

基于双目计算机视觉的自适应识别算法及其监控应用

丁 谨¹⁾ 王 新²⁾

¹⁾(复旦大学软件学院,上海 201103) ²⁾(复旦大学计算机科学与工程系,上海 200433)

摘要 双目计算机视觉是利用仿生学原理,通过标定后的双摄像头来得到同步曝光图像,然后计算获取的2维图像像素点的第3维深度信息。为了对不同环境场景进行监控提出了一种新的基于双目计算机视觉的自适应识别算法。该算法首先利用像素点的深度信息对场景进行识别判断,然后采用统计的方法为场景建模,并通过时间滤波克服光照渐变,以及通过深度算法特性克服光照突变。与单摄像头监控系统相比,利用该算法实现的视频监控原型系统,可应用于更多场合,并利用深度信息设置报警级别,来降低误检率。

关键词 双目计算机视觉 深度信息 自适应 光照变化 视频监控

中图法分类号: TP391.41 TP242.62 **文献标识码:** A **文章编号:** 1006-8961(2006)11-1708-04

An Adaptive Recognition Algorithm Based on Two-eye Computer Vision and Its Application in Monitoring

DING Jin¹⁾, WANG Xin²⁾

¹⁾(Software School, Fudan University, Shanghai 201103)

²⁾(Department of Computer Science and Engineering, Fudan University, Shanghai 200433)

Abstract Two-eye computer vision algorithm uses the principle of Bionics. It gets the depth information of each pixel by two demarcated cameras. A new adaptive recognition algorithm based on two-eyes computer vision is presented in this paper. This algorithm makes use of the depth information of each pixel. Then it adopts the statistics method to model the scene and overcomes the gradual light change using time filter. It can also overcome the sudden light change with the characteristic of the depth algorithm. The video monitor antitype system using our algorithm can be set at more places than one camera system. It can set the alarm level according to the depth. Information, which can reduce the mistake ratio.

Keywords two-eye computer vision, depth information, adaptive, light change, video monitoring

1 引言

面对日益复杂的社会和政治环境,国家安全、社会安全、个人人生安全和财产安全等都面临着不同程度的威胁,都需要各种安全保护措施,在众多场所建立切实有效的安保措施,成为一个迫切的课题。本文提出了一种基于双目计算机视觉的自适应识别

算法,将该算法应用于现有的监控系统,并赋予监控终端智能性,不仅使其脱离人而具有独立智能、自主判断的能力,而且使得视频监控系统在安防方面的作用大大提高。

在现有的背景建模方法中,大多对于背景像素点的亮度值,例如最小亮度值、最大亮度值和最大亮度差值^[1],或是对颜色信息进行建模^[2]。对于背景的更新,一般使用自适应滤波器对像素的统计特性

基金项目:上海市重点攻关项目(055115009);上海市重大科技攻关项目(04DZ15021-5A);上海市智能信息处理重点实验室开放基金项目(IPL-04-004)

收稿日期:2006-06-15; **改回日期:**2006-08-02

第一作者简介:丁谨(1982~),男,2004年复旦大学本科毕业,现为复旦大学硕士研究生。研究方向为嵌入式系统、计算机视觉。
E-mail: 042053022@fudan.edu.cn

进行递归更新,为了考虑到噪声的影响,文献[3]提出了 Kalman 滤波器的方法,该文认为系统的最优信息可通过估计获得。考虑到环境的动态缓慢改变,文献[4]利用统计模型给背景建模,即由一个时域滤波器保留着一个序列均值和一个标准偏差,并通过滤波过程统计值随时间改变来反映环境的动态特性。另外有一些方法解决了光照渐变等影响^[5-7],但计算较复杂。

2 双目计算机视觉深度算法

基于实际应用考虑,摄像头的数量关系着成本和计算量,所以选择支持双摄像头(双目视觉)的算法是最合适的。在支持双目视觉的算法中,Princeton NEC research institute 基于最大流算法(maximum-flow)的计算机视觉算法(Stereo-MF)在深度效果平滑性上做得较好^[8,9],适用于监控区域深度计算的应用背景。但原有算法所需的计算量和计算过程中的暂存数据量是较大的,虽然支持计算量的削减,但只是机械地在一块区域中选择中心点来进行计算,这样计算的结果会因选择的机械性,而出现大量的“伪点”,这些伪点错误地表现了该区域的平均深度信息。本文采用统计平均值选取计算点,通过距离因子的 Gauss 分布将块内其他点的值融合计算,从而使得计算出的值较准确的代表了这一块内的大致深度分布。

$$\bar{d} = \frac{1}{n \times m} \sum_{i=1}^n \sum_{j=1}^m d_{i,j} \quad (1)$$

$$|d_{m,n} - \bar{d}| = \min |d_{i,j} - \bar{d}| \quad (2)$$

$$q_{i,j} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{h^2}{2\sigma^2}} \quad (3)$$

$$h = |i - M| + |j - N| \quad (4)$$

$$\sigma = \sqrt{m \times n} \quad (5)$$

$$d_B = \sum_{i=1}^n \sum_{j=1}^m (q_{i,j} \times d_{i,j}) \quad (6)$$

m, n 分别是图像的长和宽所包含的像素点个数, M, N 表示像素点的横纵坐标, \bar{d} 是块内深度统计平均值, $d_{m,n}$ 为计算点的深度值, q 为距离因子, d_B 是计算所得的块深度代表值。

图 1 为改进后双目视觉深度算法与原算法识别效果比较。

由图 1 可以明显看出,修改后的算法效果在细节表现、平滑性、伪点减少上均有明显改善,而且深

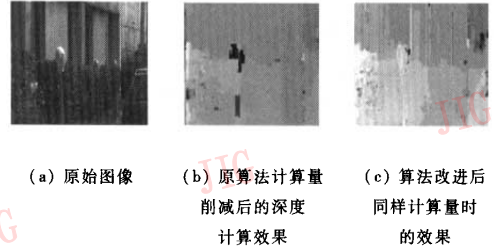


图 1 改进后双目视觉深度算法与原算法识别效果比较

Fig. 1 Effect comparison after algorithm modification

度计算精确度能够完全满足视频监控应用的需要。

3 自适应识别算法

对于一个固定的场景,场景各像素点的深度值是符合一个随机概率分布。以某一均值为基线,在其附近做不超过某一偏差的随机振荡,这种情况下的场景称之为背景。而场景环境往往是动态变化的,如环境自然光的缓变,灯光的突然熄灭或点亮,以及运动对象的出现、运动和消失等。如果能识别出场景中的动态变化,就能自适应的更新背景值,将光照的改变融合到背景值中。本文采取了用统计模型的方式给每个像素点建模,而以像素点变化的分布情况来确定光照突变引起的深度突变,并结合深度计算本身特性,解决光照缓变突变引起的误判问题,以及判别场景中对象的主次性。

3.1 背景像素点的深度值建模

由于双目计算机视觉算法得到的深度值,已经是块融合的,可以根据精度要求,来加大块面积,减少数据量。本文获得的数据量只有原像素点的 $\frac{1}{k \times l}$ (k, l 分别是块的长和宽所包含的像素点个数)。以统计的方法给每个像素点的深度值建模,设 $D_{i,j}^{(u)}$ 为第 u 帧图像的某个像素点的深度值,其中 u 代表第 u 帧图像, i, j 分别代表像素点的横坐标和纵坐标。由一个时间滤波器来保持该像素点深度值的序列均值 $\overline{D_{i,j}^{(u)}}$ 和时间偏差 $\overline{\sigma_{i,j}^{(u)}}$:

$$\overline{D_{i,j}^{(u)}} = \alpha D_{i,j}^{(u)} + (1 - \alpha) \overline{D_{i,j}^{(u-1)}} \quad (7)$$

$$\overline{\sigma_{i,j}^{(u)}} = \alpha |D_{i,j}^{(u)} - \overline{D_{i,j}^{(u)}}| + (1 - \alpha) \overline{\sigma_{i,j}^{(u-1)}} \quad (8)$$

其中, α 是一个可调增益参数,其与采样频率有关。通过滤波过程,来得到每个像素点的深度值基于时间的统计特性,由于这些统计特性反映了环境的动

态特性,据此可以了解到是环境的光照发生了突变,还是有运动对象的运动。

3.2 背景更新与场景识别

通过上述滤波过程,就可以将光照缓变融入到背景中去,实现背景的自适应更新。而对于光照突变,此时几乎所有的像素点的亮度值会同时增大或减小,但根据最大流算法的特性,同方向的变化对流量差不会引起太大变化,而对深度计算结果只会引起较小的同方向变化。这种全局的等量变化,可以认为是光照突变引起的。

$$|D_{i,j}^{(u)} - \overline{D_{i,j}^{(u-1)}}| > a \overline{\sigma_{i,j}^{(u-1)}} \quad (9)$$

$$s \times t - Q < b \quad (10)$$

$$\overline{\sigma}^{(u)} = \frac{1}{s \times t} \sum_{i=1}^s \sum_{j=1}^t \overline{\sigma_{i,j}^{(u)}} \quad (11)$$

$$\frac{1}{s \times t} \sum_{i=1}^s \sum_{j=1}^t (\overline{\sigma_{i,j}^{(u)}} - \overline{\sigma}^{(u)})^2 < c \quad (12)$$

其中, a, b 和 c 是 3 个可调节系数,他们的取值可依据场景的情况及检测光照突变的速度与误差来进行选取。 s, t 分别是深度图像的长和宽所包含的像素点个数。 Q 是符合式(9)的像素点个数。一旦检测到环境光照发生了突变,就把背景点像素的深度序列均值 $\overline{D_{i,j}^{(u)}}$, 全部以当前帧像素点的深度值的测量值 $D_{i,j}^{(u)}$ 代替,而 $\overline{\sigma_{i,j}^{(u)}}$ 以 0 取代,从而实现背景的及时更新。

如果式(10)式(12)中任意一个不成立的话,则认为像素块深度值的变化并非由光照突变引起,而是场景中有运动对象出现。

4 算法分析与实验

4.1 算法复杂度

对于光照突变检测,若有突变的话,则会立即检测出来,当有运动对象出现时,并且式(10)式(12)都接近满足时,处理会较慢,因为需要处理突变检测和运动对象两个过程。当 b 取 25% $s \times t$ 时的处理速度与变化点比例关系如图 2 所示。

相对于一般的光强、灰度的识别检测算法,本算法的优势在于不仅可以利用深度特性更容易地检测到光照的渐变与突变,而且可以判断出现的多个运动对象的主次性。

4.2 算法误检率

由于光照直接对于像素点的光强、灰度等产生影响,所以深度算法的噪声容限更大,这样可降低了

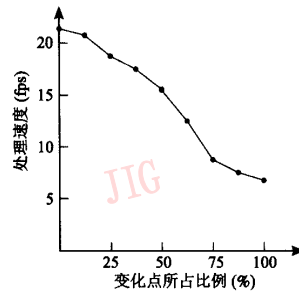


图 2 处理速度图

Fig. 2 Speed of processing

误检率,多组实验后得到的误检率对比图如图 3 所示。但是由于深度算法本身对于反光或者阴暗面会产生伪点,所以,某些时候由于光照突变中光源的位置变化而会误检为运动对象,为此算法还需进一步改进能判别伪点的出现,除去它在光照突变检测中的影响。另外,公式中可调节系数 a, b, c 的选取也会对不同场景产生影响。

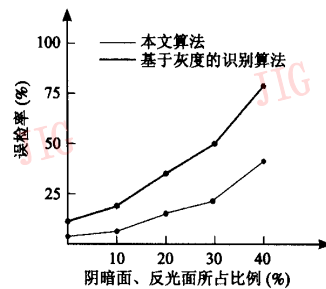


图 3 误检率对比图

Fig. 3 Comparison of mistake ratio

笔者在实验室环境下做了不同光照角度、不同环境光强度、不同运动物体的多组实验,发现在反光面或是阴暗面较多的情况下,光照突变检测不是很灵敏,而且会出错,但是在增加系数 a, c 的值后,误检率有所降低(如图 3 所示)。

5 基于算法的监控系统

我们利用该算法实现了视频监控原型系统。计算机视觉算法对于摄像头的同步曝光要求很高,所以本系统终端用一块单独的 MCU (micro control unit) 控制同步曝光。核心算法用 DSP 处理。系统结构如图 4 所示。实际系统原型图如图 5 所示。

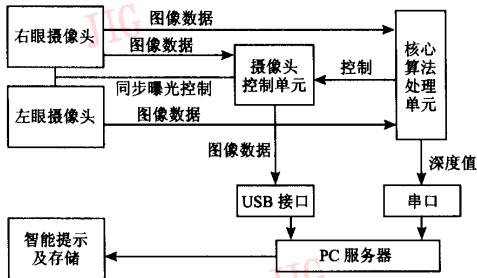


图 4 系统结构图

Fig. 4 System architecture

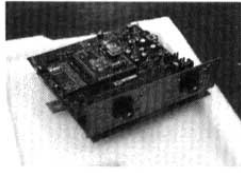


图 5 实际系统原型图

Fig. 5 System picture

参考文献 (References)

- 1 Ude A, Riley M. Prediction of body configurations and appearance for model-based estimation of articulated human motions [A]. In: IEEE SMC '99 Conference Proceeding [C], Tokyo, Japan, 1999: 687 ~ 691.
- 2 Ricquebourg Y, Bouthemy P. Real-time tracking of moving persons by exploiting spatio-temporal image slices [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 797 ~ 808.
- 3 Tsap L V, Goldof D B, Sarkar S. Nonrigid motion analysis based on dynamic refinement of finite element model [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(5): 526 ~ 543.
- 4 Haritaoglu I, Harwood D, Davis L. A real time system for detecting and tracking people [A]. In: Third International Conference on Automatic Face and Gesture [C], Nara, Japan, April 1998.
- 5 Wren C, Azarbayejani A, Darrell T. Real-time tracking of the human body [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 780 ~ 785.
- 6 Ridder C, Munkelt O, Kirchner H. Adaptive background estimation and foreground detection using Kalman filtering [A]. In: Proceedings International Conference. Recent Advances in Mechatronics, ICRAM '95, Istanbul, Turkey, 1995: 193 ~ 199.
- 7 Fujiyoshi H, Lipton A J. Real-time human motion analysis by image skeletonization [A]. In: Proceedings of the Workshop on Application of Computer Vision, Freiburg, Germany, October 1998.
- 8 Sebastien Roy, Ingemar J Cox. A maximum-flow formulation of the n-camera stereo correspondence problem [A]. In: International Conference on Computer Vision (ICCV '98) [C], Bombay, India, January 1998: 492 ~ 499.
- 9 Cox I J, Hingorani S, Maggs B M. A maximum likelihood stereo algorithm [J]. Computer Vision and Image Understanding, 1996, 63(3): 542 ~ 567.

6 结 论

利用深度信息做智能场景识别,是一种新的尝试,有其优势。将这种方法应用到智能视频监控中,能起到很好的效果,克服了其他方法较难处理的光照渐变和突变等问题。对比单摄像头监控系统,该系统可应用于更多场合。后续研发准备在系统上加上更多功能,以适用于更多的环境,并与其他保安类监控系统互联,以组成一整套功能强大、达到国内外一流水平的安防系统。