

基于半规则网格的视差估计算法

吴书凯,都思丹,李 华

(南京大学 电子科学与工程系,南京 210093)

(wushukainju@163.com)

摘 要:立体视频中视图间的视差估计是压缩的关键技术。针对视差估计中传统网格法的缺点,提出一种使用半规则的网格来估计视差的方法。该方法综合了节点均匀分布的规则网格和节点贴合物体边界的非规则网格这两种方法,选择具有最高的梯度和的行和列的交点作为网格的节点,并且利用每个节点的梯度值进一步剔除在缺少纹理区域中的节点,最后在分辨率下混合块迭代和网格迭代对剩余的节点进行匹配。该算法剔除了匹配时易于出错的节点,同时也减少了需要估计的节点数。实验结果表明该方法改善了标准网格法的匹配质量,同时也降低了匹配的计算复杂度。

关键词:立体视频会议;半规则网格;块迭代;网格迭代;视差估计

中图分类号: TP391.41 **文献标志码:** A

Improved algorithm for estimating disparity using non-uniform mesh

WU Shu-kai, DU Si-dan, LI Hua

(Department of Electronic Science and Engineering, Nanjing University, Nanjing Jiangsu 210093, China)

Abstract: The stereo video needs compression before transmission. Besides eliminating the temporal redundancy as in the mono-channel video, it can further reduce the data by eliminating spatial redundancy between different channels, so the disparity estimation is the key technique for the compression of the stereo video. A new algorithm was proposed for estimating disparity vector using non-uniform mesh. It combined parts of the merits of the standard mesh and the mesh attached to object boundary, selected the rows and the columns with the highest sum of the gradient values, and used their cross points as node points, then further eliminated the points in the area lack of textures, at last it used both of block iteration and mesh iteration to match these nodes. This algorithm eliminates the points that are hard to match precisely, and reduces the node points at the same time. The experimental result shows that this algorithm improves the matching quality, and reduces the computational complexity.

Key words: stereo video conferencing; non-uniform mesh; block iteration; mesh iteration; disparity estimation

0 引言

传统的单通道视频通过去除帧内和帧间的相关性进行数据压缩,而立体视频一般由 2 个以上的视频通道组成,因此它的数据量比普通的单通道的视频要大得多,在存储或者传输之前有必要进行压缩。由于立体视频的左右两个视图有很强的相关性,所以压缩的关键在于如何消除这样的相关性,正如单通道中的前后帧之间的运动估计,估计左右视图的视差矢量(Disparity Vector, DV),然后利用这些矢量来去相关。对于运动估计,前后帧中有运动变化的点数较少,而在视差估计中,一幅视图中除了远离镜头的区域以外几乎所有的像素点都有非零 DV,并且在运动估计中使用较多的关于块匹配方法的一个假设——块中的像素点具有相同的 DV——在视差匹配中却不太适合,因此需要一些别的方法来估计 DV。

一般进行视差估计主要有双向块匹配法、网格法、贝叶斯法和基于对象的方法。双向块匹配方法实现简单,不过估计精度不高;贝叶斯法和基于对象的方法都有较为精确的估计结果,不过计算的复杂度都偏高;网格法则是通过在原图上建立网格,通过迭代收敛的方式得到整体的视差场,网格法一般有两种:一种是标准网格法,它的网格是均匀分布的;另一种

使用的是贴合于物体边界的网格。本文针对立体视频会议的实时要求,以及在一般视频会议中背景比较固定或者简单的特点,结合了两种网格法的优点,提出了一种基于半规则网格的视差估计算法,以达到去相关压缩数据的目的。

1 改进的网格法

1.1 视差匹配的约束条件

利用一些假设条件可以得到一些约束条件用来减小匹配的复杂度。本文中使用了以下几个约束条件。

1.1.1 外极约束

空间中一点在摄像机对中的投影一定在外极线上。因此在搜索匹配点的时候只要搜索在外极线上的点即可。对于像平面平行配置的两个摄像机,它们的外极线与水平扫描线重合,因此只需要在水平方向上搜索匹配点。本文中假设图像已经被校正成平行配置的情况。

1.1.2 单向性约束

对于平行摄像机配置,我们规定 DV 为 $d = x_l - x_r$,有 $d = FB/Z$,其中 F 为焦距, B 为基线长度, Z 为空间该点到镜头的距离。可见 d 总是大于等于零的,因此 $x_l > x_r$,所以由左视图对右视图进行匹配的时候,我们只需要向左搜索即可。

收稿日期:2007-10-13;修回日期:2007-12-04。

作者简介: 吴书凯(1983-),男,福建永安人,硕士研究生,主要研究方向:图像视频处理; 都思丹(1963-),女,江苏南京人,教授,博士生导师,主要研究方向:图像工程、模式识别; 李华(1982-),女,江苏南京人,硕士,主要研究方向:图像视频处理。

1.2 标准的网格法

网格法一般有三角网格和四角网格两种。三角网格的插值函数是仿射模型,而四角网格的插值函数是用双线性函数;三角网格有较好的插值效果,不过四角网格在确定元素的边界时候较三角网格运算更简单。因此本文中使用的是四角网格,这里只介绍四角网格^[1]。

网格法首先在一个基准视图中应用网格,并试图在另一个视图中找到对应的节点位置,而 4 个节点所组成的正方形内部的点的 DV 则由四个节点的 DV 通过双线性插值的方法得出:

$$d(x, y) = \sum_{i=1}^4 \phi_n d_n \quad (1)$$

其中 ϕ_n 为与该节点对应的插值系数^[3]。

如图 1 所示,匹配节点的时候通过最小化与该节点相关的四个正方形块的视差补偿预测 (Disparity Compensated Prediction, DCP) 来估计该节点的 DV。

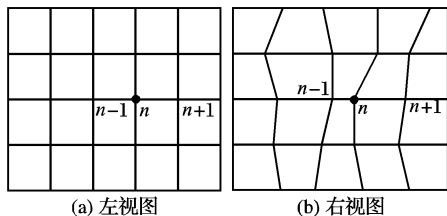


图 1 网格迭代示意图

1.3 改进的网格法

标准网格法的优点在于它估算出来的 DV 场是连续的,不过在有些情况下例如在边界上这种连续性并不是所希望的。特别是在标准网格法中在基准图像上应用网格的时候使用的是均匀分布的网格,如图 1 中所示。由于节点分布与图像内容基本无关,因此当边界出现在两个节点之间的时候就会产生模型误差。另外对于有大量单一背景存在的时候,背景上的许多节点是不必要的,增加了计算负担。而且由于背景上估计的不准确性,还会影响总体估计的效果。

解决办法之一是使用贴合物体边界的节点来使用网格。例如在视图中选择梯度最大的几个点,依据一定的法则来选择节点。不过这种方法也有一些缺陷。首先是计算复杂度高,由于这样取得的节点还是基本没有规则,所以建立网格的时候需要使用 Delaunay Triangulation 来建立三角网格,而且在迭代过程中计算迭代边界也需要进行乘除计算;其次由于立体图像中大量存在的 occlusion 的问题,有些节点可能根本就找不到正确的匹配点,而且由于如此定出的节点之间相互关系不太确定,因此可能造成严重的错误。

本文的方法是对这两种方法进行综合,采用类似于标准网格法中建立网格的方法,保持标准网格中节点的行列结构,依然是在视图中选择若干行和若干列,这些行列的交点作为节点,同时吸收了一些建立贴合物体边界的网格的概念,根据视图的梯度值来选择行和列的位置,这样可以用较少的节点提供较好的网格表述精确度,在模型精确度和性能之间取得一个较好的平衡点。

1.3.1 建立网格

依据外极约束,对于像平面平行配置的摄像机,在匹配的时候只在水平方向上进行搜索。一般来说节点匹配的准确度是和该节点水平方向上的梯度相关的^[6],因此采用水平方向上的梯度作为节点选择的主要依据。

1) 计算左视图中每个像素的垂直和水平梯度值 $grad_v$

与 $grad_h$ 。

2) 对每一行计算:

$$S_x(x) = \sum_{x_0 \in x} \sqrt{grad_v^2(x_0, y_0) + grad_h^2(x_0, y_0)}$$

对每一列计算: $S_y(y) = \sum_{y_0 \in y} grad_h^2(x_0, y_0)$ 。规定一个

上限 N ,并可以根据码率的要求来选择节点之间最小距离 k 。 k 越小,则网格的描述越精确,数据量和计算量越大。

3) $n = \arg \max(S_x(x))$,将 n 存入数组 $x1$, N 减 1,并且将 $S_x(n-k; n+k)$ 的值置零。对 $S_y(y)$ 进行相同的操作,将索引存入数组 $y1$ 。为了便于确定边界,将视图的左右边界也加入数组 $x1$,这样得到 $length(x1) \times length(y1)$ 个节点。

4) 如果 $N > 0$,并且 $\max(S_x(x))$ 或 $\max(S_y(y))$ 不为零,则返回 2)。

5) 除了左右边界上的点,对每个节点 $(x1(i), y1(j))$ 计算以该点为中心的方块(本文采用的是 5×5 方块)的平均梯度值:

$$\bar{g}(i, j) = \sum_{(il, jl) \in BLOCK} |grads(il, jl)| / sizeof(BLOCK) \quad (2)$$

其中 $grads(il, jl)$ 为 (il, jl) 点处的梯度值, $sizeof(BLOCK)$ 为该方块的像素数。如果 $\bar{g}(i, j)$ 大于一个预定义的值,则将节点加入数组 $node$,进一步抛弃无意义的点。

最后得到了一组存有节点数据的数组 $node$,作为下一段进行匹配的节点,如图 2 所示。

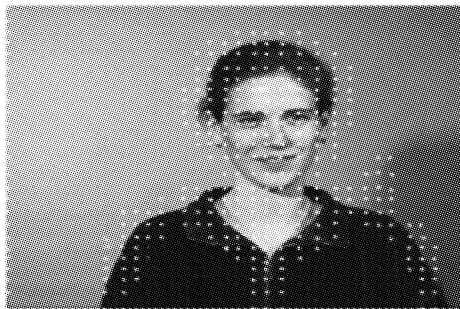


图 2 生成节点的位置

1.3.2 块迭代的预估计

标准网格的匹配方法其实就是寻找视差补偿预测 (Disparity Compensation Prediction, DCP) 误差的全局最小点的过程,由于标准的网格匹配法涉及大量的插值步骤,因此先使用块的方法进行迭代匹配搜索,以加速网格法的收敛。

考虑到单向性的约束,并且网格匹配时候节点应该保持其顺序,因此块匹配也使用迭代的方式。

1) 初始一个 DV 的数组 d ,如果 d 无定义则置零,其后用上次迭代的结果。初始一个 $i = 0$ 。

2) 取出 $node$ 数组中的第 i 个节点,如果是左右边界的节点则忽略,直到取到不在边界的节点,计算 DV 的边界: $dmin = node(i-1) + d(i-1) - node(i)$, $dmax = \min(node(i+1) + d(i+1) - node(i), 0)$ 。

3) 对于每个 $d_{test} \in [dmin, dmax]$, 计算 $E_{DCP}(d_{test}) = \sum_{(i, j) \in BLOCK} |I^l(i, j) - I^r(i + d_{test}, j)|$ 。

将 $d1 = \arg \min_{d_{test}} (E_{DCP})$ 作为该点在这次迭代中的值存入 d 。如果 $i < length(node)$,则返回 2)。

4) 如果迭代完一次后发现没有节点的 d 有变化,则退出循环,否则返回 2)。

块大小的选用会影响最后迭代收敛的效果,块太大会降低估计的精度,块取得太小则可能导致陷入局部最优。因此为了进一步提高精度,采用高斯金字塔,结合了算法复杂度和估计精度这两方面的考虑,选用2级的高斯金字塔。先在1/2分辨率下建立网格并进行块匹配;然后再在原分辨率下,采用相同大小的块,用上一级估计出的 d 值作为迭代的初始值继续重复以上的匹配步骤得到块匹配的估计值。

1.3.3 用网格迭代对块迭代估计值的修正

最后利用1.1节中介绍的标准网格的匹配方法进行最后的匹配。节点 (i, j) 匹配搜索的下限 $d_{\min} = \text{pos}(i-1, j) + d(i-1, j) - \text{pos}(i, j)$,上限 $d_{\max} = \text{pos}(i+1, j) + d(i+1, j) - \text{pos}(i, j)$ 。计算 E_{DCP} 取有着最小 E_{DCP} 的 d 值作为该节点的视差。

$$E_{DCP} = \sum_{n=1}^4 \sum_{(x,y) \in BLOCK_n} e(f^l(x, y) + f^r(x + d, y)) \quad (3)$$

$$d(i, j) = \arg \min_{d \in [d_{\min}, d_{\max}]} (E_{DCP})$$

其中 $f^l(x, y)$ 表示左视图, $f^r(x + d, y)$ 表示右视图。

迭代过程中记录下这次迭代的DCP,并与上一次记录下的DCP进行比较,如果新的DCP小于原来的,则保留这次的DCP,并更新DV;否则保留原来的DV。这样保证最后的网格迭代收敛。

2 实验结果及分析

本文对Gwen(图3)与Anne(图4)中的两个典型的视频会议中的场景进行了DV的估算,结果如图中所示。估算重建质量时候使用:

$$PSNR = 10 \times \lg(255^2 \times \text{NUMOFPOINT}(\text{IMAGE}) / \sum_{i,j} (\bar{I} - I)^2)$$

则Gwen的前景信噪比 $PSNR = 18.07$ dB, Anne的前景 $PSNR = 23.18$ dB。与其他方法的比较如表1中所示,其中标准网格法使用的节点数要大于本文中使用的节点数:在Anne这个场景(288×360)使用了 40×40 个节点,而本文方法只有 31×30 个节点,在Gwen(144×360)中,则是 23×59 与 20×32 。块匹配使用了最简单的逐点匹配的方法,Gwen中使用了 9×9 的块,Anne中使用了 16×16 的块。可以看出本文中的方法对于标准的方法有所改进,尤其在视差范围比较大的场景有比较大的改善。虽然块匹配的 $PSNR$ 较高,不过网格法不存在块效应的问题,具有较好的视觉上的恢复效果,而且速度方面也有比较大的优势。图5是标准网格法和本文提出方法的收敛曲线的对比。由于本文的方法去除了一些无意义的节点并且使用了块的方法进行了预估计,所以收敛速度要快得多。在Anne的场景中标准网格法迭代了16次才收敛,而本文方法只需要8次。表2是本文方法和标准网格法匹配速度上的比较(Matlab环境中,AMD 2800+, 512M内存)。另外视频序列中网格法可以为后续帧提供良好的有意义的初值,可以在后续帧中省略块迭代,只需进行少量的网格迭代即可,因此在视频序列中平均每帧的匹

配时间应该要远小于这个数值。

表1 半规则网格和标准网格法与块匹配法的对比

场景	PSNR/dB		
	半规则网格	标准网格法	块匹配法
Gwen	18.07	13.94	13.40
Anne	23.18	23.10	27.50

表2 本文方法和标准网格法匹配速度的比较

场景	匹配时间/s	
	本文方法	标准网格
Gwen	10.2484	41.0215
Anne	22.5783	68.0248

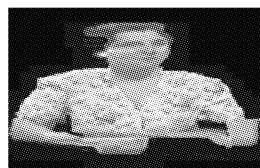
块迭代和网格迭代的混合使用也有效地改善了网格估计的效果。表3中网格迭代指用2层分辨率下迭代出的结果,可以看出混合使用块迭代和网格迭代的方法较仅使用网格迭代能够得到更好的预测结果,在Gwen中更有效地排除了次优点的干扰。

表3 混合法和仅用网格迭代的对比

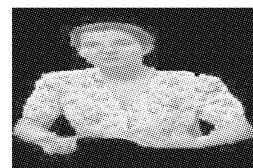
场景	PSNR/dB	
	混合法	网格迭代
Gwen	18.07	15.98
Anne	23.18	22.84



(a) 原始的左视图

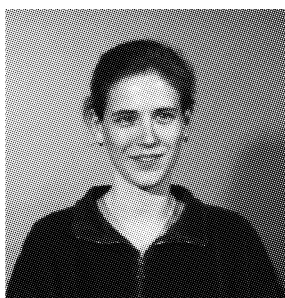


(b) 由原始右视图恢复出的左视图

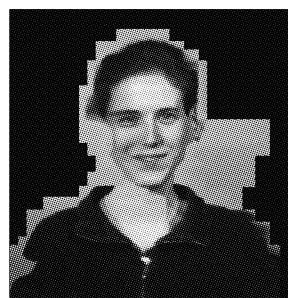


(c) 原始右视图

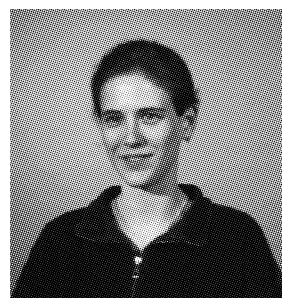
图3 “Gwen”的预测结果



(a) 原始的左视图



(b) 由原始右视图恢复出的左视图



(c) 原始右视图

图4 “Anne”的预测结果

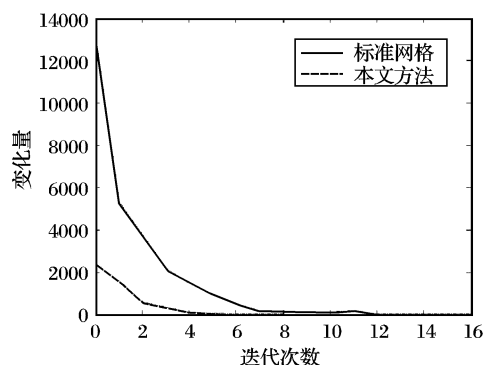


图5 标准网格法和半规则网格法的收敛曲线对比

(下转第965页)

例可以看到,采用本文提到的方法进行点云的精简,可以保留模型内部的特征区域,并且算法思想比较简单,实现起来比较容易,并且由于算法收敛速度快,简化所用时间较短。

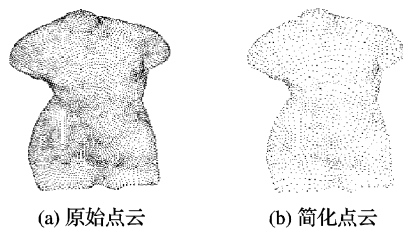


图5 模型3点模型简化结果

4 结语

本算法不需要对点云进行重构或三角化,而是根据点的特征直接对点云进行精简,减少了算法的时间和空间复杂度,更易于计算机操作。本算法不仅能快速地完成数据点云的精简,而且较好地保持了原有对象的特征。简化算法主要用在需要快速绘制复杂的点模型,对绘制质量要求不高的场合。但是,在实验中发现本文算法对部分模型的简化会造成部分区域的真空现象,个别边界区域的点数保留较少,因此,如何使处理后的点云保留更多的特征区域点并且不会形成区域真空,将是我们下一步工作的重点。

参考文献:

- [1] ALEXA M, BEHR J, COHEN-OR D, *et al.* Computing and rendering point set surfaces[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2003, 9(1): 3–15.
- [2] SCHAUFLEER G, JENSEN W H. Ray tracing point sampled geometry[C]// *Proceedings of the 11th Eurographics Workshop on Rendering*, ACM International Conference Proceeding Series. New York: ACM Press, 2000: 319–328.
- [3] PAULY M, KEISER R, KOBELT P L, *et al.* Shape modeling with point-sampled geometry[C]// *Computer Graphics Proceedings, Annual Conference Series*. Los Angeles, USA: ACM SIGGRAPH, 2003: 641–650.
- [4] OHTAKE Y, BELYAEV A, ALEXA M, *et al.* Multi-level partition of unity implicits[C]// *Computer Graphics Proceedings, Annual Conference Series*. Los Angeles, USA: ACM SIGGRAPH, 2003: 463–470.
- [5] ADAMS B, DUTRÉ P. Interactive boolean operations on surfel-bounded solids[C]// *Computer Graphics Proceedings, Annual Conference Series*. Los Angeles, USA: ACM SIGGRAPH, 2003: 26–31.
- [6] STAMMINGER M, DRETTAKIS G. Interactive sampling and rendering for complex and procedural geometry[C]// *Proceedings Eurographics Rendering Workshop*. Manchester: ACM, 2001: 151–162.
- [7] ZWICKER M, PFISTER H, BAAR J V, *et al.* Surface splatting[C]// *Computer Graphics Proceedings, Annual Conference Series*. Los Angeles, USA: ACM SIGGRAPH, 2001: 371–378.
- [8] PAULY M, GROSS M, KOBELT L. Efficient simplification of point-sampled surfaces[C]// *Proceedings of IEEE Visualization*. Washington, DC, USA: IEEE Computer Society, 2002: 163–170.
- [9] LINSSEN L. Point cloud representation[R]. Karlsruhe: University at Karlsruhe, Faculty for Computer Science, 2001.
- [10] OHTAKE Y, BELYAEV A, ALEXA M, *et al.* Multi-level partition of unity implicits[J]. *ACM Transactions on Graphics*, 2003, 22(3): 463–470.
- [11] FUKUNAGA K, HOSTETLER L. The estimation of the gradient of a density function, with applications in pattern recognition[J]. *IEEE Transactions on Information Theory*, 1975, 21(1): 32–40.
- [12] CHENG Y. Mean shift, mode seeking, and clustering[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, 17(8): 790–799.

(上接第959页)

3 结语

本文提出了一种基于特征的半规则分布的网格法,它利用图像的内容特征对标准网格的节点进行一定程度的调整,保留了标准网格法与贴合物体边界的网格法的一些优点。另外将块的迭代与网格法混合使用也有效地提升了匹配的速度。该方法在立体视频会议这样有着固定或者比较简单背景的场景下有着良好的估计性能。

本文提出的半规则的网格在模型上还是有一些缺陷,特别是在一些垂直方向的弧形的边界上还是可能带来一些模型误差,可以只保留本文中计算行位置的方法,而每行中的点可以用更灵活的方法来确定,以减小模型上的误差。在将来可以在序列中加入运动和视差的联合匹配,利用两个通道的多帧图像进行估计,更为精确地计算视差。

参考文献:

- [1] WANG RU-SHANG, WANG YANG. Multiview video sequence analysis, compression, and virtual viewpoint synthesis[J]. *IEEE Transactions on Circuits System for Video Technology*, 2000, 10(3): 397–410.
- [2] DUDON M, AVARO O, ROUX C. Triangular active mesh for motion estimation[J]. *Signal Processing: Image Communication*, 1997, 10(1/3): 21–41.
- [3] WANG Y, LEE O. Use of 2D deformable mesh structures for video-compression. Part I – The synthesis problem: Mesh based function approximation and mapping[J]. *IEEE Transactions on Circuits System for Video Technology*, 1996, 6(6): 636–646.
- [4] WANG YAO, OSTERMANN J, ZHANG YA-QIN. 视频处理与通信[M]. 侯正信, 杨喜, 王文全, 等译. 北京: 电子工业出版社, 2003: 130–137, 302–303.
- [5] CHIEN S-Y, YU S-H, DING L-F, *et al.* Fast disparity estimation algorithm for mesh-based stereo image/video compression with two-stage hybrid approach[C]// *Proceedings of 2003 SPIE International Conference on Visual Communications and Image Processing (VCIP 2003)*. Lausanne, Switzerland: SPIE, 2003: 1521–1530.
- [6] OHM J-R, IZQUIERDO E. An object-based system for stereoscopic videoconferencing with viewpoint adaptation[C]// *Proceedings European Symposium on Advanced Imaging and Network Technologies*. Berlin, German: SPIE, 1996, 2952: 29–40.
- [7] 韩军功, 卢朝阳. 基于 Delaunay 三角形网格的立体图像编码算法[J]. *计算机辅助设计与图形学学报*, 2004, 16(12): 1708–1712.