

立体视频编码中的目标分割与立体匹配算法*

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

¹(浙江大学 信息与电子工程系, 浙江 杭州 310027)

²(宁波大学 电路与系统研究所, 浙江 宁波 315211)

³(西安电子科技大学 ISN 国家重点实验室, 陕西 西安 710071)

New Algorithms for Object Segmentation and Stereo Correspondence in Stereo Video Coding

ZHU Zhong-Jie^{1,2}, YU Mei^{2,3}, JIANG Gang-Yi^{2,3+}, WU Xun-Wei^{1,2}

¹(Department of Information and Electronics Engineering, Zhejiang University, Hangzhou 310027, China)

²(Institute of Circuits and Systems, Ningbo University, Ningbo 315211, China)

³(National Key Laboratory of ISN, Xidian University, Xi'an 710071, China)

+ Corresponding author: E-mail: jianggangyi@sina.com

<http://www.nbu.edu.cn>

Received 2002-12-25; Accepted 2003-01-28

Zhu ZJ, Yu M, Jiang GY, Wu XW. New algorithms for object segmentation and stereo correspondence in stereo video coding. *Journal of Software*, 2003,14(11):1971~1976.

<http://www.jos.org.cn/1000-9825/14/1971.htm>

Abstract: Video object segmentation and stereo correspondence is a key technology in object-based stereo video coding. In this paper, based on an algorithm for extracting video object in one view-channel, a new contour tracking and correspondence algorithm is first proposed for extracting stereo video object pairs. Then, a disparity estimation algorithm based on video objects is proposed. It takes a characteristic pixel as a correspondence primitive and the disparity estimation is implemented with disparity constraints. The disadvantages of discontinuous and imprecise disparities with the traditional block-based algorithms can be improved and a precise and smooth disparity field can be finally acquired.

Key words: stereo video coding; video object segmentation; disparity correspondence

摘要: 视频目标分割与立体匹配是目标基立体视频编码中的核心技术。首先在单通道视频目标分割的基础上提出一种轮廓跟踪匹配算法,以提取立体视频目标对,然后提出一种基于目标的视差估计算法。它以具有特征

*Supported by the Project-Sponsored by SRF for ROCS, SEM under Grant No.2001331 (教育部留学归国人员科研基金); the Natural Science Foundation of Zhejiang Province of China under Grant Nos.RC01057, 601017 (浙江省自然科学基金); the Open Project Foundation of National Key Laboratory of ISN of Xidian of China under Grant No.H02025 (西安电子科技大学 ISN 国家重点实验室开放基金); the Key Doctor Foundation of Ningbo of China under Grant No.01J20300-05 (宁波市重点博士基金)

第一作者简介: 朱仲杰(1976 -),男,安徽安庆人,博士生,主要研究领域为多视域视频图像编码,无线视频编码与通信,视频水印与信息安全。

信息的像素为匹配基元,结合视差匹配约束进行视差估计,克服了传统块匹配算法视差场不连续、视差精度差的缺点,可以获得较为精确和平滑的视差场。

关键词: 立体视频编码;视频目标分割;视差匹配

中图法分类号: TP391 文献标识码: A

从黑白到彩色,从模拟到数字,再到今天的高清晰度数字视频,视频技术的发展速度极其迅猛.进入 21 世纪,人们的视觉要求将是具有立体感、有着交互操作功能的视听产品.立体视频与单目视频相比,增加了景物深度信息,能够增强视觉效果的真实感和逼真度.它一般是由两个位置略有差别的摄像机同步拍摄得到^[1].在基于对象和内容的第二代视频编码 MPEG-4 标准中,为了支持基于内容的交互性,引入了视频对象的概念.视频序列的每一帧都被分割成形状任意的视频对象面(video object plane,简称 VOP),每个对象面可能覆盖图像中特定的感兴趣区域^[2].立体视频与单目视频相比,至少要处理翻番的数据量,需要进行高效的编码压缩。

立体视频目标分割和视差估计是目标基立体视频编码中的一项关键技术.在立体视频编码中,左右两个通道中的视频目标均需进行分割.这两个视频目标被称为一个视频目标对,它们之间存在很强的相关性^[3].视差信息在编码压缩、中间视域生成中等都要用到.要进行高效、高质量的立体视频编码必须有精确的视差信息.视频目标分割和视差信息的好坏直接关系到编解码后的视频图像的视觉效果.本文的研究重点是如何进行简单、快速的立体视频目标分割和精确的视差估计.首先根据我们先前提出的单通道视频目标分割算法提取一个通道中的视频目标,然后提出一种基于目标轮廓的跟踪匹配算法提取另外一个通道中的视频目标.本文提出一种基于目标的视差匹配算法.它以具有特征信息的像素为匹配基元,采用交叉校验准则进行校验,最后得到具有精确视差信息的视差场。

1 立体视频摄像系统

立体摄像系统一般由左右两个位置略有差别但参数相同的摄像机组成,根据像机光轴之间的关系可分为平行像机系统和交叉像机系统.平行像机系统两个像机的光轴是平行的,而交叉像机系统两个像机的光轴之间有一个微小的角度。

设 C_w 代表世界坐标系, C_l 和 C_r 是左右像机的坐标系, 设 $[R_l]$ 和 t_l (或 $[R_r]$ 与 t_r) 分别代表旋转矩阵和变换矢量, 则左右像机的坐标, $\bar{X}_l = [X_l, Y_l, Z_l]^T$ 和 $\bar{X}_r = [X_r, Y_r, Z_r]^T$, 与世界坐标 $\bar{X} = [X, Y, Z]^T$ 的关系为^[4]

$$\bar{X}_l = [R_l] \bar{X} + t_l; \bar{X}_r = [R_r] \bar{X} + t_r. \quad (1)$$

最常见的像机配置系统是平行像机系统,其左右两个像机的图像平面是平行的.两个像机之间的距离称为基线距离,用 B 表示.一般人眼双目间距离为 62~76mm,当 B 接近这个值时,这样的像机配置就可以模仿人眼的双目成像系统.将世界坐标系的原点设在两像机的中点,并假定两像机的焦距 (F) 相同.对于平行像机系统可以得到:

$$X_l = X + B/2, X_r = X - B/2, Y_l = Y_r = Y, Z_l = Z_r = Z. \quad (2)$$

$$x_l = F \frac{X + B/2}{Z}, x_r = F \frac{X - B/2}{Z}, y_l = y_r = y = F \frac{Y}{Z}. \quad (3)$$

此时,视差矢量仅有水平分量,它是深度的函数:

$$d_{(R \rightarrow L)x}(X, Y, Z) = x_l - x_r = \frac{FB}{Z}. \quad (4)$$

$$d_{(L \rightarrow R)x}(X, Y, Z) = x_r - x_l = -\frac{FB}{Z}. \quad (5)$$

其中, $d_{(L \rightarrow R)x}(X, Y, Z)$ 和 $d_{(R \rightarrow L)x}(X, Y, Z)$ 分别表示以左图像和右图像中的像素为参考点时的视差。

2 立体视频目标对分割

立体视频序列分为左右两个通道,相应的视频目标提取算法需要提取视频目标对中的左右两个视频目标。

提取视频目标对的一种简单方案是从左右两个视频通道中分别各自独立地提取左右两个视频目标.但这种方案没有利用左右通道视频序列之间的相关性,提取的效率低,计算量大,不可取.本文所提方案是先从两个通道中的一个提取出运动目标,然后利用两个通道视频序列之间的相关性,根据轮廓跟踪匹配提取出另一个通道中的视频目标.这种方案计算量小,效率高.单通道视频目标提取算法可以参考我们先前的研究论文^[5],本文不作详细介绍,其算法流程如图1所示.首先,根据多帧帧差信息和高阶统计检测方法,去除噪声,提取运动区域.然后,采用改进分水岭算法进行空域分割.为了减少计算量和算法的复杂度,我们采用有限区域分割的思想,即空域分割并不是对整帧图像进行,而只是对运动区域及其附近部分进行.在空域分割和运动区域的提取完成之后,对二者进行投影运算以提取视频目标.

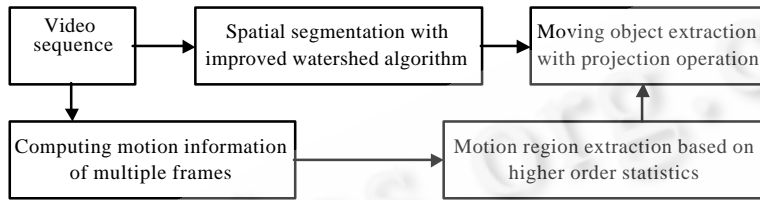


Fig.1 Flowchart of video object extraction algorithm of single view-channel

图1 单通道视频目标提取算法流程图

左通道视频目标提取后,右通道视频目标的提取不再需要进行运动区域检测、空域分割等这些复杂的步骤.由于左右两幅图像是左右两个摄像机在同一时刻对同一场景进行拍摄得到的,只是摄像机的视角有微小差异,所以两幅图像之间存在很强的相关性.因此右通道视频目标的分割就可以利用这种相关性来简化分割.本文采用二级视差匹配算法来提取右通道中的视频目标,视差匹配主要是对视频目标的轮廓进行.对于水平相机配置系统可以只考虑水平方向的视差,垂直方向的视差可以视为零.本文提出的右通道视频目标轮廓跟踪匹配算法步骤如下:

(1) 根据提取的左通道视频目标,提取其边缘轮廓.

(2) 对左视频目标的边缘轮廓进行视差匹配,寻找其在右图像中的匹配轮廓.轮廓匹配采用二级匹配算法.其中心思想是,对每一个边缘点进行匹配时,先采用较大的匹配块,在一定的窗口内搜索,得到一个初始视差值.然后将此视差值作为初始搜索位置,再采用一个较小的匹配块进行匹配,得到最后的较为精确的匹配点.

(3) 视差匹配后得到的右图像的边缘轮廓通常是不连续的,需要进行后处理,将所有边缘点连接起来,得到一个连续光滑的边缘轮廓.本文采用形态处理的方法进行处理.

(4) 将得到的轮廓进行填充,得到完整的视频目标二值掩模图像.

(5) 由二值掩模图像和右图像提取右通道视频目标.

这里,右通道视频目标提取算法非常简单,计算量很小.

3 基于目标的视差匹配算法

视差估计与运动估计类似,都是为了寻找两幅图像之间的相关性,运动估计是寻找两幅图像之间的时域相关性,而视差估计是寻找立体图像对中左右两幅图像之间的空域相关性.传统的视差匹配算法采用的是块匹配的方法.块匹配方法算法简单,但是很难得到连续、精确的视差场,而连续、精确的视差场是进行高效、高质量立体视频编码的重要保证.本文提出的基于目标的视差匹配算法仅对视频目标进行,对背景区域不作视差估计,这样会减少视差匹配的计算量和复杂度.事实上,在电视会议等虚拟会议系统中,背景一般都较为简单且一般不会改变,视差一般也是固定不变的.而精确、平滑的视频目标视差信息则很重要,它是进行虚拟场景合成和构建中间视角图像的关键信息^[5-7].本文的视差匹配算法步骤如下:

(1) 首先,对分割后的左右原始视频目标图像进行抽样.抽样的目标是为了减少计算量,设 $f_L(x, y)$ 和 $f_R(x, y)$ 分别为左右原图像, $S_L(x, y)$ 和 $S_R(x, y)$ 分别为相应抽样后的图像.二者的关系如下:

$$S_R(x, y) = f_R(2x, 2y), x = 0, 1, \dots, \lfloor row/2 \rfloor, y = 0, 1, \dots, \lfloor col/2 \rfloor, \quad (6)$$

其中, row 和 col 分别为原始图像的行列数, 符号 $\lfloor x \rfloor$ 表示不大于 x 的最大整数.

(2) 令 $S_{OR}(x, y)$ 表示 $S_R(x, y)$ 中属于视频目标部分的像素区域. 对 $S_{OR}(x, y)$ 进行特征检测, 将像素分为特征点和非特征点, 为了较少计算量, 采用如下的简易特征判别准则:

$$T(x, y) = \begin{cases} 1, & \frac{1}{N_w^2} \sum_{(p,q) \in w(x,y)} (S_{OR}(p, q) - \bar{S}_{OR}(x, y))^2 \geq T_0 \\ 0, & \text{其他} \end{cases} \quad (7)$$

$$\bar{S}_{OR}(x, y) = \frac{1}{N_w^2} \sum_{(p,q) \in w(x,y)} S_{OR}(p, q) \quad (8)$$

其中, $T(x, y)$ 为像素点 (x, y) 的特征值, $T(x, y) = 1$ 表示像素点 (x, y) 为特征点, 否则为非特征点, T_0 为检测阈值. $w(x, y)$ 表示以 (x, y) 为中心的检测窗, N_w 为窗口尺寸.

(3) 对特征点进行视差估计, 寻找它在 $S_{OL}(x, y)$ 中的匹配点. 设 $\{S_{ORi}(x_i, y_i)\} (i = 0, 1, \dots, M)$ 表示所有特征点集合, M 为特征点的总数; 设 $S_{OLi}(x_i, y_i)$ 表示点 $S_{ORi}(x_i, y_i)$ 的匹配点, $D_{(R \rightarrow L)i}(x_i, y_i)$ 表示相应 $R \rightarrow L$ 的视差矢量.

(4) 对 $S_{OL}(x, y)$ 中的 $S_{OLi}(x_i, y_i)$ 进行视差估计, 得到相应的 $L \rightarrow R$ 视差矢量 $D_{(L \rightarrow R)i}(x_i, y_i)$.

(5) $D_{(L \rightarrow R)i}(x_i, y_i)$ 和 $D_{(R \rightarrow L)i}(x_i, y_i)$ 按如下规则进行校验:

$$\|D_{(L \rightarrow R)i}(x_i, y_i) + D_{(R \rightarrow L)i}(x_i, y_i)\| \leq 2 \quad (9)$$

如果式(9)成立, 则认为第 i 个特征点是准确匹配的.

(6) 重复(3)-(5), 找出 $\{S_{ORi}(x_i, y_i)\}$ 中所有满足式(9)的 $S_{ORi}(x_i, y_i)$, 记为 $S_{ORd}(x, y)$, 相应的视差记为 $D_{(R \rightarrow L)d}(x, y)$. 这样有下式成立:

$$S_{OR}(x, y) = S_{ORd}(x, y) + S_{OR\bar{d}}(x, y) \quad (10)$$

式中 $S_{ORd}(x, y)$ 表示已经进行准确匹配的像素区域, $S_{OR\bar{d}}(x, y)$ 则表示还没有进行准确匹配的像素区域.

(7) 由 $D_{(R \rightarrow L)d}(x, y)$ 对 $S_{OR\bar{d}}(x, y)$ 中的像素进行视差插值, 得到完整的视差图 $D_{R \rightarrow L}(x, y)$.

视差匹配搜索时可以利用视差估计的约束条件进行快速搜索, 特别是采用平行像机系统时, 搜索可以只在水平方向进行, 垂直方向的视差可以不予考虑^[8,9].

4 实验结果

为了检验新算法的有效性, 本文应用立体图像测试序列 Sergio 对新算法进行了计算机仿真, 该序列具有简单静止背景, 图像尺寸为 512×512 . 首先, 基于单通道视频目标提取算法提取左图像中的视频目标, 并根据轮廓跟踪匹配算法提取右图像中相应的视频目标. 然后, 根据提出的基于目标的视差匹配算法对视频目标区域进行视差估计, 得到了较为精确和光滑的视差场. 图 2 和图 3 给出了部分实验结果. 其中, 图 2(a) 是左原图像, 图 2(b) 为最终提取的左视频目标, 图 2(c) 为左视频目标的边缘轮廓, 图 2(d) 为轮廓匹配得到的右视频目标边缘轮廓, 图 2(e) 为右原图像, 图 2(f) 为最终提取的右视频目标, 图 2(g) 为应用本文提出的目标视差匹配算法得到的 $D_{R \rightarrow L}$ 视差图; 图 2(h) 为应用传统块匹配算法得到的视差图. 比较而言, 图 2(g) 比图 2(h) 更为平滑. 图 3(a) 为右原图像, 图 3(b) 为应用块匹配算法得到的视差信息进行重建所得右视频目标与背景合成后的图像, 其 PSNR(峰值信噪比) 为 32.3780dB, 图 3(c) 为应用本文算法所得视差信息重建的右视频目标与背景合成后的图像, 其 PSNR 为 36.4130dB. 可见本文算法重建图像的信噪比明显提高. 此外, 从局部放大图明显看出, 应用本文算法所得重建图像的衬衫纹理基本接近原图像, 而采用块匹配算法所得图像的衬衫纹理出现断裂现象, 主观视觉效果较差. 实验结果显示, 本文算法是有效的.

5 结论

在基于 MPEG-4 的立体视频编码中, 立体视频目标分割是编码的前提; 而精确、平滑的视差场是进行虚拟场景合成和构建中间视角图像的关键信息, 是进行高效、高质量立体视频编码的关键. 本文首先根据我们先前提出的单通道视频目标分割算法提取一个通道中的视频目标, 接着利用左右视频目标之间的相关性, 根据我们

提出的目标轮廓跟踪匹配算法提取另外一个通道中的视频目标,然后,提出一种基于目标的视差匹配算法,它以具有特征信息的像素为匹配基元,采用交叉检验准则进行校验,得到最终具有精确视差信息的视差场.最后对本文算法进行了模拟测试,并给出了部分实验结果.应用本文算法得到的重建后的图像信噪比明显提高,实验结果说明了本文提出的算法是有效的.



Fig.2 Experimental results of stereo video sequence Sergio (the sixth image pair)

图 2 Sergio 立体图像序列的实验结果(第 6 图像对)

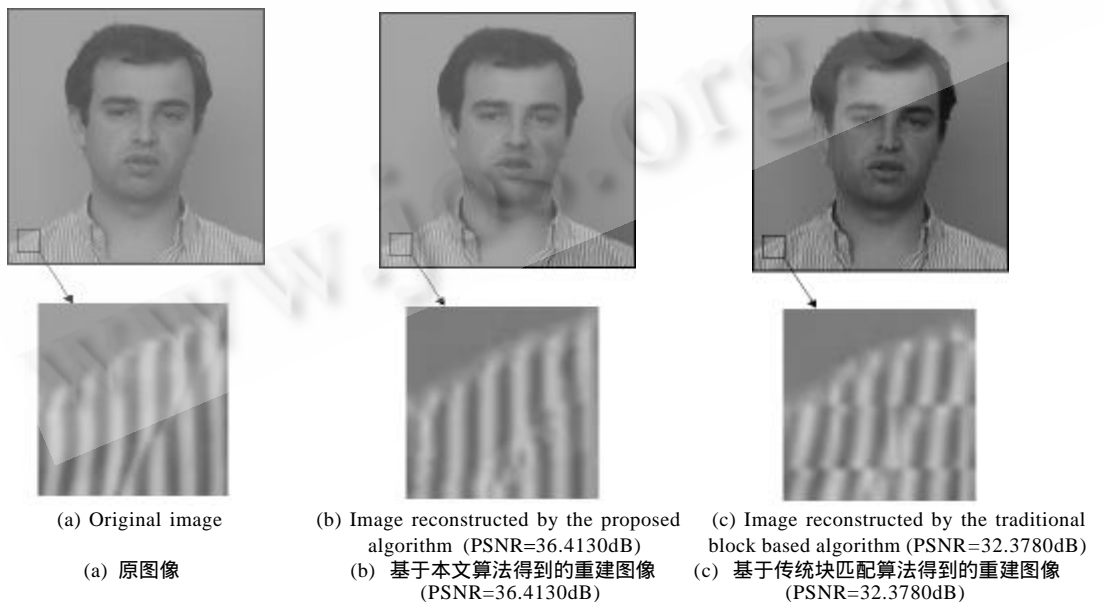


Fig.3 Reconstruction images of the proposed algorithm and the traditional block based algorithm

图 3 基于本文算法与传统块匹配算法重建的图像

References:

- [1] Naemura T, Kaneko M, Harashima H. Compression and representation of 3-D images. *IEICE Transactions on INF.& SYST*, 1999, E82-D(3):558~565.
- [2] Strintzis MG, Malassiotis S. Object-Based coding of stereoscopic and 3D image sequences. *IEEE Signal Processing Magazine*, 1999,16(3):14~28.
- [3] Ohm JR. Encoding and reconstruction of multiview video objects. *IEEE Signal Processing Magazine*, 1999,12(7):47~54.
- [4] Pedersini F, Sarti A, Tubaro S. Multi-Camera systems. *IEEE Signal Processing Magazine*, 1999,16(3):55~65.
- [5] Zhu ZJ, Jiang GY, Yu M, Wu XW. New algorithm for extracting moving object based on spatio-temporal information. *Acta Electronica Sinica*. 2003,31(9):1426~1428 (in Chinese with English abstract).
- [6] Lin CW, Fei EY, Chen YC. Hierarchical disparity estimation in stereo sequence. *IEEE Transactions on Consumer Electronics*, 1998, 44(3):630~637.
- [7] Grammalidis N, Strintzis MG. Disparity and occlusion estimation in multi-ocular systems and their coding for the communication of multiview image sequences. *IEEE Transactions on Circuits and System Video Technology*, 1998,8(3):328~344.
- [8] Jeong H, Oh Y. Fast stereo matching using constraints in discrete space. *IEICE Transactions on Information and Systems*, 2000,E83-D(7):1592~1600.
- [9] Zhu ZJ, Jiang GY, Yu M, Wu XW. Fast disparity estimation algorithm for stereoscopic image sequence coding. In: Yuan BZ, Tang XF, eds. *Proceedings of the IEEE Region 10 Technical Conference on Computers, Communications, Control and Power Engineer (TENCON 2002)*. Beijing: IEEE Press and POSTS&TELECOM Press, 2002. 28~31.

附中文参考文献:

- [5] 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.电子学报,2003,31(9):1426~1428.