

自由视点视频的虚拟图像合成

李 寒,程 浩 冯先成

(武汉工程大学电气信息学院,湖北 武汉 430205)

摘 要:通过一种有效的基于深度图像描绘技术的虚拟图像合成方法评价一种快速深度图像估计算法.此快速算法采用区域匹配 SSD 算法得到初始视差,构造出图割算法的能量函数求深度图.仿真结果证明了快速算法能获得高质量的 PSNR,可以合成较好的虚拟图.

关键词:图割法;SSD 算法;深度图像估计;深度图像描绘

中图分类号:TU502

文献标识码:A

doi:10.3969/j.issn.1674-2869.2011.11.019

0 引 言

近几年,随着 3D 视频技术的不断发展,学者们开始关注一种叫多视点视频的新 3D 表示技术.这种技术中,场景信号是通过一系列不同视点的摄像机同步获得的,即 3D 视频信号是由一系列序列不同摄像机拍摄视频序列组成的.这种技术被广泛的用在自由视点视频(Free Viewpoing Video:FVV)中^[1].在应用中,用户可以自由的选择视点,而不受视频视角的限制.目前商业广播系统还不能支持多视频信号的处理和传输;而且,在接受端计算合成虚拟视频也是非常低效的.为了解决这一问题,研究者提出了不同的编码方法和虚拟视频合成方法,并被证明是有效的.文献[2]中,利用视频间视差与 3D 几何特性编码和合成虚拟视频;同样的,在文献[3]中,利用小波变换去除视频中的时间和空间的冗余信息,并且对子带进行熵编码.这两种方法有效的去除冗余信息,满足了 3D 信号的传输要求.但是,这两方法对 3D 几何特性要求很高,这也成为了他们应用的瓶颈.相比之下,在基于图像描绘(Image-based rendering:IBR)^[4]中,多视点视频信号由一些彩色视频信号和利用这些彩色视频插值得到的中间视频信号组成的.这件技术不需要附加条件和几何条件的约束就能合成出十分逼真的虚拟视频来.然而,它对带宽的要求很高.基于深度图像的自由视点编码是一种具有网格基编码^[5]和 IBR 的优点的编码方法.在选择视点时,它做了一个光场函数的近似,这种近似允许减少摄像机视点数目,减少带宽.

1 基于深度图像的视点合成

1.1 自由视点 DIBR 算法

基于深度图像描绘(Depth Image Based Rendering:DIBR)是利用参考视点的图像及其每个像素对应的深度信息,并且结合了相机的内外参数,通过三维变换来合成场景中虚拟视点图像,该过程包括以下四步^[6]:

第一步变换(warping):通过多个视频点图像和对应深度图像信息,即相机的内外参数,进行三维变换合成粗虚拟图像.由于深度信息是灰度图像表示,其中像素点灰度值在 0 到 255 之间,灰度越大表现越远,亦则相反.公式将灰度信息和深度信息联系起来.

$$d(x_{\text{ref}}) = \frac{255}{p(x_{\text{ref}})} \left(\left(\frac{1}{d_{\min}} - \frac{1}{d_{\max}} \right) + d_{\max} \right) \quad (1)$$

其中 $p(x_{\text{ref}})$ 是深度图的灰度值, $d(x_{\text{ref}})$ 是像素点得深度值, d_{\min} 和 d_{\max} 分别是透视投影所变换所采用的远近剪切值.得到深度信息后,再利用参考视点相机的内外参数进行三维空间投影,公式如下:

$$\mathbf{P}_{\text{wc}} = \mathbf{R}_{\text{ref}}^{-1} \cdot \mathbf{A}_{\text{ref}}^{-1} \cdot \mathbf{P}_{\text{ref}}^{-1} \mathbf{D} - \mathbf{R}_{\text{ref}}^{-1} \mathbf{t}_{\text{ref}} \quad (2)$$

其中 \mathbf{R}_{ref} 、 \mathbf{A}_{ref} 、 \mathbf{t}_{ref} 分别表示相机内参数、旋转矩阵和平移矩阵, \mathbf{D} 向量矩阵(深度值), 矩阵 \mathbf{P}_{wc} 是三维空间中坐标.最后根据公式(3)将三维空间点投影到虚拟图像平面.

$$\mathbf{P}_{\text{ta}} = \mathbf{A}_{\text{ta}} \cdot \mathbf{R}_{\text{ta}} \cdot (\mathbf{P}_{\text{wc}} + \mathbf{R}_{\text{ta}}^{-1} \cdot \mathbf{t}_{\text{ta}}) \quad (3)$$

在这个过程中,先找出不连续点,并在投影物体时,去除这些点.

第二步:处理裂缝(Cracks). 三维变换合成会

收稿日期:2011-09-15

作者简介:李 寒(1983-),男,湖北应城人,硕士研究生.研究方向:嵌入式系统及应用、网络通信、图像处理.

指导老师:冯先成,男,副教授,硕士,硕士研究生导师.研究方向:网络通信.

导致虚拟出得图像中有空洞或裂缝现象,而这部分通过中值滤波对空洞和裂缝进行平滑处理。

第三步:混合(Blending). 这个部分是将两个被投影的图像混合成为可见的前景景物. 这个步骤可以减少因遮挡的误投影。

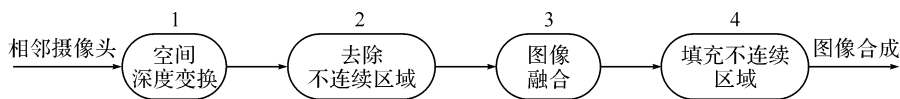


图 1 DIBR 算法流程

Fig. 1 DIBR algorithm

1.2 深度图像估计

深度图像估计是 DIBR 算法的关键技术之一,因为深度图像的质量好坏直接影响整个三维投影以及最后的虚拟图像的质量. 为了获得高质量的深度图像,学者们已经提出了深度图有基于区域的算法(local-based algorithms)和基于全局的算法(global-based algorithms)两种方法. 区域算法主要有 SAD^[7], SSD^[7], NCC^[8]等算法. 全局算法主要有图割(Graph Cut: GC)^[9]、动态规划(Dynamic Programming: DP)^[10]、信念传播(Belief Propagation: BP)^[11-12]等算法. 本文采用的快速图割算法步骤如下。

1.2.1 求初视差 设 $S(xy)$ 代表平方差值算法. 利用改进 SSD 算法求取视差的初始值,假设 $S(x, y, d)$ 表示 $I_1(x, y)$ 与 $I_1(x, y+d)$ 区域的 S 值时,则 $I_1(x, y+1)$ 与 $I_1(x, y+1+d)$ 区域的 $S(x, y+1, d)$ 值可以用 $S(x, y, d)$ 值表示:

$$S(x, y+1, d) = S(x, y, d) + f(x, y+1, d) \quad (5)$$

其中 $f(x, y+1, d)$ 为:

$$f(x, y+1, d) = \sum_{i=-n}^n [I_1(x+i, y+m+1) - I_2(x+i, y+m+1)]^2 \sqrt{b^2 - 4ac} - \sum_{i=-n}^n [I_1(x+i, y-m) - I_2(x+i, y+d-m)]^2 \quad (6)$$

1.2.2 校正初始视差 校正初始视差是为了提高视差的精度. 通过左右一致验证法^[13]对视差图进行校正. 左右一致验证法是假设左右图分别为右左的参考图,也就是分别求出以左右参考图的右左深度视差($D_L(p), D_R(p)$),然后求两幅视差图的一致性。

$$D_L(p) = D_R(p - D_L(p)) \quad (7)$$

如果视差值相同,就是可以确定这个值为视差值,如果不一致则是遮挡点,那么将这个点做为图割的限制项进行优化能量函数,使能量函数最

第四步:内绘(Inpainting). 和第二、三个步骤一样,这个步骤也是对虚拟出图像中问题进行修复. 在这个部分,利用深度图像和视频图像的权重关系对虚拟图像中的黑点像素进行插值修复. 如图 1 所示。

小.

1.2.3 构建能量函数求最终的深度图 将左右一致性求出遮挡点作为约束项 f 初始,能量函数如下:

$$E(f) = E_{\text{data}}(f) + E_{\text{smooth}}(f) + E_{\text{occlusion}}(f) + E_{\text{frange}}(f) \quad (8)$$

其中 $E_{\text{frange}}(f)$:

$$E_{\text{frange}}(f) = \begin{cases} 0 & f \in (f_{\text{初始}} \pm f_{\text{range}}) \\ \infty & \text{其它} \end{cases} \quad (9)$$

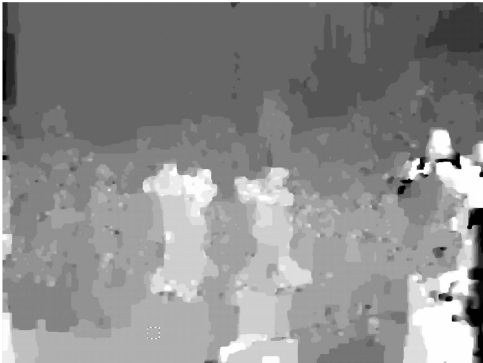
2 实验结果与分析

为了验证本文提出的算法的性能,本文将采用美国 Middlerbuy 大学计算机视觉研究中心提供 8 视点的 640×480 的 Ballroom 测试序列和微软研究所提供的 8 视点的 1024×768 的 Ballet 测试序列进行算法耗时和合成虚拟视点的仿真实验. 在本试验中,将选择 Ballroom 和 Ballet 测试序列的视点 0 和做为虚拟视点,视点 2 为辅助视点,视点 1 为虚拟视点. 计算机硬件条件为:单核 CPU 主频为 2.93 GHz,内存为 1 G,软件编译环境为 vs2005。

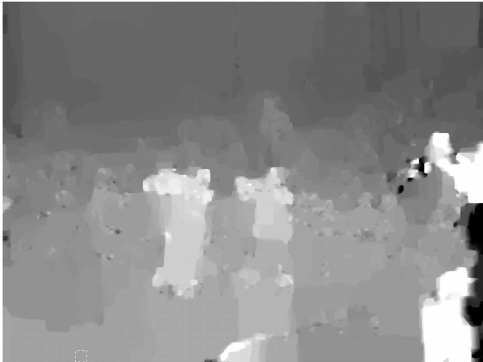
因为采用了初始像素视差的计算结果来控制网络的规模,在允许视差范围内增加的节点要比搜索全部视差范围的节点要小得多,所以这样就减少了网络中的节点数,降低了计算复杂度. 表 1 是两种算法对 Ballroom、Ballet 的匹配耗时情况,不难看出本算法明显减少匹配时间. 图 2 为不同算法估计出的深度图. 将虚拟合成出来的图像进行噪声峰值性噪比和结构相似指标(SSIM)评价,发现利用本文算法产生的深度图像合成出得虚拟图像有较高的 PSNR,有较高的相似度,结果如表 2 和表 3 所示. 图 3 中依次是 Ballet 序列的第 1 虚拟合成图像, Ballroom 序列的第 1 帧和第 95 帧虚拟合成图像。

表 1 算法耗时

Tab 1	Algorithm time-consuming		ms
	Ballroom	Ballet	
GC	7 507.5	102.96	
本文算法	4 237.5	72.36	



(a) Gc算法估计的深度图



(b) 本文算法估计的深度图

图 2 不同算法估计的深度图

Fig. 2 GC and proposed algorithm estimate depth maps

表 2 虚拟图像质量评估

Tab 2	Virtual image quality assessment		
	“Ballet”0frame	“Ballroom”0frame	“Ballroom”95frame
PSNR	32.736 3	31.157 3	30.987 1
SSIM	0.948 9	0.905 5	0.911 8

表 3 虚拟图像质量评估比较

Tab 3	Comparison of the PSNR GC and proposed algorithm		
	“Ballet”0frame	“Ballroom”0frame	“Ballroom”95frame
本文算法	32.736 3	31.157 3	30.987 1
GC	25.121	25.173	26.986



(a) "Ballet"squeeze1 第0帧 (b) GC算法合成第0帧视点 (c) 本文算法合成第0帧视点



(a) "Ballroom"视点1 第0帧 (b) GC算法合成第0帧视点 (c) 本文合成第0帧视点



(a) "Ballroom"视点1 第95帧 (b) GC算法合成第95帧视点 (c) 本文算法合成第95帧视点

图 3 合成虚拟序列

Fig. 3 Synthetic virtual sequence

3 结 语

上述采用的 DIBR 算法合成虚拟图像客观和主观上证明了快速立体匹配算法的性能. 采用改进的 SSD 匹配算法快速获得初始视差图, 并将其作为 GC 算法的参考项, 减少网络的规模, 优化了能量函数, 在保证深度图像质量的情况下, 进一步提高了估计效率. 而虚拟合成的图像会有裂纹、重影、空洞等问题, 这些问题在高清视频将被放大, 不能满足高清 3D 视频观看要求.

参考文献:

[1] 程浩,王龙海. 三维视频中深度信息估计算法[J]. 武汉工程大学学报,2011,33(19):105-110.

[2] Mueller K, Smolic A, Kautzner M, et al. Rate distortion-optimized predictive compression of dynamic 3-D mesh sequences [J]. Signal process image comm,2007,21(9):812-828.

[3] Cho J W, KIM M S, Valett, et al. A 3-D mesh sequence coding using the combination of spatial and temporal wavelet analysis [J]. Lecture notes in computer science,2007,21(9):812-828.

[4] Shum H Y, Chan S C, Kang S B. Image-based rendering [M]. Berlin Germany: Springer-Verlag, 2007:344-350.

[5] Merkle P, Smolic A, Mueller K, et al. Efficient prediction structures for multiveiw video coding[J]. Circuits syst video technol,2007,17(11):1461-1473.

[6] Jung C, Jiao L. Depth-preserving DIBR based on disparity map over T-DMB[J]. Eelectronics letters, 2010,46(9):628-629.

[7] Yu Hsun Lin, Ja-Lingwu. A digital blind water marking for depth-image-based rendering 3D images [J]. Broadcasting: IEEE transcation on, 2011, 57 (2):602-611.

[8] Tanimoto M. Overview of free viewpoint television [J]. Signal processing: Image Comm, 2006, 21 (6): 454-461.

[9] 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.

[10] 杨海涛. 三维视频编码技术研究[D]. 西安:西安电子科技大学,2009.

[11] Yin Zhao, Ce Zhu, et al. Depth no-synthesis-error model for view synthesis is in 3D video[J]. Image processing: IEEE transcation on, 2011, 20 (8): 2221-2228.

[12] Daribo I,Saito H. A novel inpainting-based layered depth video for 3DTV [J]. Broadcasting: IEEE transactions on,2011,57(2):553-541.

[13] Cheol kon Jung, Jiao L. Disparity-map-based rendering for mobile 3DTV. [J]. Consumer electronics: IEEE transactions on, 2011, 57 (3): 1171-1175.

Virtual view synthesis for free viewpoing video

LI Han , CHENG Hao ,FENG Xian-cheng

(School of Telecommunication&Information Engineering,Wuhan Institute of Technoloy,Wuhan 430205,China)

Abstract: An effective virtual view synthesis approach was proved by utilizing the technology of depth-image-based rendering(DIBR). This papers was utilized a fast depth images estimation based on SSD algorithm and graph cuts algorithm. First,the SSD matching method was used as a similarity decision criterion to determine the initial disparity map. When the energy function is constructed to calculate the depth map. Simulation results demonstrate that the fast algorithm provides high PSNR and synthesised good views.

Key words: graph cuts;SSD algorithm;depth images estimation;DIBR.

本文编辑:陈小平