

一种基于立体视邻接帧时空相关性的 最小代价函数帧估计算法

骆 艳, 张兆扬

(上海大学通信学院电子信息系, 上海 200072)

摘 要: 为了在立体视频序列编码中获得高的压缩率, 需要对立体视频序列中一个视的序列按传统方法进行独立编码; 另一个视的序列中, 只对其中一些参考帧(I 帧或 P 帧) 按视差补偿预测的方法进行编码, 其余帧不进行编码和传输, 而在解码端用立体视帧估计的方法得到重建. 本文提出了一种基于立体视中邻接帧在图像、视差场和运动矢量场之间高度相关性的方法. 对于因遮挡而缺乏估计的区域, 则结合了图像强度的连续性和运动、视差矢量的分布特性, 构造了代价方程并估计出该部分的运动矢量及强度值. 实验证明, 重建出来的帧图像在视觉和信噪比意义上均具有较好的效果.

关键词: 立体视频; 遮挡; 帧估计

中图分类号: TN911. 1 **文献标识码:** A **文章编号:** 0372-2112 (2003) 10-1513-05

A Smallest Cost Function Frame Estimation Algorithm Based on Spatial Temporal Correlation in Neighboring Frames of Stereoscopic Sequences

LUO Yan, ZHANG Zhao2yang

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

Abstract: In order to achieve high compression ratio in coding stereo video sequences, one sequence of stereoscopic sequences is independently coded by a conventional MPEG2 type scheme, only a few reference frames in another sequence are coded. The rest of frames, however, are neither coded nor transmitted. They are reconstructed at the decoder using a stereoscopic frame estimation technique. This paper studies the highly spatial and temporal correlation between images, disparity fields and motion vector fields in the neighbor frames of stereoscopic sequences and proposes a frame estimation technique to explore these correlations. As to those un2est2mated regions due to occlusion, this paper also constructs a cost function combined with the continuity of image intensity and the distr2bution characteristic of motion and disparity vector to estimate the motion vectors and intensities of them. The experiment results show that this frame estimation algorithm has good results on view perception and PSNR.

Key words: stereoscopic; occlusion; frame estimation

1 引言

随着通信及信息技术的发展, 以平面图像描述的 2D 视频远未满足人们日益增长的对场景更真实、自然再现的需求, 因此基于 3D 视觉^[1, 2] 的立体/多视点视频技术越来越受到重视. 立体/多视点视频技术主要是利用人体的生理视觉特性, 即人的大脑可以对左右两只眼睛所捕获图像的微小差异进行处理并形成具有深度感的 3 维世界. 使用一个立体图像对, 可以为我们提供一种简捷的从左右两幅图像中获得真实世界相对深度感的方法. 但是与立体视频传输伴随的主要问题在于图像/视频传输与存储的数据量比普通的单通道图像/视频加倍. 这样大的带宽在实际应用中是不可接受也是不实际的. 行

之有效的解决数据量增加的方法之一是使用图像/视频压缩技术. 为适应新的发展需求, 新一代多媒体信号编码标准 MPEG- 4 和视频检索标准 MPEG7 也支持多视点并发数据流的编码以实现对象的环视和立体视频的传输. 因此, 与独立编码每一个视的序列相比, 为了使得压缩性能获得显著提高, 需要仔细考虑和利用在立体视序列中存在的附加特性, 并在图像质量与带宽之间进行合理平衡. 因此, 高效立体压缩编码技术在三维技术中起到重要作用. 在多媒体通信中, 3D 图像通信将成为一个重要的发展方向. 立体视频技术将在 3D 可视化、3D 远程医学、3D 可视通信、3D 高清晰度电视、3D 影院以及虚拟现实等众多领域得到广泛应用.

针对人眼在一眼给出清晰图像, 另一眼给出模糊图像时,

其综合感觉仍然是清晰并且具有三维观感的生理视觉特性,为获取高效的压缩,一种对立体视频序列编码的可取方法是对其中的一个视的序列按传统的 MPEG 方法进行独立编码,而对另一视的序列只对某些参考帧(I 帧或 P 帧)利用视差补偿进行预测编码,对其余的大部分帧不作编码传输,用帧估计的方法在解码端重建.这样,利用很小的附加带宽,可以传输和获得具有理想图像质量的立体视频序列^[1~4].因此帧估计技术在立体视频序列编码方案中十分重要.目前对这方面的研究主要是通过了解码端进行固定块匹配搜索完成^[4],但固定块匹配搜索方法不仅有着较大的视差遮挡估计误差,还具有明显的方块效应,同时在解码端进行匹配搜索十分耗时,这在实时解码时是难以接受的.

本文研究了邻接帧立体图像对视差矢量与运动矢量的关系,提出了一种将视差和运动进行联合估计的立体视目标帧合成估计方法.对于因视差遮挡(一个视中的点在另一个视中不可见)而造成帧估计失败的区域,提出了一种最小代价方程解决方法.节 2 给出邻接帧立体图像对视差矢量与运动矢量的关系,节 3 讨论可视区域的目标帧合成估计法,节 4 是对遮挡区域的最小代价方程处理法,且给出了对目标帧进行帧估计的具体算法,节 5 和节 6 分别是实验结果和总结.

2 邻接帧立体图像对视差矢量与运动矢量的关系

设一个立体透视投影模型的几何参数(基线距离 B , 摄像机焦距 f 及会聚角度 H) 已知,同一个三维空间的点在左右两个图像平面上均有对应的投影点(本文中称为可视区域的点).在 $n-k$ 时刻投影点坐标分别为 (X^L, Y^L, f) 和 (X^R, Y^R, f) ,其中 X 为投影平面中沿扫描线的横坐标, Y 为纵坐标,上标 L 和 R 表示左右图像平面.若该点到成像平面的深度为 z^L ,则两个投影点坐标具有如下关系式^[5,6]:

$$\begin{cases} X^R = f \frac{\cos H K^L + f \sin H H - B \cos(H/2)(f/z^L)}{-\sin H K^L + f \cos H H - B \sin(H/2)(f/z^L)} \\ Y^R = f \frac{Y^L}{-\sin H K^L + f \cos H H - B \sin(H/2)(f/z^L)} \end{cases} \quad (1)$$

当立体图像视频序列采用平行光轴立体摄像机获得时,则会聚角 $H=0$,式(1)可以简化为:

$$\begin{cases} X^R = X^L + (Bf)/z^L \\ Y^R = Y^L \end{cases} \quad (2)$$

令 $d_{n-k}^{LR} = X^L - X^R = (Bf)/z^L$,则 d_{n-k}^{LR} 为 $n-k$ 时左右图像间视差矢量的水平分量,而视差矢量的垂直分量为 0.

如果该 3D 点在 $n-k$ 到 n 的时段中存在运动位移,且该位移投影到左图像平面上的值为 $v^L = (v_{n,x}^L, v_{n,y}^L)$,则该 3D 点在左图像平面上的投影坐标为 $(X^L + v_{n,x}^L, Y^L + v_{n,y}^L, f)$.若新的深度值为 z^L ,则 3D 点在右平面上投影点新坐标为:

$$\begin{cases} X^R = X^L + v_{n,x}^L + (Bf)/z^L \\ Y^R = Y^L + v_{n,y}^L \end{cases} \quad (3)$$

此时 $d_n^{LR} = X^L - X^R = (Bf)/z^L$ 为在 n 时刻左图像到右图像视差矢量的水平分量.因此,该 3D 点在右图像平面上的二维运动矢量为:

$$\begin{aligned} \begin{pmatrix} v_{n,x}^R, v_{n,y}^R \end{pmatrix} &= \begin{pmatrix} X^R - X^R, Y^R - Y^R \end{pmatrix} \\ &= \begin{pmatrix} v_{n,x}^L + \frac{Bf}{z^L + z^L} (z^L - z^L), v_{n,y}^L \end{pmatrix} \\ &= \begin{pmatrix} v_{n,x}^L + d_{n-k}^{LR} - d_n^{LR}, v_{n,y}^L \end{pmatrix} \end{aligned} \quad (4)$$

从式(4)我们得到可视区域中点运动矢量与视差估计间的闭环关系:

$$(v_{n,x}^R + d_{n-k}^{LR}, v_{n,y}^R) = (v_{n,x}^L + d_n^{LR}, v_{n,y}^L) \quad (5)$$

由于实际的立体视频序列中视差矢量变化十分缓慢^[7,8],因此对于小的时间间隔 k 有:

$$d_n^{LR} = d_{n-k}^{LR}, v_n^R = (v_{n,x}^L, v_{n,y}^L) = v_n^L \quad (6)$$

这表明在满足精确平行光轴立体摄像机几何情况下(会聚角度 $H=0$),对于非遮挡区域中的点,在小的时间间隔 k 内,深度变化十分缓慢,相对于运动矢量的变化来说可以忽略.则对于立体视频序列,非遮挡区域中的 3D 点在左视频流中的运动矢量与右视频流运动矢量相等; n 时刻得到的视差矢量与 $n-k$ 时刻得到的视差矢量相等.

3 可视区域的目标帧合成估计法

设已知的 3 帧参考图像分别为左视第 $n-k$ 帧以及右视的 $n-k$ 帧,待估计的目标帧图像为右视的第 n 帧.设左视第 n 帧位于可视区域点 (x, y) 处的某像素点为 p_n^L ,其强度值为 $I_n^L(x, y)$.根据式(5), (6),则 p_n^L 在左视第 $n-k$ 帧和右视 $n-k$ 帧图像平面上的对应点强度值分别为 $I_{n-k}^L(x - v_{n,x}^L, y - v_{n,y}^L)$ 和 $I_{n-k}^R(x - v_{n,x}^L - d_{n-k}^{LR}, y - v_{n,y}^L)$.如果 n 到 $n-k$ 帧没有外界光照变化,并且左右摄像机参数一致的情况下,则像素点在这 4 个邻接帧的强度值之间应有如下等式:

$$\begin{aligned} I_n^L(x, y) &= I_{n-k}^L(x - v_{n,x}^L, y - v_{n,y}^L) \\ &= I_{n-k}^R(x - v_{n,x}^L - d_{n-k}^{LR}, y - v_{n,y}^L) \\ &= I_{n-k}^R(x - d_n^{LR}, y) \end{aligned} \quad (7)$$

由于右视的第 n 帧未知,与其对应的运动和视差矢量也未知,但可以根据式(6),用 d_n^{LR} 取代 d_{n-k}^{LR} ,且 $v_n^R = v_n^L$,这样 p_n^L 在右视的第 n 帧图像上所处的位置、运动矢量及视差矢量均可直接由式(7)确定,这意味着对于立体视频序列中任意非遮挡区域的像素点 p_n^L ,我们只要得到两个立体对中的其中 3 幅图像,就可以确定第 4 幅图像的所有信息.考虑到成像噪音,摄像机参数等因素,本文中强度值选取右视第 $n-k$ 帧像素点的值作为参考值:

$$I_{n-k}^R(x - d_{n-k}^{LR}, y) = I_{n-k}^R(x - v_{n,x}^L - d_{n-k}^{LR}, y - v_{n,y}^L) \quad (8)$$

这样根据式(5)的闭环关系,就可得到右视第 n 帧中对可视区域部分的像素值以及运动和视差矢量的估计值.图 1 的 a_2, b_2 分别是对图 1 中 a_1, b_1 所示的 2 种立体视频序列完成可视区域帧估计后的结果.图中高亮度区域为在视差估计中检测到的因遮挡而导致缺乏匹配的区域.

4 遮挡区域的处理

对于视差遮挡区域,上述式(1)~(8)的关系是不成立的.这是由于在右视图像中可见而在左视图像中被遮挡的区域,在左视图像序列中找不到对应点,因而这部分的运动矢量不

能从左视图序列中确定,从而第 n 帧像素点所处的位置和像素值也不能确定。但是我们可以根据图像像素强度的平滑变化和运动矢量的时空连续特性,利用相邻可视区域已获得的运动矢量来估计这部分的运动矢量,从而获得视差遮挡区域强度及运动矢量的估计值。问题是如何在这些众多的参考运动矢量中选择最优。本文将这种求解最优值的方法转换为构造基于多个特征的匹配代价方程并求取该代价方程最小值。当一个参考运动矢量成为遮挡点运动矢量的可信度越高,则它的代价方程的值就越小,具有最小代价值的参考运动矢量即为最优的运动矢量值。

设位于右 n 帧平面坐标 (x, y) 处的一个遮挡点 P_n^R | occlusion, 该点到右 $n-1$ 帧对应点的真实运动矢量为 (v_x, v_y) , 则按照运动矢量进行逆向映射可得到该点的估计值 $I_n^R(x, y) = I_{n-k}^R(x - v_x, y - v_y)$ 。设 (x, y) 邻域中某个位于可视区域的像素点,利用前节的方法可以得到该像素点坐标 $(x+h, y+1)$, 运动矢量 (v_x, v_y) , 视差值 d^{LR} 以及像素值 $I_n^R(x+h, y+1)$, 则 (v_x, v_y) 为估计 (v_x, v_y) 的参考运动矢量。

引入第一个实际观察: 图像中运动和视差矢量几乎处处平滑变化^[8]。这样,我们可以在 (x, y) 邻域的可视区域中找到 (v_x, v_y) 的估计值。

首先假定匹配窗口内像素 $(x+h, y+1)$ 处的运动矢量服从如下的条件分布模型:

$$v \sim N(0, R_d \sqrt{h^2 + 1^2}) \quad (9)$$

其中 R_d 代表运动变化程度的常量。该模型说明: 像素在 $(x+h, y+1)$ 处的运动期望值等于像素 (x, y) 处的运动期望值,但他们差值之方差却随着二者之间的距离增大而增大。随着像素点之间距离的增加,运动矢量的空间相关性也随之减弱,则运动矢量 (v_x, v_y) 成为真实运动矢量的可信度也随之下降。

当选定 (v_x, v_y) 成为真实运动矢量后,可得到右 n 帧点的估计值:

$$I_n^R(x, y) = I_{n-k}^R(x - v_x, y - v_y) \quad (10)$$

根据立体摄像机成像的几何特性,当自然场景中的物体深度发生变化,投影到图像中为同一扫描行上的视差值发生从大到小的跃变时产生视差遮挡,因此视差遮挡是全局分布并且主要分布于物体垂直边沿。根据这种视差分布特性,在遮挡区域左右两边的参考运动和视差矢量往往分属于不同物体表面,因而也具有不同的运动矢量(视差)值。为了提高参考运动矢量的可信度,引入第二个实际观察: 邻域中具有相似强度的像素点趋向于属于同一个物体表面,因而也趋向于具有相同的运动和视差矢量^[8]。

进一步假设右 n 帧中 (x, y) 处的估计值 $I_n^R(x, y)$ 与点 $(x+h, y+1)$ 的图像强度之间差分对运动矢量可信度的影响服从如下统计模型:

$$I_n^R(x+h, y+1) - I_n^R(x, y) = I_{n-k}^R(x+h, y+1) - I_{n-k}^R(x - v_x + h, y - v_y + 1) \sim N(0, R_t) \quad (11)$$

R_t 是表示右 $n-1$ 帧邻域内图像强度局部方向波动程度的常量。该模型说明: (v_x, v_y) 处像素图像强度的期望值等于 (x, y) 处强度期望值,该假设的不确定性也随着像素间的强

度值差分增大而增大。

最后引入第三个实际观察: 立体视中视差遮挡区域主要属于自然场景中的远景区域,表现在图像中为视差矢量值较小的物体表面^[9, 10]。因此点 $(x+h, y+1)$ 处视差值 d^{LR} 对运动矢量可信度的影响服从如下统计模型:

$$d^{LR} \sim N(0, R_0) \quad (12)$$

R_0 表示视差变化程度的常量。该模型说明参考点 $(x+h, y+1)$ 处视差越大,则该点运动矢量的可信度随之下降。

为了在众多参考矢量中选择最优,使用贝叶斯决策检验。根据(9), (11), (12)的三个统计概率模型,以及图像强度与运动矢量分布相互统计独立的假设,可以得到贝叶斯分类器的决策函数由三部分特征组成:

$$C = L_1 d_1 + L_2 d_2 + L_3 d_3 \quad (13)$$

式中, C 为接受假设的平均代价, $L_i, i = 1 \sim 3$ 为加权系数, $d_i, i = 1 \sim 3$ 表示如下:

$$\begin{cases} d_1(v) = \frac{1}{\sqrt{2\pi} R_d \sqrt{h^2 + 1^2}} \exp \left[-\frac{(v - 0)^2}{2R_d^2(h^2 + 1^2)} \right] \\ d_2(v) = \frac{1}{\sqrt{2\pi} R_t} \exp \left[-\frac{(I_n^R(x+h, y+1) - I_{n-k}^R(x+h - v_x, y+1 - v_y))^2}{2R_t^2} \right] \\ d_3(v) = \frac{1}{\sqrt{2\pi} R_0} \exp \left[-\frac{(d^{LR})^2}{2R_0^2} \right] \end{cases} \quad (14)$$

当选定一个候选运动矢量时, $(v_x, v_y) = (v_x, v_y)$, 此时 $v = 0$ 。

考虑到高斯函数的指数形式,对式(14)采用自然对数形式来表达,并省略掉其中的常数项后,得到最终的代价函数表达式:

$$C = L_1 \left(I_n^R(x+h, y+1) - I_{n-k}^R(x+h - v_x, y+1 - v_y) \right)^2 + L_2 \ln \left[\sqrt{2\pi} R_d \sqrt{h^2 + 1^2} \right] + L_3 (d^{LR})^2 \quad (15)$$

通过式(15)我们估计遮挡点相邻可视区域多个参考运动矢量的代价函数,从中选择具有最小代价的运动矢量和像素值作为估计值。考虑到遮挡区域是沿物体垂直边缘分布的,从下往上地进行遮挡区域的估计,已完成估计的下一行像素点参与上一行的估计,选择位于遮挡区域左右两边的运动矢量作为候选值。

图1中 a_3, b_3 为利用最小代价函数对遮挡区域处理后得到最终重建结果,可以看出该方法基本准确地恢复了遮挡的内容,尤其对那些容易出现误匹配的,处于物体边界的像素值及运动矢量值的判断和选取上具有理想效果,另一方面对于处于同一物体内部的可视区域到遮挡区域的过渡上基本保持了图像强度本身的连续性,使得重建质量与 a_1, b_1 相近。

5 实验结果及说明

节3与4中仿真实验中所使用的两种立体视频序列^[11]见图1中 a_1, b_1 。其中 Booksale(90 帧/视)和 Crowd(180 帧/视)为自然场景序列,图像大小为 320×240 ,帧到帧之间具有多个

物体的运动,在 Booksale 序列中还出现了摄像机的快速移动,而 Crowd 序列中摄像机移动比较缓慢,但是场景中既有近景物体也有远景物体,视差变化范围较大。

实验中,为快速获得一个能自动检测视差遮挡的视差场,视差估计采用了动态规划及最大似然代价函数方法^[9,10]。编

码帧结构为 IBBP, 在 I(P) 与 P 之间有两个 B 帧, 以左图像流为参考帧流, 右图像流为目标帧流。利用本文进行帧估计时, 选择与其时间间隔最近的参考帧以及运动(视差)矢量分别进行前向或后向帧估计和内插。式(15)中的 $L_1BL_2BL_3 = (1 \sim 10)B(10 \sim 50)B1$ 。



图1 立体视频重建及比较

a_1 : Booksale 原始图像(右视第 58 帧); b_1 : Crowd 原始图像(右视第 121 帧); a_2 ; b_2 : 完成可视区域帧估计后的图像(高亮度区域为遮挡区域); a_3 ; b_3 : 处理遮挡区后帧估计法重建图像; a_4 ; b_4 : 使用固定块匹配搜索法得到的重建图像。

为进一步验证本文方法,本文还与传统的基于固定块匹配搜索方法进行了实验比较。图 1 中的 a_4 ; b_4 为使用固定块匹配方法后得到最终重建结果,与 a_3 ; b_3 相比,可以看出图像具有较为明显的方块效应⁰,且本文方法得到图像的视觉效果明显好于固定块匹配方法。图 2 显示了用这两种方法获得的每个重建 B 帧与原始 B 帧的峰值信噪比的比较。从实验结果看来,在平均峰值信噪比意义上,本文方法比固定块匹配搜

索方法提高了 0.8~1.94dB。另一方面,从图 2 曲线可以看出,用本文方法获得的重建 B 帧受帧到帧之间的时间相关性变化影响较小,表现为 PSNR/帧序号曲线比较平坦;而固定块匹配搜索方法则对帧间变化比较敏感,PSNR/帧序号曲线起伏较大。这是因为固定块匹配搜索法假定了图像中物体与场景的运动均为匀速运动,这在出现变速运动时帧估计失败,导致 PSNR 迅速下降;而本文方法利用了与被重建 B 帧相对应的左视序列的运动矢量,可以很好地抑制上述问题的产生,尤其是对 Booksale 序列这种帧间变化较大的序列,可以获得比固定块匹配搜索方法好的多的重建结果。

6 结论

在帧间可视区域,我们可以利用立体视频运动和视差估计的闭环关系,用帧估计合成法直接获得这部分内容的估计;而对遮挡区域的运动矢量则使用邻域区域的运动矢量来近似,实验结果证明,本文中利用特征代价函数判断最优矢量的方法具有理想的效果。还应指出,本文的帧估计算法不仅比固定块匹配搜索法所重建的图像质量要好,而且可以直接利用解码端获得的运动和视差矢量值进行恢复和重建,无需进行匹配搜索,具有高压缩率编码和快速解码的特点。

参考文献:

- [1] Sriram Sethuraman. Stereoscopic image sequence compression Using multiresolution and quadtree decomposition based disparity and motion adaptive segmentation [D]. Ph. D Thesis, Dept of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburg PA (1996).

图2 用两种方法进行立体视频帧估计的结果比较

- [2] Puri, R V Kollarits, B G Haskell. Stereoscopic Video Compression Using Temporal Scalability [A]. SPIE Conf. Visual Communications and Image Processing (VCIP. 95) [C]. vol SPIE22501. 1995. 745- 756.
- [3] P D Gunatilake, M W Siegel, A G Jordan. Compression of Stereo Video Streams [A]. International Workshop on HDTV. 93 [C]. Ottawa, Ontario, CANADA, Elsevier, Oct 1993.
- [4] Qin Jiang and Monson H. Hayes. Stereo sequence coding [J]. SPIE 3309: 277- 786X/ 97.
- [5] D Tzovaras, N Grammalidis, M G Strintzis. Objectbased coding of stereo image sequences using joint 3D motion/ disparity compensation [J]. IEEE Trans. on Circuits and System for Video Technology 1997, 7 (2): 312- 328.
- [6] J Liu, R Skerjanc. Stereo and motion correspondence in a sequence of stereo images [J] Signal Processing: Image Commun, 1993, 5: 305- 318.
- [7] R Szeliski. A MultiView Approach to Motion and Stereo [OL]. Technical Report MSR2TR29219, Microsoft Research, May 1999. <http://www.research.microsoft.com/pubs/>
- [8] M Ebroul Izquierdo. Stereo image analysis for multiviewpoint telepresence applications [J]. Signal Processing: Image Communication 1998, 11: 231- 254.
- [9] I J Cox, S L Hingorani, S B Rao. A maximum likelihood stereo algorithm [J]. in Computer Vision and Image Understanding, 1996, 63(3): 542- 567.
- [10] I J Cox, S Hingorani, B M Maggs, S B Rao. Stereo without regularization

[R]. Tech. Rep., NEC Research Institute, Princeton, USA, October 1992.

- [11] DISTIMA stereoscopic test sequences, generated and distributed by CCETT [S]. France under the RACE2DISTIMA European project, October 1994.

作者简介:



骆 艳 女, 1974 年 11 月生于贵州省安顺市, 1999 年获清华大学工学学士学位, 现为上海大学通信与信息工程学院博士生, 主要研究领域为视频信号处理、三维图像编码及其应用。



张兆扬 男, 1938 年 9 月生于浙江省海宁, 上海大学通信与信息工程学院教授、博士生导师、学科带头人, 主要研究领域为数字视频处理、多媒体通信和计算机视觉。