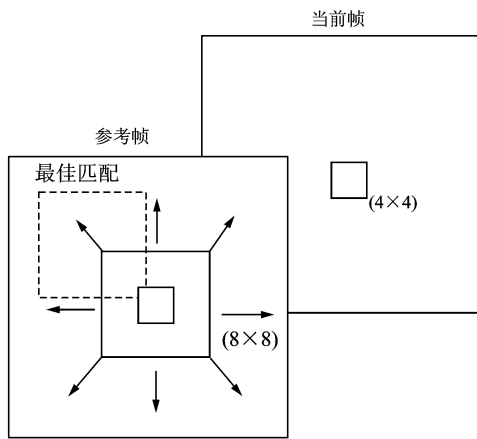


图 2 改进的运动估计(两种不同大小的匹配块)

(2)采用颜色值信息。大部分运动估计中只采用亮度信息,因而亮度信息变化不大的区域会发生误匹配的情况。文献[9]表明在视频解码器,颜色信息可有效修正不可靠的运动矢量。本文提出的第 2 点改进为使用颜色信息进行运动估计获取更可靠

的运动矢量,则块匹配的衡量准则如下

$$E_{m,n} = \sum_{(i,j) \in b_{m,n}^Y} |r_Y(i,j)| + \alpha^* \sum_{(i,j) \in b_{m,n}^{C_b}} |r_{C_b}(i,j)| + \beta^* \sum_{(i,j) \in b_{m,n}^{C_r}} |r_{C_r}(i,j)| \quad (1)$$

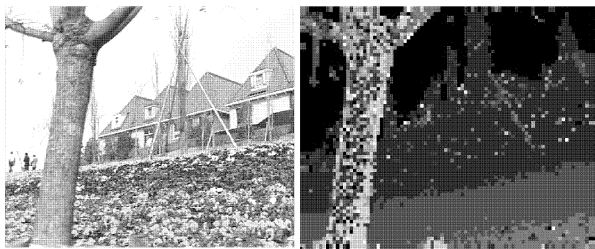
式中: $r_Y(i,j)$, $r_{C_b}(i,j)$, 和 $r_{C_r}(i,j)$ 为当前帧和参考帧 8×8 方块 ($b_{m,n}$) 的 Y , C_r 和 C_b 分量的平均绝对差值 (Mean absolute differences, MAD); α 和 β 为 C_r 和 C_b 分量的权重。因为亮度信息仍然是块匹配最基本的信息,因此选择 α 和 β 值时,不应取过大的权重。

为了降低运动估计的运算量,本文使用菱形搜索算法。像素点 (i,j) 的深度值 $D(i,j)$ 与运动矢量的转换公式如下

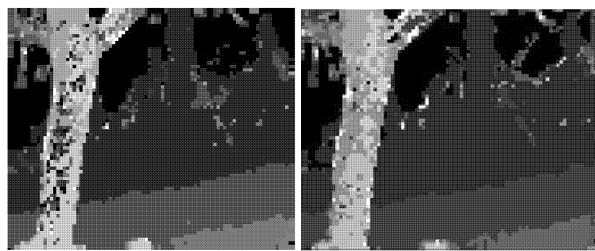
$$D(i,j) = C \sqrt{MV(i,j)_x^2 + MV(i,j)_y^2} \quad (2)$$

式中: $MV(i,j)_x$ 和 $MV(i,j)_y$ 分别为运动矢量的水平和垂直分量; C 为常数。

经过改进的基于块匹配的运动估计算法取得了更好的性能,如图 3 所示。



(a) 原始二维视频 “flower” (b) 由块匹配算法提取的深度图



(c) 采用点(1)的改进方法后的块匹配算法提取的深度图 (d) 采用点(2)的改进方法后的块匹配算法提取的深度图

图 3 块匹配算法提取的深度图

1.1.2 基于颜色分量 C_r 的深度图

研究表明^[10],不同的物体有不同的颜色分量 C_r ,因此可以利用 C_r 作为一个深度索引,获得的深度图中不同物体的边界将有不同的深度值。

将二维图像的 C_r 分量图映射至边界为 $[D_a, D_b]$ 的线性空间便可达到基于 C_r 分量的深度图,其中 D_a, D_b 是运动视差的间隔。

由 C_r 分量提取得到的深度图并不是精确的,

但利用这个深度图可以修复基于运动估计的深度图中错误的深度值。

1.1.3 基于线条透视的深度图

在现实生活中,人眼观看远近景物有如下的透视规律:(1)物体远近不同,人感觉它的大小不同,愈近愈大,愈远愈小,最远的小点会消失在地平线上;(2)有规律地排列形成的线条或互相平行的线条,越远越靠拢和聚集,最后会聚为一点而消失在地平线上;(3)物体的轮廓线条距离视点越近越清晰,越远则越模糊。

在绘画和摄影领域,艺术家常用线条表现画面深度和距离的方法,使二维图像产生三维空间感,这就是线条透视。因此,二维图像的线条透视规律也是深度信息的一个线索,且研究^[11]表明,基于线条透视的深度信息能使三维效果更舒适,更适合人眼观看。本文提出的方法中,提取基于线条透视的深度图实现的方法基于了文献^[11]的工作,即赋予二维图像从上到下呈阶梯状线性增加的深度值。

1.1.4 深度图融合

提取以上3种深度图后,使用线性模型融合这3种深度图,最终的深度图由以下融合公式所得

$$D_{all} = D_m \times W_m + D_c \times W_c + D_l \times W_l \quad (3)$$

$$W_m + W_c + W_l = 1 \quad (4)$$

式中: D_m , D_c , D_l 分别对应基于运动、基于颜色分量和基于线条透视获得的深度值,而 W_m , W_c 和 W_l 是它们对应的权重。

基于运动估计获得的深度信息是主要的深度信息,而基于颜色分量 C_r 和基于线条透视的深度信息则是辅助信息,因此 W_m 值相对较大,而 W_c 和 W_l 值相对较小。对于不同的类别的视频,各权重可能会略有改变。

1.1.5 联合双边滤波器

联合双边滤波器^[12]具有修复融合深度图中的阶梯状边缘的效果。联合滤波器是在双边滤波器的基础上改进。融合深度图中的像素点 p ,其深度值为 $D(p)$, q 点是 p 点邻域 $N(p)$ 的点。 I_{ref} 表示参考帧图像,则联合双边滤波器表示为

$$D(p) = \frac{1}{\sum_{q \in N(p)} W_q} \times \sum_{q \in N(p)} W_q \times D(q) \quad (5)$$

式中: W_q 是像素点 q 的深度值的权重,定义如下

$$W_q = F_{os}(\|p - q\|) \times G_{sr}(|I_{ref}(p) - I_{ref}(q)|) \quad (6)$$

式中: $F_{ss}()$ 和 $G_{sr}()$ 分别为空间和深度值的权重函数。

1.2 DIBR 算法模块

DIBR^[13]是目前在3D-TV系统中倍受关注的图像处理技术。图4显示了一个基本的DIBR系统包含的3个阶段:(1)对深度图进行滤波处理,将物体边缘的深度变化值平滑化;(2)3D图像扭曲阶段是将中间视角图像和深度图合成左右眼视图;(3)填洞处理用于填补左右眼视图中的非闭塞区域。

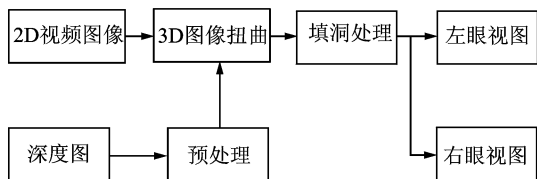


图4 DIBR 算法框图

1.2.1 深度图预处理

预处理的目的是为了在合成图像中减少非闭塞区域的数量。对深度图的平滑处理,选择非对称高斯滤波器。因为深度信息的失真,高斯平滑滤波器会在合成的虚拟图像中产生明显的纹理缺陷。为了解决这个问题,采用了非对称高斯滤波器,它基于高斯滤波器的架构对纵向的深度信息使用较大的平滑力度,对横向的深度信息使用较小的平滑力度。这样,竖直的物体,在平滑处理后将具有相似的深度值。图5为经非对称高斯滤波器平滑后的深度图。

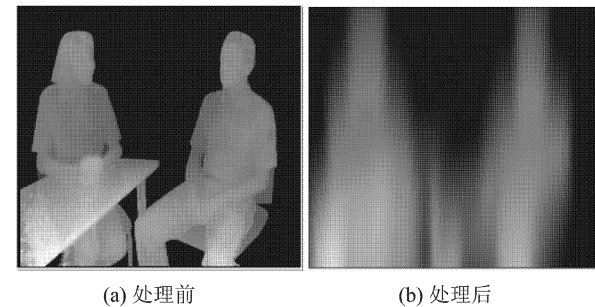


图5 非对称高斯滤波器平滑效果

1.2.2 3D 图像扭曲

如图6所示, c_c 是单目摄像机,即原始的拍摄二维视频的摄像机;而 c_l, c_r 则是模拟人的双眼的双目摄像机; f 是人眼的焦距; z 是物体的实际深度; x_c 是物体在2D图像中的位置; x_l, x_r 则分别是物体在左眼视图和右眼视图的位置。由图6可得

$$x_l = x_c + \frac{t_x}{2} \frac{f}{Z} \quad x_r = x_c - \frac{t_x}{2} \frac{f}{Z} \quad (7)$$

因此,根据式(7)来进行图像扭曲,虚拟的左右眼可以由中心图像和它的深度图提供 t_x 和 f 的值来产生。为了不失去一般性,实验中的 f 值取1。

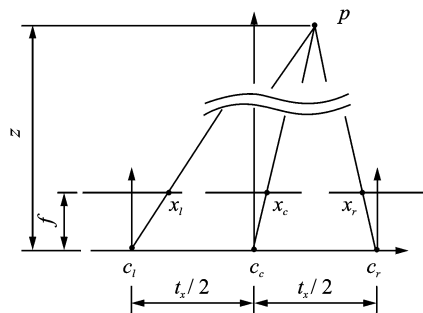


图6 DIBR 原理

1.2.3 “填洞”算法

使用3D图像扭曲的DIBR技术有两个主要的问题需要解决,这两个问题就是闭塞和非闭塞问题。

闭塞是指原图像中两个不同位置的像素在虚拟视图中被移动到了同一个位置;非闭塞是指原图像中被遮挡的部分场景在虚拟视图中出现,没有信息填补造成空洞。闭塞问题容易解决,可以使用具有更高深度值(距离镜头更近)的像素来产生虚拟视图。而非闭塞问题则不容易解决,因为在原图像和深度图中都没有相应信息来产生这些新出现的像素,而这些新视图中空白的像素,也就是空洞。

因此完整的DIBR技术需要填洞处理来填补这些空白的区域。本文使用的是水平内插填洞算法。在水平内插填洞法中,空洞由水平方向上的边界像素值内插得到。在图7中可以对比图像填洞前后的区别。

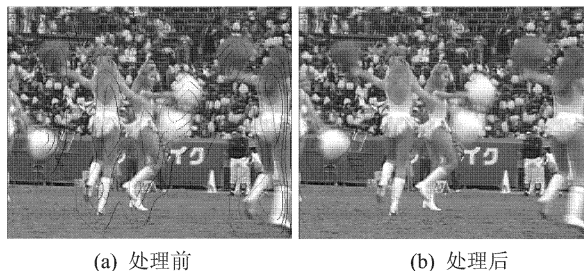


图7 填洞处理效果

2 系统实现

2.1 系统方案

本文的目标是实现实时高效地将2D视频转换为3D视频,因此要求处理器能快速高效地处理视

频这类大数据,因此选用TI公司的一款高效的多媒体处理器平台——多核处理器OMAP3530。OMAP3530采用720MHz ARM®Cortex™-A8核与520MHz C64x+DSP核,并且支持SGX 2D/3D图形加速器。

2D 视频转 3D 视频系统的系统方案如图 8 所示,主要包括视频输入、系统控制、数据处理、视频输出显示和电源管理等几个部分。

OMAP3530 的 ARM 核运行基于 Linux 操作系统的应用程序,所用的外围设备都由 ARM 负责控制。ARM 端响应外部请求,并发送 2D 视频数据至 DSP,DSP 完成视频转换,ARM 输出显示 3D 视频。

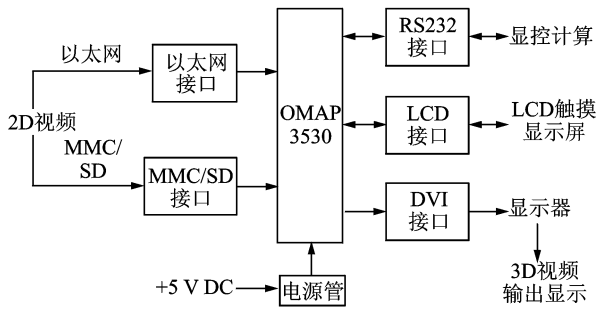


图 8 2D 转 3D 系统图

2.2 硬件设计

根据系统方案设计,选择 TI 与 Mistral Solutions 联合开发的 OMAP35x 评估模块 EVM 为硬件开发平台。OMAP35x EVM 采用模块化的可扩展架构进行设计,主要使用了 OMAP35x EVM 其中的 3 个电路板模块:主板,处理模块和电源模块。

2D 转 3D 视频系统的硬件设计框图如图 9 所示。

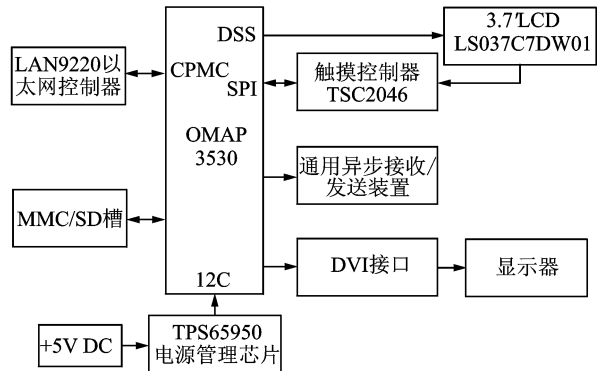


图 9 2D 转 3D 系统的硬件设计框图

2.3 软件设计

系统软件设计框图如图 10 所示。视频转换的

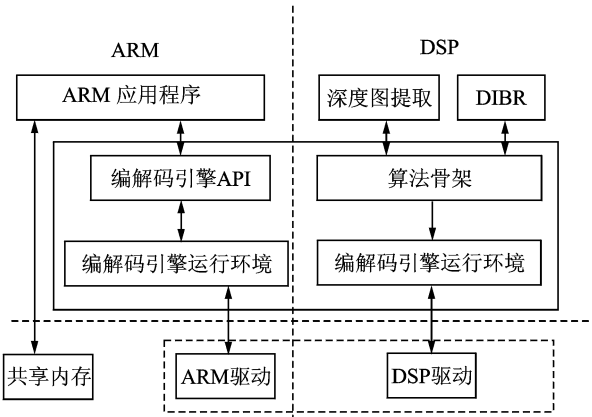


图 10 系统软件设计框图

应用程序由 4 个线程组成,分别是转换请求线程、视频转换线程、视频显示线程和系统控制线程。转换请求线程把视频数据送到 ARM 和 DSP 的共享缓冲内存,并通知 DSP 执行视频转换算法,包括深度图获取算法、深度信息渲染算法以及编码输出算法。转换线程负责控制 DSP 侧的转换算法并从共享内存中读取 2D 视频数据。视频显示线程从视频缓存中读取已完成转换的 3D 视频数据帧,最后通过设备驱动输出显示。系统控制线程负责响应来自触摸屏或鼠标或扩展计算机接口的请求,并执行相应操作。

OMAP3530 的 DSP 核上运行 DSP/BIOS 实时操作系统和深度图获取算法、深度信息渲染算法。所有算法的接口都符合 TI xDAIS 标准,由 Codec Engine 调用。除了算法,DSP 核上还集成了管理内存和 DMA 的 Framework Component。

ARM 核和 DSP 核的通信由 TI 提供 Codec Engine 软件框架负责。Codec Engine 是介于应用程序和具体算法之间的软件模块,其中 VISA API 通过 Stub 和 Skeleton 访问 Engine SPI 最终调用算法。ARM 和 DSP 的所用共享缓存内存都是通过 CMEM 模块在 DDR 中分配的,缓存内存地址连续且与 DSP 核 Cache 对齐。

另外,为了提高 DSP 的运行速度,还采取了一系列的优化策略:如使用 TI C6000 的编译工具提供的编译选项对代码进行各种优化;使用内联函数,进行循环优化、定点优化和数学函数优化等。

3 系统评测

3.1 效果测评

使用 352×288 的 avi 视频 out3 作为 2D 图像输

入,获得初步深度图、平滑后的深度图和渲染生成的3D图像,如图11所示。

可见,本系统实现了深度图获取以及DIBR等主要核心技术算法,能够自动地将2D视频转化为3D,并且合成的3D视频有明显的立体效果。

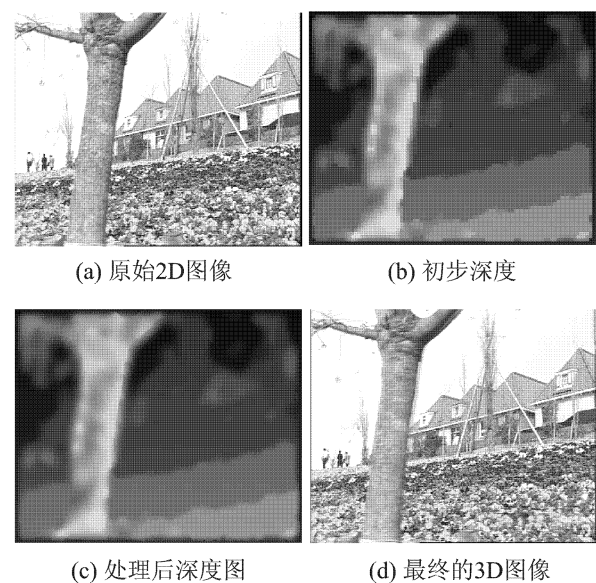


图11 系统效果图

3.2 效率测评

使用3段帧率不同的2D视频Girl,Out3,Resize作为输入数据,其中视频大小均为 352×288 ,分别测评深度图生成算法、DIBR算法和整个2D到3D视频转换系统的效率。测评结果如表1所示。

根据表1的数据,对于测试视频,深度图获取算法模块的运行速度最差为8帧/s,最好为12帧/s。DIBR算法模块的运行速度最差为27帧/s,最好为30帧/s。整个2D到3D视频自动转换系统的平均处理速度为8帧/s。

表1 测评数据

时间/ (ms·帧 ⁻¹)	视频		
	Girl	Out3	Resize
提取深度图	115.3	83.4	79.9
DIBR	42.2	39.0	35.4
2D to 3D	36.7	33.0	34.6

可见,本系统在不牺牲转换效果的情况下具有简便、速度快的特点。

4 结束语

为了能获得更真实、更舒适的立体效果,本文

采用了深度信息融合的方法,在OMAP3530上实现了2D到3D的自动转换系统。2D到3D转换算法由深度图提取算法和DIBR算法组成。深度图提取算法提取了基于运动估计、线条透视和颜色分量3种深度图,并做有效融合。其中,针对基于块匹配的运动估计算法存在的问题提出两点改进:使用两种不同大小的匹配块和采用颜色。实验表明,由此深度图提取算法可获得更准确的深度图,因而通过DBIR渲染而得的立体图像更真实。该算法在OMAP3530上做了合理分配和优化,对于大小为 320×288 的图像序列,平均处理速度为8帧/s,处理速度较理想。2D转3D技术可把现有的大量传统2D视频转换成3D视频,简化和节省了用摄像机直接捕捉3D场景的复杂性和高成本,可以极大解决3D片源不足的问题,具有良好的发展前景。

参考文献:

[1] Kubota A, Smolic A, Magnor M, et al. Multiview imaging and 3DTV [J]. IEEE Signal Processing Magazine, 2007,24(6):10-21.

[2] Saxena A, Sun M, Ng A Y. Make3D: learning 3D scene structure still image[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009,31(5):824-840.

[3] Saxena A, Sun M, Ng A Y. 3-D Reconstruction from sparse views using monocular vision[C]//IEEE 11th International Conference on Computer Vision. [S. l.]:IEEE,2007:1-8.

[4] Saxena A, Chung S H, Ng A Y. 3-D depth reconstruction from a single still image[J]. International Journal of Computer Vision (IJCV), 2008, 76(1): 53-69.

[5] Fehn C. A 3D-TV system based on video plus depth information[C]. Proc of 37th Asilomar Conference on Signals, System, and Computers. Pacific Grove, CA, USA:[s. n.],2003:1529-1533.

[6] Tam W J, Zhang L. 3D-TV content generation: 2D-to-3D conversion [C] // Multimedia and Expo 2006 IEEE International Conference. [S. l.]:IEEE, 2006:1869-1872.

[7] Fehn C. Depth-image-based rendering (DIBR)[C]. Compression, and transmission for A New Approach 3D-TV[C]//Proc SPIE Conf Stereoscopic Displays and Virtual Reality Systems XI, CA, USA: GPIE,2004:93-104.

[8] Ideses I, Yaroslavsky L P, Fishbain B. Real-time

2D to 3D video conversion[J]. Journal of Real-Time Image Processing, 2007,2(1):3-9.

[9] Huang A M, Nguyen T. Motion vector processing using the color infor-mation[C]//IEEE International Conference on Image Processing (ICIP). Cairo: IEEE,2009:1605-1608.

[10] Wa J T, Carlos V, Filippo S. Three-dimensional TV: a novel method for generating surrogate depth maps using colour information[J]. SPIE Electronics Imaging, 2009,7237A:1-8.

[11] Tsai S F, Cheng C C, Li C T. A real-time 1080p 2D-to-3D video conversion system[J]. IEEE Transactions on Consumer Electronics, 2011,57(2):915-922.

[12] Kopf J, Cohen M F, Lischinski D, et al. Joint bilateral upsampling[J]. ACM Trans Graph, 2007,26(3):96-103.

[13] Fehn C. A 3D-TV approach using depth-image-based rendering (DIBR)[C]//Proceedings of Visualization, Imaging, and Image Processing. Spain:[s. n.], 2003:482-487.

作者简介:黄转娣(1988-),女,硕士研究生,研究方向:计算机视觉,E-mail: zhuandi. h@gmail. com; 农镇铭(1988-),男,硕士研究生,研究方向:嵌入式系统、计算机视觉;赖豪(1990-),男,硕士研究生,研究方向:计算机视觉;刘佩雯(1989-),女,硕士研究生,研究方向:计算机视觉;王卫卫(1987-),女,硕士研究生,研究方向:计算机视觉;朱跃生(1959-),男,博士,教授,研究方向:宽带网络通信与信息安全、新媒体信号处理及版权保护技术。