

文章编号: 1007-2780(2011)02-0229-05

基于背景重构的视频分割技术及应用

范小聪, 李大海*, 王琼华, 刘曦, 张映权, 刘效勇

(四川大学 电子信息学院, 四川 成都 610065, E-mail: 22958325@qq.com)

摘 要: 针对背景静止的立体视频提出了一种快速的基于背景重构的视频分割算法。先利用帧差法分别确定出左右视频的前景运动区域, 再重构出该区域的背景图像, 最后, 通过视频图像和背景图像的对比来准确地提取运动前景。然后, 先匹配左右图像序列的背景图像, 并保存它们的匹配结果, 再分别对各立体图像对的前景进行匹配。采用该方法将立体视频分割可以减少图像数据的传输量和储存空间; 同时, 视频分割之后再匹配运算, 减少了立体匹配的时间。

关 键 词: 视频分割; 背景重构; 视差图像; 立体匹配

中图分类号: TP391.4; TN911.73

文献标识码: A

DOI: 10.3788/YJYXS20112602.0229

Video Object Segmentation Algorithm Based on Background Reconstruction and Its Application

FAN Xiao-cong, LI Da-hai*, WANG Qiong-hua,

LIU Xi, ZHANG Ying-quan, LIU Xiao-yong

*(School of Electronics and Information Engineering, Sichuan University,**Chengdu 610065, China, E-mail: 22958325@qq.com)*

Abstract: A fast video segmentation algorithm based on background reconstruction is proposed for stereo video with static backgrounds. Firstly, the moving foreground areas are found out by the frame difference respectively. And then, the backgrounds shaded by foreground areas are reconstructed. At last, the moving foregrounds could be separated by comparing the backgrounds information with the current frame images and it could achieve the segmentation accurately. After that, two background images reconstructed from the left and right video sequences are matched in the initialization and the parallax image is saved, so that only the foreground parallax images are necessary to be updated. Therefore, the proposed method can improve image compression ratio and reduce storage. Meanwhile, the time for stereo matching is also decreased.

Key words: video segmentation; background reconstruction; parallax image; stereo matching

1 引 言

双目立体视觉技术^[1-4]是人们研究较多且最具实用价值的计算机立体视觉技术。一个完整的

双目立体视觉系统通常可以分为图像获取、摄像机定标、特征提取、立体匹配、深度确定和内插等 6 大部分^[5]。由于立体视觉图像信息量大、立体匹配时间较长, 所以目前的立体视觉系统对计算

收稿日期: 2010-11-17; 修订日期: 2010-12-09

基金项目: 四川省科技厅支撑计划(No. 2008FZ0003)

作者简介: 范小聪(1981-), 男, 河南济源人, 硕士研究生, 主要从事三维立体显示及三维测量方面的研究。

* 通讯联系人, E-mail: lidahai@scu.edu.cn

机硬件要求较高,也很难做到实时处理。针对双目相机拍摄的立体视频,为减少图像的数据量,人们将视频图像分为静止的背景图像和运动的前景图像,并根据左右通道视频序列间的相关性分别对背景图像和前景图像进行压缩编码^[6]。其中,立体匹配是立体视觉的瓶颈,如何提高匹配结果、减少匹配时间一直是研究的重点^[2]。Hansung Kim 等把前景分割应用于立体序列的实时深度重建^[7],取得了较好的效果。

本文针对背景静止的运动视频,提出了一种快速的基于背景重构的视频分割算法,并把它应用到双目相机的立体匹配实验中,既减小了重构的背景区域,又减少了所需要的图像数量,实现了实时分割。

2 视频分割方法

2.1 分割原理

视频分割的主要目的是把人们感兴趣的运动前景从背景图像中分割出来。目前主要的视频分割方法有差分图像法^[8]、时间差分法^[9]和光流法^[10]等。差分图像法是利用原始图像与重构后的背景图像进行差分来实现视频分割的一种方法。时间差分方法是在差分图像法的基础上,引入了时空学的关系。光流法是基于运动目标随时间变化的光流特征来有效地提取和跟踪运动目标。对比这几种方法,差分图像法计算简单,较少受光照影响,对硬件要求低,在大多数情况下检测效果较好^[11]。

差分图像法的关键是如何完整地重构出视频图像的背景。文献^[12]中提到的背景重构方法需要对至少 25 幅视频图像的同一坐标点的像素值进行优选,重构出背景图像。这种方法耗时较长,不利于实时分割。由于各帧视频图像运动前景区域内同一坐标点的灰度值一般不同,即运动前景区域的帧差应当比静止背景区域的帧差大,因此,通过计算连续帧之间的灰度差就可以得到前景的运动区域。本文提出的算法是用帧差法先确定出读入内存的第一帧视频图像的运动前景区域,该区域的背景信息被前景所遮挡住;然后,随着前景的移动,判断出前景运动过后露出的背景像素值,把露出的背景像素值赋给第一帧视频图像中相对应的坐标点。只要被遮挡的背景部分完全露出,就可以重构出完整的背景图像。文章提出了一种

快速的重构背景的方法,既减小了重构的背景区域,又减少了所需要的图像数量,从而减少了计算量,并用该方法对 MPEG-4 测试序列 HallObject 做了实验,得到了可靠的实验结果,实现了实时分割。

2.2 背景图像的重构

重构背景图像的具体步骤如下:

(1)采用 5×5 大小的均值滤波器对读入内存的每一帧视频图像进行预处理以去除高斯噪声。

(2)设视频图像序列各帧图像为 $I_i(x, y)$,建立与之相对应的二值变化检测模板 $CDM_i(x, y)$,矩阵大小与原始图像相同。其中 (x, y) 表示像素位置, i 表示帧数 ($i=1, 2, 3 \dots K, K+1, K+2, \dots$)。依次对第 i 帧和第 $i+1$ 帧视频图像灰度图做帧差。设置阈值 T ,根据式(1)可以得到一组与视频序列相对应的二值变化检测模板:

$$CDM_i(x, y) = \begin{cases} 1, & \text{if } |I_i(x, y) - I_{i+1}(X, Y)| \geq T \\ 0, & \text{if } |I_i(x, y) - I_{i+1}(X, Y)| < T \end{cases} \quad (1)$$

阈值 T 的选取采用文献^[13]所提出的一种自动计算最佳阈值的方法来确定。二值变化检测模板中灰度值为 1 的区域就是相邻两帧灰度图像中灰度值变化较大的区域。针对 HallObject 中的两帧连续视频图像灰度图,如图 1(a)和图 1(b),根据式(1),文章得到了图 1(c)所示的帧差图。

(3)采用一种基于噪声点统计的去噪算法^[12]对二值变化检测模板 $CDM_i(x, y)$ 进行去噪处理。分别以二值变化检测模板 $CDM_i(x, y)$ 中的像素点 (x, y) 为中心取一个 $N \times N$ ($N=3, 5, 7, \dots$) 的去噪窗口,统计去噪窗口内灰度值为 1 的像素数,如果去噪窗口内灰度值为 1 的像素数小于 M (M 一般取窗口内像素数的一半左右),则像素点 (x, y) 的灰度值 $CDM_i(x, y)$ 被赋值为 0;反之, $CDM_i(x, y)$ 被赋值为 1。这样就除去了二值变化检测模板 $CDM_i(x, y)$ 中的孤立噪声,模板中保留的白色区域就粗略地表示了前景运动区域。其中, N 和 M 的选取根据视频图像噪声点的大小和光照变化而定。文章中测试序列的分割实验 N 选取的是 3, M 选取的是 5。图 1(d)是对图 1(c)做上述处理的结果。

(4)上下左右扫描经过去噪处理的 $CDM_i(x, y)$,找出前景运动区域的外接矩形,矩形内所有像

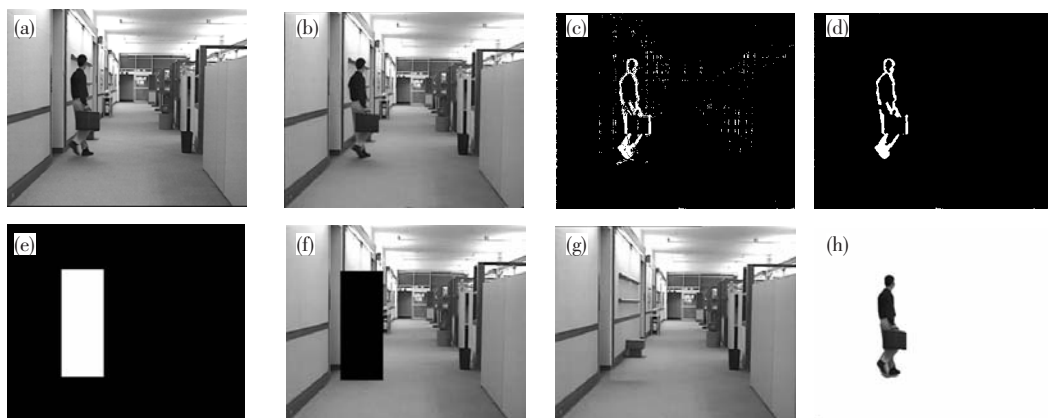


图1 测试序列实验结果。(a) 第1帧图像;(b) 第2帧图像;(c) $CDM_1(x, y)$;(d) 去噪后的 $CDM_1(x, y)$;(e) $DDM_1(x, y)$;(f) 重构的背景部分;(g) 重构的背景图像;(h) 图(a)的前景。

Fig. 1 Results of test video sequence. (a) The first frame; (b) The second frame; (c) $CDM_1(x, y)$; (d) De-noised $CDM_1(x, y)$; (e) $DDM_1(x, y)$; (f) Background area to be reconstructed; (g) Reconstructed background image; (h) Foreground of figure (a).

素点灰度值全部赋为1,矩形外面所有像素点灰度值全部赋为0。为了把被遮挡的背景部分和前景运动产生的阴影全部包含在矩形区域内,可以适当地向上下左右4个方向增加大外接矩形的面积,这个区域记为 $B_i(m, n)$, (m, n) 是矩形区域内像素点的坐标。把经过上述过程处理的二值变化检测模板 $CDM_i(x, y)$ 记为 $DDM_i(x, y)$ 。图1(e)是对图1(d)做上述过程处理后得到的,记为 $DDM_1(x, y)$ 。 $DDM_1(x, y)$ 中灰度值为1的区域记为 $B_1(m, n)$ 。读入内存的第1帧视频图像 $I_1(x, y)$ 中与 $B_1(m, n)$ 相对应的区域记为 $OB(m, n)$ 。 $OB(m, n)$ 就是要重构的背景区域,如图1(f)中黑色部分所示。

(5) 依次对二值变化检测模板 $DDM_i(x, y)$ 中与 $OB(m, n)$ 区域相对应的坐标点的灰度值进行检测,如果第 K 帧、第 $K+1$ 帧、第 $K+2$ 帧连续3帧的 $DDM_i(x, y)$ 二值变化检测模板中同一坐标点灰度值都为0,则确定该坐标点的背景图像已被露出,把 $I_K(x, y)$ 中该点的像素值赋给第1帧图像中 $OB(m, n)$ 里相对应的点,直到 $OB(m, n)$ 中所有点都被赋值,背景就被完全重构出来。图1(g)就是重构后的背景图像。

2.3 提取运动前景

重构出完整的背景图像后,通过当前帧和背景图像的像素值的对比,可准确地完成对运动前景的提取^[12]。首先,建立一个空的矩阵,矩阵大小与原始视频图像大小相同。设置阈值 T ,对原

始视频图像灰度图和重构的背景图像灰度图相减;如果差值大于 T ,则判断该点为前景信息,把原始图像中该点的 RGB 值赋给空矩阵中相对应的点。如果差值小于或等于 T ,则判断该点为背景信息,空矩阵中对应点的 RGB 值赋为255。 T 值的选取与式(1)相同。空矩阵的所有坐标点被赋值后,就粗略地提取了运动前景。然后,将该前景图像转化为二值灰度图,先用噪声门限法消除噪声,再做数学形态学处理,得到没有噪声、前景内部没有空洞、前景边缘比较平滑的二值灰度图。最后,利用原始图像和处理过的二值灰度图,得到比较理想的前景图像。图1(h)就是图1(a)和图1(g)对比后提取出的前景图像。

3 视频分割在立体匹配中的应用

用两个型号为 WAT-221S(PAL) 的 CCD 感应器和焦距为 8 mm 的镜头组成的双目相机,在自然光照条件下拍摄立体视频。视频帧率为 25 帧/s,视频图像大小为 768×576 pixels。先以右通道视频图像为参考,采用一种不基于摄像机标定的非外极几何校正的方法^[14]对左通道中的视频图像进行校正,再分别重构出右通道背景图像和校正后的左通道背景图像,最后提取左右通道的运动前景。图2(a)和图2(b)是用双目相机拍摄的立体视频中的一帧立体图像对。图2(c)是对图2(a)校正后得到的。图2(d)是对校正后左通道图像重构的背景图像。图2(e)是右通道背景

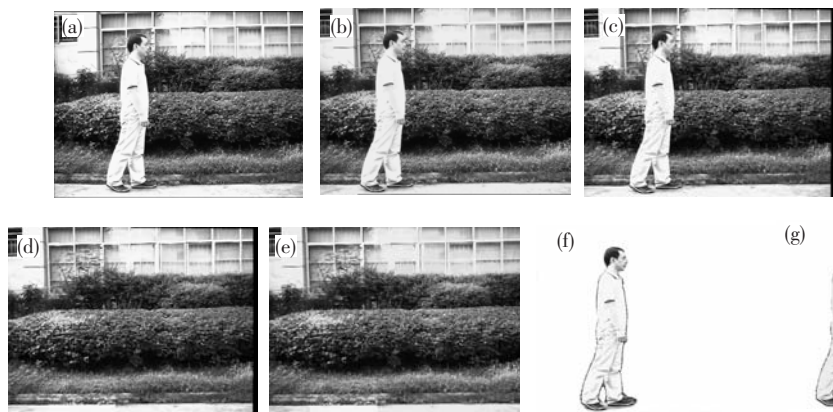


图 2 立体视频分割实验结果。(a) 左通道图像;(b) 右通道图像;(c) 图(a)的校正结果;(d) 左通道背景图;(e) 右通道背景图;(f) 图(c)的前景图像;(g) 图(b)的前景图像。

Fig. 2 Results of stereo video segmentation. (a) One frame of the left video; (b) The corresponding right image of figure (a); (c) Rectification result of figure (a); (d) Reconstructed background of the left video; (e) Reconstructed background of the right video; (f) Foreground of figure (c); (g) Foreground of figure (b).

图像。图 2(f)和图 2(g)分别是图 2(c)和图 2(b)的前景图像。

对校正后的立体图像对采用基于自适应权值算法、Rank 变换和视差校准的彩色图像立体匹配算法^[15]进行密集型匹配。在初始阶段,对重构的两幅背景进行匹配,对左右通道提取的相对应的运动前景图像依次进行匹配。这样就可以得到一幅背景深度图和多幅连续的前景深度图。因为左右通道中相对应的前景被提取出来后,视差并没

有改变,而且背景深度图和前景深度图的灰度值的赋值方法是一样的,所以,可以将背景深度图与各帧前景深度图分别进行合成,得到分割后视频图像的深度图。图 3(a)是图 2(d)和图 2(e)的匹配结果。图 3(b)是图 2(f)和图 2(g)的匹配结果,图 3(c)是图 3(a)和图 3(b)的合成图,图 3(d)是图 2(b)和图 2(c)没有经过视频分割直接匹配的结果。通过对比图 3(c)和图 3(d)可以发现,这两幅深度图的差异较小。因此,基于背景重构的视

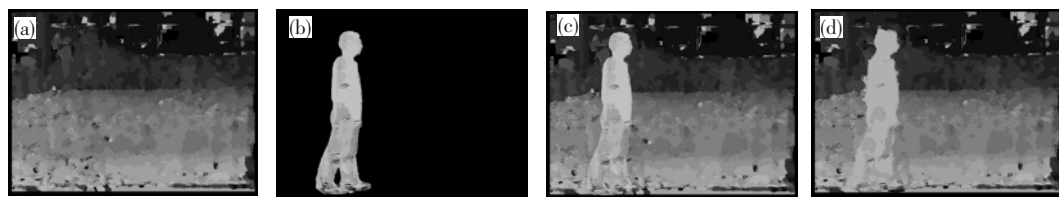


图 3 立体匹配实验结果。(a) 背景深度图;(b) 前景深度图;(c) 图(a)图(b)合成图;(d) 原始图像深度。

Fig. 3 Results of stereo matching. (a) Depth map of the background; (b) Depth map of the foreground; (c) Combined image of figure (a) and (b); (d) Depth map of the original image.

频分割可应用于立体图像匹配。

假设要对 N 帧视频图像做立体匹配,匹配每帧图像对需要的时间平均为 t_0 ,则匹配 N 帧图像对需要的总时间:

$$T_0 = N \times t_0 \quad (2)$$

如果前景面积占原始图像面积的 $1/n$,则只匹配一次前景图像对需要时间为 t_0/n 。实验中,匹配前景所消耗的时间约为匹配原始图像所需时

间的 $1/9$ 。设重构背景的时间为 t_1 ,提取前景的时间为 t_2 ,视频分割之后,只需对重构的背景匹配一次,匹配背景需要的时间和匹配原始图像需要的时间 t_0 基本相同。设背景深度图与每帧前景深度图进行合成的平均时间为 t_3 ,则匹配 N 帧图像需要时间为

$$T_1 = t_1 + N \times t_2 + t_0 + N \times t_0/n + N \times t_3 \quad (3)$$

由实验可知, t_1 远小于 t_0 , t_2 和 t_3 都是毫秒量级, 所以

$$T_1 \approx t_0 + N \times t_0/n \quad (4)$$

由式(2)和式(4)可得

$$T_0 - T_1 = (N - 1 - N/n) \times t_0 \quad (5)$$

由于前景在图像中占的面积较小, 即 n 值较大, 所以 $T_0 > T_1$, 且 N 越大, 即匹配的图像对越多, 节省的时间就越多。

4 结 论

针对背景静止的运动视频图像, 首先用差分

图像法确定出运动前景区域, 然后重构出该区域的背景图像, 最后利用原始图像和重构的背景图像提取每一帧的运动前景。因为只是重构被前景遮挡的背景部分, 所以计算量小, 利于实时分割。对视频图像的静止背景只做一次存储和匹配, 对运动前景做多次存储和匹配, 较大程度地提高了数据压缩效率, 减少了匹配的时间消耗, 从而使得实时匹配成为可能, 进而缩短基于匹配的立体视频对的三维测量和三维重建的时间。

参 考 文 献:

- [1] Tsai R Y. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV camera and lenses [J]. *IEEE Journal of Robotics and Automation*, 1987, 3(4): 323-344.
- [2] 何友兵, 李大海, 李良玉, 等. 彩色立体图像匹配算法研究 [J]. *液晶与显示*, 2007, 22(4): 417-422.
- [3] 张超, 王琼华, 李大海, 等. 从平面图生成多视点自由立体显示视差图的研究 [J]. *液晶与显示*, 2009, 24(2): 258-261.
- [4] 张浩, 李大海, 王琼华, 等. 自由立体显示拍摄系统中摄像机空间自由度的确定 [J]. *液晶与显示*, 2010, 25(2): 287-291.
- [5] Barnard S T, Fischler M A. Computational stereo [J]. *ACM Computing Surveys*, 1982, 14(4): 552-572.
- [6] 王世刚, 林青, 付庆兴, 等. 基于视差和帧差运动检测的立体视频对象提取与压缩编码 [J]. *中国电视学与图像分析*, 2009, 14(1): 67-72.
- [7] Kim H, Choi S, Sohn K. Real-time depth reconstruction from stereo sequences [J]. *Proc. of SPIE*, 2005, 6016: 60160E(1-12).
- [8] Haritaoglu I, Harwood D, Davis L. Real-time surveillance of people and their activities [J]. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2000, 22(8): 809-830.
- [9] 李彬, 刘冀伟, 韩鸿哲, 等. 基于步态特征的快速身份识别方法 [J]. *计算机工程与应用*, 2004, 22(2): 60-62.
- [10] 夏侯玉娇, 龚声蓉, 刘纯平, 等. 结合 Gaussian 分布和 LK 光流法的视频对象分割算法 [J]. *微电子学与计算机*, 2009, 26(6): 239-245.
- [11] 伏思华, 张小虎. 基于序列图像的运动目标实时检测方法 [J]. *光学技术*, 2004, 30(2): 215-217.
- [12] 袁静. 动目标的提取分析与研究 [D]. 西安: 西安电子科技大学, 2007.
- [13] 王军, 沙芸, 吴裕树. 基于背景模型的自动视频分割方法 [J]. *计算机工程与应用*, 2004, 40(9): 60-62.
- [14] Wang H M, Chang C W, Yang J F. An effective calibration procedure for correction of parallax unmatched image pairs [J]. *IET Image Process*, 2009, 2(3): 63-74.
- [15] Gu Z, Su X Y, Liu Y K, et al. Local stereo matching with adaptive support-weight rank transform and disparity calibration [J]. *Pattern Recognition Letters*, 2008, 29(9): 1230-1235.