

基于视差稠密化的视图合成^①

张 倩^{②*} 安 平^{***} 张兆杨^{***} 吕朝辉^{***}

(* 上海大学通信与信息工程学院 上海 200072)

(** 上海大学新型显示技术及应用集成教育部重点实验室 上海 200072)

(*** 中国传媒大学信息工程学院 北京 100024)

摘 要 针对现有三维视频处理系统需要有较高的传输带宽才能合成高质量的中间视图的情况,提出了一种新颖的解码端中间视合成方法,该方法仅利用解码端获得的稀疏视差场来合成高质量的中间视图,这有助于减少传输带宽。该方法以解码端获得的稀疏视差场作为初始值,通过均值偏移算法和总体最小二乘法对其作后处理,获得解码端的稠密视差场,从而合成解码端高质量的中间视画面。实验结果证明,所提方法能仅利用稀疏视差场在解码端获得平滑稠密视差场和高质量的中间视图。

关键词 视差,中间视,图像分割,分割误差,均值偏移算法

0 引 言

随着计算机和通信技术的发展,三维(3D)视频显示将成为下一代主流视频显示技术。与传统 2D 视频相比,3D 视频使人们在观看图像时富有立体感和沉浸感。自由视点电视(free viewpoint television, FTV)是进入本世纪后迅速发展起来的 3D 视频新技术,立体感和交互性是其两个基本特点,交互性体现在当用户选择视点画面时,系统能根据用户的选择生成相应的画面,即生成任意位置的中间视图图像^[1,2]。中间视图生成是三维视频处理系统的关键技术之一。在现有系统中,解码端需要利用稠密视差场才能合成高质量的中间视图。显然,三维视频处理系统中编码端和解码端传输的数据量越高,则解码端可能获得的视差场越稠密。因此,现有系统需要较高的传输带宽才能保证所合成的中间视图图像的质量。

本文通过充分利用 3D 视频编解码系统的解码信息并进行信息修正,提出了一种解码端中间视合成方法,该方法仅利用解码端获得的稀疏视差场来合成高质量的中间视图,有助于减少传输带宽。该方法实质是一种后处理方法,它以解码端的稀疏视差场为初值,利用均值偏移算法分割参考图像,随后利用总体最小二乘(total least squares, TLS)算法消

除每个分割区域上由传输噪声和分割误差所带来的影响,从而获得该分割区域上稠密视差的精确估计,最后基于稠密化后的视差场完成中间视合成,获得高质量的中间视画面。仿真实验证明所提方法是有效的,能仅利用稀疏视差场,在解码端获得平滑稠密视差场和高质量的中间视图。

1 基于图像分割的稠密视差场估计

由于视差估计的准确性是获得高质量的中间视画面的关键,因此目前的中间视生成算法主要集中于如何获得精准和稠密的视差场来改善中间视内插的效果,但是求精准的视差场需要很大的计算量,这在很多实时性要求较高的应用中并不合适。降低计算量的主要途径之一是避免复杂的视差场估计,对于立体或多视点视频编码系统来说,为了提高编码效率,往往传送的是稀疏视差场,例如我们在文献[3]中编码的视差场就是这样的稀疏场。这一类解码视差场往往含有较多的伪匹配,如果直接用于中间视内插,将会产生低质量的中间视图。

因此,我们提出一种基于图像分割的稠密视差场精确估计方法,以修正初始的解码视差场。不失一般性,我们考虑基线较小的平行相机系统,以文献[3]中立体视频编解码系统解码出的稀疏视差场

① 国家自然科学基金(60832003, 60672052),上海市曙光计划(06SG43),上海市教委重点项目(09ZZ90)和教育部重点实验室开放性课题(07k04)资助项目。

② 女,1983年生,博士生;研究方向:多视点视频编码及其应用;联系人, E-mail: zhangqianleslie@yahoo.com.cn (收稿日期:2009-04-01)

为初值。可以认为视差图是由一系列不重叠的小平面组成,每个小平面对应于一个分割区域。首先,本文采用图像分割的方法将参考图像分割为相似区域,随后利用总体最小二乘(TLS)算法去除每个分割区域上由传输和分割误差所带来的噪声。

在图像分割中^[4,5],Comaniciu 和 Meer 提出的均值偏移算法^[6]是一种新颖的分割算法,通过对特征空间中的样本点进行聚类,得到图像特征空间中密度函数梯度为零的模式点,从而完成分割。该方法运算量低且能很好地保留图像边缘,本文将用于参考图像的分割。

均值偏移算法的核心是核密度函数的估计,其理论基础如下:假设 d 维空间有 n 个点 $p_i, i = 1, 2, \dots, n$, 则在点 p 处定义的用核函数 $K(\cdot)$ 和窗函数 H 得到的核密度估计算子为

$$\hat{f}(p) = \frac{1}{n} \sum_{i=1}^n K_H(p - p_i) \quad (1)$$

其中 K_H 是一个积分为 1 的以零为原点的非负函数:

$$K_H(p) = |H|^{-1/2} K(H^{-1/2}p) \quad (2)$$

本文中为了减少计算的复杂度,我们选择窗宽矩阵为单位矩阵的比例阵 $H = h^2 I$, 这样只需要指定一个大于零的窗宽参数 h , 在这种情况下核密度估计算子转化为

$$\hat{f}(p) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{p - p_i}{h}\right) \quad (3)$$

其中,核函数 $K(\cdot)$ 是一个关于图像位置和该位置处的色度(亮度)信息的函数,通常用 Epanechnikov 核^[6]作为核函数通过对核密度算子求梯度来跟踪到密度最大点,即完成对特征空间中的样本点进行聚类,得到图像特征空间中所谓的模式点。由于不同模式之间有可能存在很大相似性,我们把距离相近的模式点标识为同一类,赋予相同的色彩值,从而完成分割。图 1 是 akko 的右视点图像,图 2 是用均值偏移算法分割后的图像。



图 1 右视点图像



图 2 分割后的图像

由于传输和分割误差会导致区域噪声,我们采用 TLS 算法去除每个分割区域上的噪声。假设色彩相似的区域视差不会发生大的变化,根据文献[7]每个分割区域都可以用一个 3D 模型平面来建模:

$$\frac{1}{z} = a'x + b'y + c' \quad (4)$$

其中, (x, y) 是图像中点的位置, z 是该点在参考相机坐标下的深度值。假设 f 为焦距,令 λ 为两个摄像机光心的基线距离,则根据平行光轴双目成像几何原理,每点像素的视差 d 与深度值 z 的倒数成比例,即 $d = \frac{\lambda f}{z}$, 对某个分割区域中的点 $[x, y]$, 可得到

$$d = ax + by + c \quad (5)$$

其矩阵形式为

$$\mathbf{B} = \mathbf{A}[a, b, c]' \quad (6)$$

其中 $\mathbf{A}$ 中每一行为 $[x, y, 1]$, $[x, y]$ 为该分割区域中每个点的位置, $\mathbf{B}$ 中每行对应于该点的视差值,初始值采用解码端解码生成的稀疏视差。

然而,式(6)所示的模型两边存在误差。一方面,在解码端由于受到传输带来的噪声影响以及编码时误匹配的影响,使得到的视差与其真实值之间存在误差:

$$\mathbf{B}_2 = \mathbf{B} + \mathbf{B}_1 \quad (7)$$

其中, $\mathbf{B}$ 是真实视差, $\mathbf{B}_1$ 是由于噪声和误匹配所带来的误差, $\mathbf{B}_2$ 是解码端得到的实际视差。另一方面,在物体边界处同一个前景物体可能由于颜色差异大而被分割为不同的块,前景物体也可能由于与背景物体颜色相近而被分割在同一个块中,从而产生误差:

$$\mathbf{A}_2 = \mathbf{A} + \mathbf{A}_1 \quad (8)$$

其中, $\mathbf{A}_1$ 是误差项, $\mathbf{A}_2$ 是实际得到的分割块矩阵, $\mathbf{A}$ 是理想状态的分割块矩阵。综上所述,实际环境中式(7)就被修正为

$$\mathbf{B}_2 = \mathbf{A}_2[a, b, c]' \quad (9)$$

为了克服式(9)两边扰动误差的影响,本文用总体最小二乘(TLS)法来估计每个分割块上的 $[a, b, c]^T$ 向量,总体最小二乘的基本思想是,不仅用扰动向量 $B1$ 去干扰数据向量 B ,而且用扰动矩阵 $A1$ 来干扰数据矩阵 A ,以便校正 A 和 B 二者之间存在的扰动^[8]。进而得到该分割块每点的新视差。我们以解码端解码生成的稀疏视差为初值,利用 TLS 算法对初始稀疏场进行修正更新,即生成精确的稠密视差场。

令 a_i^T 是矩阵 $A2$ 的第 i 行, b_i 是矩阵 $B2$ 的第 i 个元素,则总体最小二乘解是指

$$\min_x \sum_{i=1}^n \frac{|a_i^T x - b_i|^2}{x^T x + 1} \quad (10)$$

的极小变化量,其中, $|a_i^T x - b_i| / (x^T x + 1)$ 是从点 $\begin{pmatrix} a_i \\ b_i \end{pmatrix} \in C^{n+1}$ (C^{n+1} 为 $n+1$ 维有理数空间)到子空间 P_x 内的最近点距离,子空间 P_x 定义为

$$P_x = \left\{ \begin{pmatrix} a_i \\ b_i \end{pmatrix} \mid a_i \in C^{n+1}, b_i \in C, b_i = x^T a_i \right\} \quad (11)$$

2 中间视的生成

通过上述步骤后,可以得到稠密化的视差图。稠密化的视差图建立了左右视点图像间匹配像素的位置关系。我们借鉴在文献[9]的方法,利用匹配像素的位置关系合成新视点。

设右图像为参考图像,从右到左的视差为 α^{RL} ,假设 α 为视角参数, α 为 0 表示右视点位置, α 为 1 表示左视点位置。不失一般性,假设在解码端只得到从右到左的视差,则在视差和虚拟的中间摄像机位置已知的条件下,中间视图 I^{RI} 就可以通过

$$I^{RI}(X + \alpha d^{RL}(X, Y), Y) = \alpha I_L(X + d^{RL}(X, Y), Y) + (1 - \alpha) I_R(X, Y) \quad (12)$$

得到。其中, (X, Y) 为任意点的像素坐标, I^{RI} 为生成的中间视图, I_R 为右视点图像, I_L 为左视点图像。

3 实验结果与分析

图 3 和图 4 分别是 akko 左图像和右图像,图 5 是我们在文献[3]中得到的稀疏视差场,图 6 是用本

文算法进行稠密化处理后的结果。比较两图可以看出,经过本文算法处理后,使原来的视差场更加平滑并且修正了一些误匹配区域,物体轮廓更突出,视差场得到明显改善。图 7 是直接用初始视差得到的中间视图,图 8 是用本文方法稠密化后的视差合成的中间图像。通过对比可发现,当解码端仅能获得稀疏视差场时,由于存在大量的误匹配,现有系统直接合成的中间视图(图 7)主观质量较差。图 9 是用文献[10]方法生成的稠密化视差场所得的中间视图,文献[10]是根据窗口内各个像素与待匹配点的关系来调整支持权重得到稠密视差。从合成的中间视图来看,本文算法相较于直接用稠密化视差场方法的主观效果并无明显差异。在算法复杂度方面,对比同样硬件环境下的算法时间复杂度,本文算法的主要时间消耗在图像分割上,用于图像分割的时间仅约为 2s,而文献[10]中主程序 LASW 算法的消耗时间约为 71s;另外本文算法的空间复杂度(内存开销)也远低于文献[10]的算法,因此本文方法可以在降低算法复杂度的基础上保持较好的主观效果。用本文算法作为一种后处理,则能从稀疏视差场合成主观质量良好的中间视图。

我们也将本文算法应用于 xmas 图像,图 10 是 xmas 左图像,图 11 是用本文方法合成的 xmas 图像



图 3 akko 左图



图 4 akko 右图

的中间视合成图 ($\alpha = 0.5$), 从主观质量上来看依据本文算法合成的中间视(图 11)具有良好的主观视觉效果。

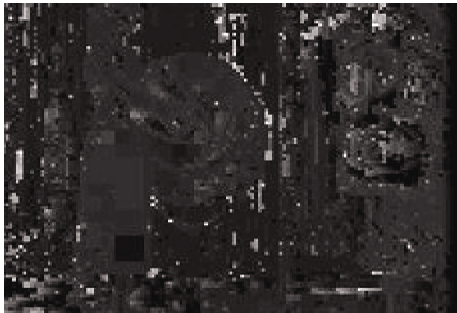


图 5 解码端稀疏的视差场



图 6 本文算法处理后的视差场



图 7 直接由初始视差得到的中间视图



图 8 本文方法合成的中间视图



图 9 文献[10]所生成的稠密化视差场所得的中间视图



图 10 xmas 左图像



图 11 $\alpha = 0.5$ 时生成的 xmas 中间视图

4 结 论

本文针对基线较小的平行相机系统,提出了一种新颖的解码端中间视合成的方法,该方法能仅利用解码端获得的稀疏视差场来合成高质量的中间视图。该方法对解码的稀疏视差场用均值偏移算法和总体最小二乘估计进行稠密化后处理,以获得稠密视差场,根据稠密视差场合成中间视图。作为一种后处理方法,本文算法仅需较小的传输带宽就能获得精确的视差图和主观效果良好的中间视图。下一步工作,我们将在此稠密视差的基础上利用多视点间的相关性和引入运动信息,来获得快捷的虚拟视点序列的绘制。

参考文献

- [1] Klaus A, Sormann M, Karner K. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In: Proceedings of the International Conference on Pattern Recognition, Hong Kong, China. 2006. 3: 15-18
- [2] Yang Q, Wang L, Yang R, et al. Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(3):492-504
- [3] 安平, 张兆扬. 立体图像分层交叠块视差估计与自适应补偿. 通信学报, 2003, 24(3): 1-7
- [4] Bleyer M, Gelautz M. Graph-based surface reconstruction from stereo pairs using image segmentation. In: Proceedings of the SPIE Symposium on Electronic Imaging, San Jose, CA, USA, 2005, 5665:288-299
- [5] Hong L, Chen G. Segment-based stereo matching using graph cuts. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington DC, USA, 2004, 1:74-81
- [6] Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(5):603-619
- [7] Tao H, Sawhney H S. Global matching criterion and color segmentation based stereo. In: Proceedings of the IEEE Workshop on Applications of Computer Vision, Palm Springs, CA, USA, 2000. 246-253
- [8] 张贤达. 矩阵分析与应用. 北京:清华大学出版社, 2004. 403-418
- [9] 吕朝辉, 袁惇. 基于视差估计的中间视合成. 光电子·激光, 2007, 18(7):855-858
- [10] Yoon K J, Kweon I S, 2006. Locally adaptive support weight approach for visual correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 4: 650-656

Generation of intermediate view based on clensification disparity map

Zhang Qian*, An Ping***, Zhang Zhaoyang**, Lv Chaohui***

(* School of Communication and Information Engineering, Shanghai University, Shanghai 200072)

(** Key Laboratory of Advanced Displays and System Application, Shanghai University, Shanghai 200072)

(*** School of Information Engineering, Communication University of China, Beijing 100024)

Abstract

In view of the fact that the existing 3D video processing systems can not synthesize higher quality intermediate view images unless they have a larger bandwidth, this paper proposes a new algorithm for synthesis of intermediate view images at the decoder of the systems. It can produce a synthesized intermediate view image with high quality only by a sparse disparity, which means only a small bandwidth is necessary. The algorithm uses the sparse disparity generated by the decoder as the initial value, and employs the meanshift algorithm and the total least squares (TLS) technique to post-process it to obtain a dense disparity at the decoder, thus synthesizing an intermediate view with higher quality from a sparse disparity. The experimental results verified the validity of the algorithm.

Key words: disparity, intermediate view, image segmentation, segmentation error, meanshift