

基于压缩域信息的立体视频对象分割方法

张倩¹⁾ 张兆杨^{1),2)} 安平^{1),2)}

¹⁾(上海大学通信与信息工程学院,上海 200072)

²⁾(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)

摘要 立体视频对象分割是交互式多视点视频应用的关键技术。为了提高对象分割的时效性和精确性,提出了一种利用压缩域视差和运动信息的立体视频对象分割算法。该算法首先对运动矢量场和视差场进行提取和修正处理,然后对视频帧进行分割作为初始值,最后用均值偏移算法聚类得到最终的对象分割结果。实验结果表明,对于纹理复杂的场景有很好的分割效果,可以获得与语义一致的对象。

关键词 立体视频对象 运动矢量场 视差 均值偏移

中图法分类号: TP391. 41 **文献标识码**: A **文章编号**: 1006-8961(2008)10-2047-04

Stereo Video Object Segmentation Based on Compressed Domain Information

ZHANG Qian¹⁾, ZHANG Zhao-yang^{1),2)}, AN Ping^{1),2)}

¹⁾(School of Communication and Information Engineering, Shanghai University, Shanghai 200072)

²⁾(Key Laboratory of Advanced Displays and System Application, Shanghai University, Shanghai 200072)

Abstract Stereo video object segmentation is a key technology in the multi-view video interacting. In this paper, based on the compressed domain information a novel stereo video object segmentation algorithm is proposed to improve the validity and accuracy of the segmentation. Firstly, we extract motion vectors and disparity field to segment the initial video object. Then we used the mean-shift algorithm to cluster the video object. Experimental results shows that the proposed algorithm could obtain the object with semantic conformity from the scene with complex texture.

Keywords stereo video, motion vectors, disparity, mean-shift

1 引言

随着多媒体技术的发展,人们对多媒体信息的需求越来越大,不再满足于对多媒体信息的简单浏览,而提出了基于对象的交互等需求,因此视频对象分割便应运而生。由于视频检索和视频监控等应用逐步从2D向3D发展和从像素域向压缩域发展,压缩域3D立体视频对象的分割也随之受到关注。在现有的单视点分割算法中^[1,2],视频对象分割的精

度往往在很大程度上受限于单视序列的可用信息量。而对于立体视频的对象分割可以利用单视点视频所没有的视差信息,视差信息能够较准确的保留物体内部的连续性及边缘的特点,对于提高对象分割的准确性十分有利。视频对象分割往往是非常耗时的,而压缩域视频对象分割由于避免了求取运动矢量场的过程从而大大降低了分割的复杂度^[3]。在单通道(2维)视频对象分割中已提出一些基于压缩域的方法,但由于从压缩域提取的运动场的稀疏性和不准确性,导致其视频对象的分割精度远远低

基金项目:国家自然科学基金项目(60572127,60672052);上海市曙光计划项目(06SG43);上海市经委产学研示范工程(07-040C);新型显示教育部重点实验室开放性课题资助项目(07k04)

收稿日期:2008-07-11;**改回日期**:2008-07-24

第一作者简介:张倩(1983~),女。上海大学信息与通信工程专业博士研究生。主要研究方向为多视点视频编码及其应用。

E-mail: zhangqianleslie@yahoo.com.cn

于像素域分割方法。因此,为兼顾视频对象分割对精度和复杂度的要求,首次尝试利用立体视频压缩域信息来进行视频对象分割,提出一种新的立体视频对象分割方法。通过提取立体视频压缩域的运动矢量场和视差场并进行校正处理,再用均值偏移算法对边缘像素聚类,分割出视频对象。由于分割所用的视差场和运动场直接取自立体视频压缩域,省去了视差和运动估计的过程,大大提高了分割的时效性;而将视差信息用于对象分割,提高了分割的准确性。

2 视差场及运动矢量场的处理

2.1 视差场的分类处理

所谓视差是指由于摄像机距离、拍摄角度等原因造成 3 维空间中同一点在不同底片投影造成的位置差异^[4]。由于视差与物体深度成反比,同一物体内部的视差场应该变化不大,理想情况下视差场只有在对象边缘才会发生突变,因此视差场可以较好的保留物体边缘并保持物体内部的连续性,此特性可以用于对象分割。

采用立体视频编码方法^[5]编码立体视频,从压缩域直接提取视差场。但此视差场是稀疏的并含有较多误匹配,不能直接用于视频对象的分割,需要进行预处理来减少噪声和视差场内的误匹配区域。将视差分为背景视差、前景视差和噪声。如果视差小于预置的阈值(T_1)则被定义为背景视差,一般背景的景深大,因而视差较小。如果视差值大于预置的阈值(T_2)则被定义为噪声,如果视差值位于两个阈值之间则表示对象视差值。阈值 T_1 和 T_2 根据所拍摄序列的相机间距和场景噪声情况由经验而定。最后为了得到更可靠的视差场,将分类后的视差场进行区域增长得到新的视差场 $disp$ 。

2.2 运动矢量场的预处理

利用相邻帧之间的相关性进行运动估计可以得到运动矢量场,运动矢量场可以确定对象的大概位置。本文的运动矢量场是采用 8×8 块匹配得到,可以直接从立体视频压缩域提取。对于分割来说,取自压缩域的运动矢量场的精度不够高,所以必须对运动矢量场做预处理。

为了得到可靠的运动信息,本文利用当前帧的前后两帧对压缩域运动矢量进行叠加得到运动矢量

场 MV 。为了消除噪声并有效的保留图像的边缘和轮廓信息,用数学形态学的有关方法来去除噪声,同时有效地保护了图像的边缘信息。

2.3 形态滤波

数学形态学运算^[6]分为二值形态学和灰度形态学两类,此处,采用灰度数学形态学。数学形态学的运算主要有膨胀、腐蚀两个基本运算,以及由这些基本运算组合而成的其他运算。设 A 为给定的灰度图像, B 为结构元素, A 被 B 的膨胀记为 $A \oplus B$, \oplus 为膨胀算子,膨胀的定义为

$$A \oplus B(x) = \max\{A(x-z) + B(z)\} \quad (1)$$

腐蚀定义为

$$A \ominus B(x) = \min\{A(x+z) - B(z)\} \quad (2)$$

数学形态学的闭运算是先对图像进行膨胀运算再腐蚀其结果,定义为

$$A \cdot B = (A \oplus B(z)) \ominus A \quad (3)$$

由于叠加得到的运动矢量场 MV 在对象内部存在缺口,为了得到连通的运动矢量场。本文用闭运算对运动矢量场 MV 进行处理去除掉与形态不吻合的结构,得到最终的运动矢量场 MV' 。

3 基于运动信息的视差矢量场修正

一般压缩域视差信息比运动信息能够较好地保留对象边缘,而对于背景中的相似区域可能会由于误匹配而造成视差场失真。为了消除这些由于误匹配所带来的视差,同时保留对象边缘和内部的视差,用运动矢量场 MV' 来限制视差场 $disp$ 。即如果在视差场中检测到的点同时也是运动矢量场中的运动点,那么保留该点,否则在视差场中去除该点。利用数学形态学中的腐蚀、膨胀进行处理来填充和平滑视差图内部的空洞和不连续区域。经上述处理可以得到最终的视差场 D 。利用视差场 D 进行分割,可以得到粗糙的分割结果 S ,为了突出边缘,采用均值偏移算法对分割后的图像进一步处理。

4 均值偏移算法聚类

用均值偏移算法^[7]来对立体视频对象细化边缘。由于边缘可能存在于相邻的像素里,对边缘扩展 8 个像素,以涵盖边缘所在区域,并得到新的分割图像 S' 。首先用视差场 D 将图像分为背景层和对

象层,然后应用均值偏移算法对新的分割图像 S' 进行聚类。均值偏移算法的核心是核密度函数的估计,其要点如下:假设 d 维空间有 n 个点 $p_i, i = 1, 2, \dots, n$, (例如:对于灰度图像每个点 p_i 是 3 维向量,代表了该点的位置信息和亮度信息)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

$$\hat{f}(p) = \frac{1}{n} \sum_{i=1}^n K_H(p - p_i) \quad (4)$$

式中, K_H 是一个积分为 1 的、以零为原点的非负函数:

$$K_H(p) = |H|^{-1/2} K(H^{-1/2} p) \quad (5)$$

通常情况下当 H 所选单位矩阵的比例矩阵为 $H = h^2 I$ 时,核密度估计算子转化为

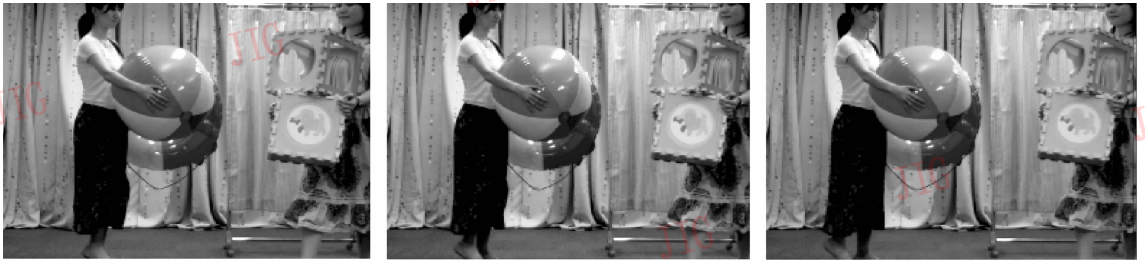
$$\hat{f}(p) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{p - p_i}{h}\right) \quad (6)$$

式中,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。理论上讲属于同一模式的像素在空域上是更加接近的,因此通常用 Epanechnikov 核^[7]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,因此把距离相近的模式点归为同一类。同类的点中如果属于背景点的比例大于比例参数 ρ 就归为背景, ρ 由实验而定。

5 实验结果分析

在内存为 1G 主频为 2.8GHz 的机器上,用本文算法分割立体视频序列 Akko 前景中的 2 个人物对象。图 1 分别为 Akko 序列右视点图像的第 1、2、3 帧,图 2 分别为原始图像的左视点图像和右视点图像。对右图像第 2 帧的分类处理后的压缩域视差场如图 3(a) 所示;用叠加后得到的压缩域运动场如图 3(b) 所示。利用压缩域运动场通过第 3 节中的方法对视差场 $disp$ 进行修正后得到的视差场 D 如图 3(c) 所示;经均值偏移算法聚类后的分割结果如图 3(d) 所示;最终的对象分割结果如图 4 所示。实验中各阈值为 $T_1 = 6, T_2 = 12, \rho = 0.1$ 。由实验结果可知,对图像中的 2 个对象的分割效果主观质量较好。不仅把背景去除了,而且分割所得前景对象的边缘也较准确平滑。由于尚无立体视频对象分割的客观评价准则,这里只给出主观评价。



(a) 第1帧

(b) 第2帧

(c) 第3帧

图 1 原始右视点序列

Fig. 1 Original right image sequence



(a) Akko左图像

(b) Akko右图像

图 2 Akko 序列第 2 帧的左右图像对

Fig. 2 Left and right images of the 2th frame

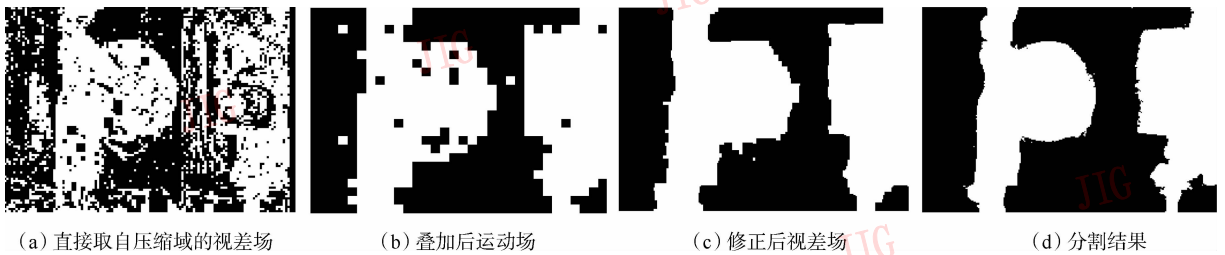


图 3 分割结果

Fig. 3 The segmentation result

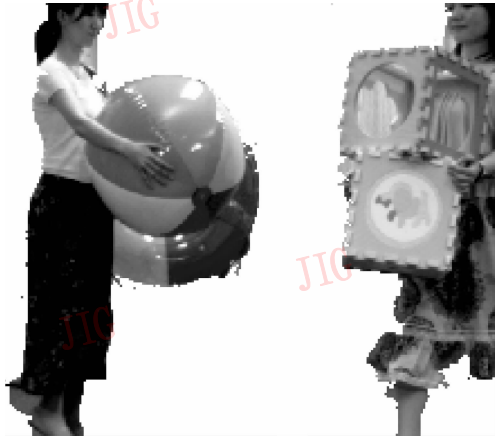


图 4 分割对象

Fig. 4 The segmented object

参考文献 (References)

- 1 Papadimitriou T. Video scene segmentation using spatial contours and 3D robust motion estimation [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2004, 14(4): 55 ~ 62.
- 2 Ren Y G, Yu G. Moving object segmentation method based on color feature [J]. Computer Engineering, 2004, 30(7): 126 ~ 127.
- 3 Liu Fang-qing, Shi Xu-li, Zhang Zhao-yang. Real-time video object segmentation in the compressed domain based on EM clustering [J]. Journal of Image and Graphics, 2007, 12(10): 1819 ~ 1822. [刘方青, 石旭利, 张兆杨. 基于 EM 聚类的 H. 264 压缩域视频对象实时分割算法 [J]. 中国图象图形学报, 2007, 12(10): 1819 ~ 1822.]
- 4 Gao Xin, An Ping, Liu Jia, et al. Stereo video object segmentation based on disparity and change detection [J]. Journal of Shanghai University (Natural Science Edition), 2006, 12(2): 116 ~ 119. [高欣, 安平, 刘佳等. 基于视差和变化检测的立体视频对象分割 [J]. 上海大学学报 (自然科学版), 2006, 12(2): 116 ~ 119.]
- 5 An Ping, Zhang Zhao-yang. Hierarchical overlapped block disparity estimation and adaptive compensation for stereo images [J]. Journal of China Institute of Communications, 2003, 24(3): 1 ~ 7. [安平, 张兆杨. 立体图像分层交叠块视差估计与自适应补偿 [J]. 通信学报, 2003, 24(3): 1 ~ 7.]
- 6 Salembia P, Torres L, Meyer F, et al. Region-based video coding using mathematical morphology [J]. Proceedings of IEEE, 1995, 83(6): 843 ~ 857.
- 7 Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(5): 603 ~ 619.

6 结 论

本文提出了一种基于压缩域视差和运动信息的立体对象视频分割方法,对视差场和运动场进行修正和融合,并用均值偏移算法聚类将背景与前景边缘进行细化,最终分割出对象。该方法兼顾对象分割的时效性和精确性,实验结果显示对于纹理复杂的场景有很好的分割效果,可以获得与语义一致的对象。进一步的工作计划是完善对象边缘的分割。