

多视点视频编码中的视频目标提取与视差匹配

朱仲杰^{1,2}, 蒋刚毅^{1,3}, 郁梅^{1,3}, 吴训威^{1,2}

(1. 宁波大学电路与系统研究所, 浙江宁波 315211; 2. 浙江大学信息与电子工程系, 浙江杭州 310027;
3. 西安电子科技大学 ISN 国家重点实验室, 陕西西安 710071)

摘 要: 在单通道视频目标分割算法的基础上, 首先提出一种多视点视频编码中的视频目标提取方案. 然后提出一种基于多级块匹配视差估计算法. 应用 Moravec 算子对视差矢量图进行检测, 提取特征视差矢量并对其进行精确度检验. 若特征视差矢量不够精确, 则其所在块将重新进行小范围的块匹配. 若块匹配误差仍太大则将其分裂为四个子块重新匹配. 与传统块匹配算法相比, 新方法能获得较为精确和平滑的视差场, 提高重建后的图像质量.

关键词: 多视点视频编码; 视频目标提取; 视差估计

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2004) 05-0848-03

Video Object Extraction and Disparity Estimation in Multi-View Video Coding

ZHU Zhong-jie^{1,2}, JIANG Gang-yi^{1,3}, YU Mei^{1,3}, WU Xun-wei^{1,2}

(1. Institute of Circuits and Systems, Ningbo University, Ningbo, Zhejiang 315211, China;
2. Dept. of Information and Electronics Engineering, Zhejiang University, Hangzhou, Zhejiang 310027, China;
3. National Key Laboratory of ISN, Xidian University, Xian, Shaanxi 710071, China)

Abstract: Based on algorithm for extracting video objects in one view-channel, a new scheme of extracting video objects in multi-view coding is proposed. A disparity estimation algorithm based on multi-level block matching is proposed. Moravec operator is used to find out characteristic disparity vectors in disparity images for precision confirmation. If a characteristic disparity vector is not precise enough, the block it belongs to will be matched again around the area indicated by the vector. If a block is still not matched precisely, it is split into four equal sub-blocks for block matching. Compared with traditional block-matching based algorithm, more precise and smooth disparity field can be acquired with the proposed algorithm, so that the quality of reconstructed image can be greatly improved.

Key words: multi-view video coding; video object extraction; disparity estimation

1 引言

交互式多媒体的出现使人们不再是完全被动地接收和欣赏视听节目, 它提供了对视听对象的编辑功能, 人们能够根据自己的喜好去选择和操作视听对象, 可以人工合成视听场景^[1]. 多视点视频是一种新型的具有立体感和交互操作功能的视频^[2,3], 它由多个摄像机拍摄得到, 其数据量随着摄像机的增加而成倍增加, 所以在存储和传输时必须进行高效的编码压缩^[4]. 在多视点视频编码中, 视频目标提取和视差估计是一种关键技术^[5,6]. 传统的块匹配视差估计算法由于算法简单、计算量小而被普遍采用. 然而, 由于块内部所有像素赋予相同的视差矢量, 匹配块选择过大会使视差估计的结果不够精确, 匹配块选择过小, 又会降低视差估计的可靠度. 本文在我们先前提出的单通道视频目标分割算法和立体视频目标分割算法的基础上^[7], 首先给出了多视点视频编码中的视频目标提取方案. 然后提出了一种多级块匹配视差估计算法. 先采

用较大的匹配块根据 MAD 准则进行匹配, 再将 MAD 值大于给定阈值的图像块分成较小的块进行匹配, 直到所有匹配块的 MAD 值都小于给定阈值. 每一级匹配完成后, 用 Moravec 算子对视差矢量进行检测, 提取具有特征值的视差矢量, 对特征视差矢量再用校验准则进行校验, 以保证提取的视差矢量的可靠性. 实验结果说明了本文算法是有效的.

2 多视点视频编码中的视频目标提取方案

多视点视频中有多个视频通道, 在进行视频目标提取时每个视频通道中的视频目标都要进行分割. 一种简单的方案是各自独立地提取每一个视频通道中的视频目标, 但这种方案没有利用相邻通道视频序列之间的相关性, 提取的效率低, 计算量大. 本文采用另外一种较为简单的方案, 首先提取一个通道中的视频目标, 然后基于边界轮廓的视差匹配算法提取相邻下一视频通道中的视频目标. 这种方案充分利用了相邻视频通道之间的相关性, 计算量小, 提取效率高. 其算法流程

收稿日期: 2003-04-04; 修回日期: 2003-10-30

基金项目: 浙江省青年人才基金 (No. RC01057); 西安电子科技大学 ISN 国家重点实验室开放基金; 宁波市重点博士基金 (No. 01J20300-05, No. 2003A61001); 教育部留学回国人员科研基金

如图 1 所示,图中通道 0 和通道 1 是两个相邻的视频通道。

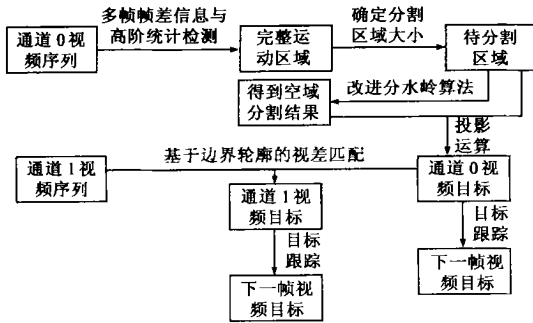


图 1 多视点视频编码中的视频目标提取方案

3 基于多级块匹配算法的视差估计算法

3.1 MAD 匹配准则与 Moravec 算子

常用的块匹配准则有很多,本文采用绝对平均误差函数(MAD),MAD 准则具有较小的计算量和较好的匹配性能,其计算公式如下:

$$MAD(i, j) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |f_0^{(t)}(x+m, y+n) - f_1^{(t)}(x+m+i, y+n+j)| \quad (1)$$

式中, M 和 N 为匹配窗口的尺寸, $f_0^{(t)}$ 和 $f_1^{(t)}$ 分别表示在 t 时刻通道 0 和通道 1 图像的灰度值。

Moravec 算子常用于象素的灰度特征值检测,其基本思想是:以像元的四个主要方向上最小灰度方差表示该像元与邻近像元的灰度变化情况,即像元的兴趣值,然后在图像的局部选择具有最大的兴趣值的点作为特征点,具体算法如下:

(1) 计算各像元的兴趣值 IV (Interest value), 例如计算像元 (x, y) 的兴趣值,先在以像元 (x, y) 为中心的 $N \times N$ 的窗口中,计算四个主要方向相邻像元灰度差的平方和:

$$V_1 = \sum_{i=-k}^{k-1} (f(x+i, y) - f(x+i+1, y))^2 \quad (2)$$

$$V_2 = \sum_{i=-k}^{k-1} (f(x+i, y+i) - f(x+i+1, y+1))^2$$

$$V_3 = \sum_{i=-k}^{k-1} (f(x, y+i) - f(x, y+i+1))^2 \quad (3)$$

$$V_4 = \sum_{i=-k}^{k-1} (f(x+i, y-i) - f(x+i+1, y-i+1))^2$$

其中 $k = \text{INT}(N/2)$. 取其中最小者为像元 (x, y) 的兴趣值:

$$IV_{x, y} = \min\{V_1, V_2, V_3, V_4\} \quad (4)$$

(2) 根据给定的阈值,选择兴趣值大于该阈值的点作为特征点的候选点. 阈值的选择应使候选点中包括需要的特征点,而又不含过多的非特征点。

(3) 在候选点中选取局部极大值点作为需要的特征点. 在一定大小的窗口内(可不同于兴趣值计算窗口),去掉所有不是最大兴趣值的候选点,只留下兴趣值最大者,该象素即为一个特征点。

3.2 多级块匹配视差估计算法

本文提出的视差估计算法采用多级块匹配思想,并应用

Moravec 算子对视差矢量进行检测,提取具有特征值的视差矢量,再对特征视差矢量进行校验,以期能最终获得较为精确的视差信息,其算法步骤如下:

(1) 设图像尺寸为 $W \times H$, I_0 和 I_1 分别为通道 0 和通道 1 中的图像, l 为匹配块的级别,初始时 $l = 0$. N_l 为图像块的尺寸,其中 $N_0 = 32$.

(2) 将图像 I_0 分成 $N_l \times N_l$ 图像块,每个图像块据 MAD 匹配准则在给定搜索窗口进行视差匹配,搜索它在 I_1 中的最佳匹配块,得到一组运动矢量 $\{V_{xN_l}[i], V_{yN_l}[i]\} (i = 0, 1, \dots, S - 1)$ 和运动矢量所对应的 $MAD_{N_l}[i] (i = 0, 1, \dots, S - 1)$, 其中, $S = (W \times H) / (N_l \times N_l)$, S 为整数. $MAD_{N_l}[i]$ 由下式给出:

$$MAD_{N_l}[i] = \frac{1}{N_l \times N_l} \sum_{m=0}^{N_l-1} \sum_{n=0}^{N_l-1} |f_0^{(t)}(x_{iN_l} + m, y_{iN_l} + n) - f_1^{(t)}(x_{iN_l} + m + V_{xN_l}[i], y_{iN_l} + n + V_{yN_l}[i])| \quad (5)$$

(3) 如果 $l = 1$, 执行该步骤, 否则跳过. 检查每个图像块的标志, 如果为 1 则以 $(V_{xN_{l-1}}[i], V_{yN_{l-1}}[i])$ 为初始搜索点, 在一个小范围内进行搜索, 得到 $\{V_{xN_l}[i], V_{yN_l}[i]\}$, 并计算相应的 $MAD_{N_l}[i]$.

(4) 把 $\{V_{xN_l}[i]\}$ 和 $\{V_{yN_l}[i]\}$ 看作两幅行和列分别为 H/N_l 和 W/N_l 的视差矢量图像. 对这两幅图像应用 Moravec 算子进行特征检测, 找出具有特征信息的视差矢量作为特征视差矢量. 设 $T_{xN_l}[i]$ 和 $T_{yN_l}[i]$ 分别表示与 $\{V_{xN_l}[i]\}$ 和 $\{V_{yN_l}[i]\}$ 相对应的特征信息, 由下式给出:

$$T_{xN_l}[i] = \begin{cases} 0, & V_{xN_l}[i] \text{ 是非特征点} \\ 1, & V_{xN_l}[i] \text{ 是特征点} \end{cases} \quad (6)$$

$$T_{yN_l}[i] = \begin{cases} 0, & V_{yN_l}[i] \text{ 是非特征点} \\ 1, & V_{yN_l}[i] \text{ 是特征点} \end{cases}$$

进一步定义 $T_{N_l}[i]$ 表示与图像块 i 相对应的视差矢量的特征信息, 取值为:

$$T_{N_l}[i] = \begin{cases} 0, & T_{xN_l}[i] + T_{yN_l}[i] = 0 \\ 1, & \text{其他} \end{cases} \quad (7)$$

(5) 对所有 $T_{N_l}[i] = 1$ 图像块 i 进行视差校验. 设图像块 i 在 I_1 中最佳匹配块为 k , 对图像块 k 进行视差匹配, 寻找它在 I_0 中匹配块, 得一视差矢量 $(V_{xN_l}[k], V_{yN_l}[k])$. 对 $(V_{xN_l}[i], V_{yN_l}[i])$ 和 $(V_{xN_l}[k], V_{yN_l}[k])$ 按下式进行视差校验:

$$((V_{xN_l}[i] - V_{xN_l}[k])^2 + (V_{yN_l}[i] - V_{yN_l}[k])^2)^{1/2} \leq 2 \quad (8)$$

如果上式成立, 则认为图像块 i 得到了可靠的视差匹配, 否则认为图像块 i 匹配不可靠。

(6) 如果图像块 i 匹配不可靠, 则进行重新搜索, 重新搜索是利用前面的视差矢量确定一个初始搜索值, 进行一个小范围的搜索, 设 (T_{0x}, T_{0y}) 表示初始搜索值, 由下式给出:

$$T_{0x} = (V_{xN_l}[i-1] + V_{xN_l}[i-W/N_l-1] + V_{xN_l}[i-W/N_l])/3 \quad (9)$$

$$T_{0y} = (V_{yN_l}[i-1] + V_{yN_l}[i-H/N_l-1] + V_{yN_l}[i-H/N_l])/3 \quad (10)$$

(7) 检查所有 $MAD_{N_l}[i]$ ($i = 0, 1, \dots, S-1$), 给定阈值 MAD_0 , 如果 $MAD_{N_l}[i] > MAD_0$, 则认为图像块 i 匹配较精确, 否则认为匹配不够精确. 其中, MAD_0 的选取可由实验确定.

(8) 如果所有的图像块都认为是精确匹配或者 $l = 4$, 则算法结束. 否则将匹配不够精确的图像块平分成四个子图像块, 并给每个子图像块分配标志号 1, 对匹配精确的图像块子图像块分配标志号 0, $l = l + 1$, $N_l = N_l / 2$, 返回(3).

4 实验结果

为了检验新算法的有效性, 本文对具有两个通道的视频图像 sergio 进行了仿真测试, 图像尺寸为 512×512 . 首先, 基于单通道视频目标提取算法提取通道 0 第 6 帧中的视频目标, 根据轮廓跟踪匹配算法提取通道 1 第 6 帧中相应的视频目标, 并应用跟踪算法跟踪提取通道 0 和通道 1 中的后续视频目标. 然后应用本文提出的多级块匹配视差估计算法进行视差估计仿真测试. 图 2 和图 3 给出了部分实验结果, 其中图 2(a) 和图 2(b) 分别是通道 0 和通道 1 第 6 帧视频原图像, 图 2(c) 和图 2(d) 分别为分割出的通道 0 和通道 1 第 6 帧视频目标, 图 2(e) 为应用传统块匹配算法得到的视差信息进行重建所得通道 0 第 6 帧视频目标与背景合成后的图像, 其 PSNR 为 31.46dB, 图 2(f) 为应用本文算法所得视差信息重建的



图 2 sergio 视频序列实验结果(第 6 帧)

道 0 第 6 帧视频目标与背景合成后的图像, 其 PSNR 为 35.97dB. 图 3(a) 和图 3(b) 是跟踪得到的通道 0 和通道 1 的第 10 帧图像的视频目标, 图 3(c) 为应用传统块匹配算法得到的视差信息进行重建所得通道 0 第 10 帧视频目标与背景合成后的图像, 其 PSNR 为 31.20dB, 图 3(d) 为应用本文算法所得视差信息重建的通道 0 第 10 帧视频目标与背景合成后的图像, 其 PSNR 为 36.13dB. 实验结果显示, 本文算法重建图像的信噪比明显提高. 从主观视觉质量看, 采用本文算法重建图像的衬衫纹理较为清晰, 没有发生断列, 接近于原图像, 而应用传统块匹配算法重建图像的衬衫纹理则出现明显的断列现象.



图 3 sergio 视频序列实验结果(第 10 帧)

5 结论

多视点视频是一种具有立体感和交互操作功能的新型视频, 基于对象的多视点视频编码使用户能编辑和操作视频场景, 从而实现真正的人机交互. 本文提出的基于 MPEG4 的多视点视频编码中的视频目标提取方案, 能有效地提取多个视频通道中的视频目标. 所提出的基于多级块匹配视差估计算法将 MAD 值大于给定阈值的图像块分成较小的块进行多级匹配, 直到所有匹配块的 MAD 值都小于给定阈值, 以保证视差估计精度. 每一级匹配完后, 对应用 Moravec 算子提取的具有特征值的视差矢量进行校验, 以保证视差矢量的可靠性. 与传统块匹配算法相比, 本文算法可以获得较为精确和平滑的视差场, 能够显著提高重建后的图像质量.

参考文献:

- [1] Naemura T, Kaneko M, Harashima H. Compression and representation of 3-D images[J]. IEICE Trans INF & SYST, 1999, E82-D(3): 558 - 565.
- [2] Srinatzis M G, Malassiotis S. Object-based coding of stereoscopic and 3D image sequences[J]. IEEE Signal Processing Magazine, 1999, 16(3): 14 - 28.
- [3] Ohm J R. Encoding and reconstruction of multiview video objects[J]. IEEE Signal Processing Magazine, 1999, 12(7): 47 - 54.
- [4] Pedersini F, Sarti A, Tubaro S. Multi-camera systems[J]. IEEE Signal Processing Magazine, 1999, 16(3): 55 - 65.
- [5] Lin CW, Fei EY, Chen YC. Hierarchical disparity estimation in stereo sequence[J]. IEEE Transactions on Consumer Electronics, 1998, 44(3): 630 - 637.
- [6] Grammalidis N, Srinatzis M G. Disparity and occlusion estimation in multi-ocular systems and their coding for the communication of multi-view image sequences[J]. IEEE Trans on Circuits and Syst. Video Technol, 1998, 8(3): 328 - 344.
- [7] 朱仲杰, 蒋刚毅, 郁梅, 吴训威. 目标基视频编码中的运动目标提取与跟踪算法[J]. 电子学报, 2003, 31(9): 1426 - 1428.

作者简介:



朱仲杰 男, 1976 年生于安徽安庆, 浙江大学与宁波大学联合培养博士研究生, 主要研究领域为多视点视频编码、无线视频编码与通信、视频水印与信息安全等.



蒋刚毅 男, 1964 年生, 博士, 教授, 中科院计算所兼职博士生导师, 主要研究领域为图像处理与视频信号编码、多媒体信息传输与信息安全、基于视觉的智能控制.