

基于二维主成分分析的运动目标检测

王 路 王 磊 卓 晴 王文渊

(清华大学自动化系 北京 100084)

摘 要 运动目标检测是计算机视觉研究的重要领域,在视频监控和智能交通等领域应用广泛。本文提出了一种自适应的运动目标检测方法。该方法采用二维主成分分析建立背景模型,通过比较重建图像和原图像的差异来检测运动目标。为了自适应动态变化的复杂背景,该模型由增量算法在线更新。实验结果表明,提出的方法可以在复杂变化的动态背景环境下进行有效的运动目标检测。

关键词 背景建模,运动检测,二维主成分分析

Moving Object Detection Based on Two-dimensional PCA

WANG Lu WANG Lei ZHUO Qing WANG Wen-yuan

(Department of Automation, Tsinghua University, Beijing 100084, China)

Abstract Moving object detection is a basic component of many computer vision applications. It has a critical impact on the performance of visual surveillance and intelligent transportation. This paper proposes an adaptive moving object detection approach that can deal with dynamic scenes. This approach uses 2D-PCA to describe the background model and on-line updates it using incremental learning method. Experimental results demonstrate that the proposed method is able to detect moving objects under various types of dynamic scenarios.

Keywords Background modeling, Motion detection, 2D-PCA

1 引言

运动目标检测是一种重要的计算机视觉技术,在视频监控和智能交通等领域有着广泛的应用。当前运动目标检测的主流方法是背景差分法。该方法首先对视频场景进行建模,然后将待检测的图像与背景模型作比较,进而分辨出前景运动目标。

由于光照的变化、背景物体的运动、摄像机的晃动等因素,实际场景总是处于不断的变化之中,因此背景差分法的核心问题在于怎样建立有效的背景模型。目前应用最广泛的一类建模方法是基于像素点的统计建模,主要包括混合高斯法(Mixture of Gaussian)^[1,2]和核密度估计法(Kernel Density Estimation)^[3,4]。混合高斯法对图像中每一个像素点的像素值按高斯混合模型进行统计建模,然后用 EM 算法不断更新每个高斯模型的参数。核密度估计法则采用变带宽的非参数核密度估计来计算每一个像素点的像素值随时间变化的概率密度,以达到建模的目的。虽然这两种方法取得了较不错的效果,但由于这类方法都是基于独立像素点的,因此这些建模方法忽略了像素点之间的相关性,丢失了有价值的信息。此外由于要对每个像素点进行统计建模,使得这类方法的计算量都较大。

针对这些问题,本文提出了一种基于二维主成分分析(2D-PCA)的自适应运动目标检测方法。这里采用二维主成分分析来描述背景模型,有效地捕捉了二维空间上像素点之间的相关性,取得了更好的建模效果。同时本文提出了一种增量方法来在线更新背景模型,从而使得该方法能适应不断变化的复杂场景。由于不需要对每个像素点进行统计建模,

该方法大大降低了计算量,取得了较好的实时性效果。

2 2D-PCA 方法

二维主成分分析^[5](2D-PCA)是最近提出的一种图像特征提取方法,该方法在人脸识别等领域中得到了成功的应用。与传统 PCA 方法相比,2D-PCA 直接以二维的图像矩阵为分析对象,不需要像传统 PCA 那样事先将图像转变成一维向量再计算,因此这种方法更好地保存了图像的二维空间信息。同时,在构造图像协方差矩阵时,2D-PCA 直接用原始图像矩阵来构建,与传统 PCA 的协方差矩阵相比,2D-PCA 中的图像协方差矩阵的维数要小很多,这样不仅降低了计算量,也避免了训练样本数较少时 PCA 可能出现的矩阵奇异问题。2D-PCA 的基本原理如下:

设 X 表示一个 n 维的列向量,将图像 A (表示为一个 $m \times n$ 的矩阵)向 X 方向做线性投影: $Y=AX$, 得到一个 m 维列向量 Y ,称之为图像 A 的投影特征向量。2D-PCA 方法通过最大化投影特征向量的总离散度(协方差矩阵的迹)来选择一个最优的投影向量 X_{opt} :

$$X_{opt} = \operatorname{argmax}_X J(X) = \operatorname{argmax}_X (\operatorname{tr}(S_X)) \quad (1)$$

其中 $J(X) = \operatorname{tr}(S_X)$, 表示投影特征向量的总离散度; $S_X = X^T [E(A-EA)^T (A-EA)] X$, 表示投影特征向量的协方差矩阵; $\operatorname{tr}(S_X)$ 表示 S_X 的迹。

定义矩阵 G_t :

$$G_t = E[(A-EA)^T (A-EA)] \quad (2)$$

称之为图像协方差矩阵。从定义可以看出, G_t 是一个非负定的 $n \times n$ 矩阵,可以从训练图像样本直接得到。设有 M 个训

练图像样本 $\{A_1, A_2, \dots, A_M\}$, 训练图像的样本平均用 $\bar{A}$ 表示, 则有:

$$G_i = \frac{1}{M} \sum_{j=1}^M (A_j - \bar{A})^T (A_j - \bar{A}) \quad (3)$$

将 G_i 引入(1)式, 得到如下判定准则:

$$J(X) = X^T G_i X \quad (4)$$

这个判定准则称为广义总体散度准则。使判定准则取最大值的向量 X_{opt} 称为最佳投影轴, 也就是说将图像矩阵投影到最佳投影轴 X_{opt} 上之后, 投影特征向量的总离散度(协方差矩阵的迹)为最大。

一般说来, 只有一个最佳投影轴是不够的, 通常需要寻找一组满足正交条件且使判定准则 $J(X)$ 取最大值的投影轴 $\{X_1, X_2, \dots, X_d\}$, 这样的最优投影向量组可以通过求解 G_i 的前 d 个最大特征值对应的特征向量而得到。

有了一组最优投影向量后, 对任意图像样本 A , 令 $Y_k = AX_k, k=1, 2, \dots, d$, 得到投影特征向量组 $\{Y_1, Y_2, \dots, Y_d\}$, 以此构成的 $m \times d$ 矩阵 $U = \{Y_1, Y_2, \dots, Y_d\}$, 称为图像 A 的特征矩阵或特征图像。值得注意的是, 2DPCA 的每个主成分都是向量, 而 PCA 的主成分是标量。

3 自适应运动目标检测

3.1 建立初始背景模型

我们使用 2D-PCA 方法来建立一个投影特征空间用以描述背景模型。首先收集一个包含 M 幅场景图像的训练集 $I = \{A_1, A_2, \dots, A_M\}$, 其中每幅图像用一个 $m \times n$ 的矩阵 A_i 表示。这些场景图像涵盖了背景的一些变化, 如天气、光照的变化等等。按(3)式计算图像协方差矩阵 G_i^M , 对 G_i^M 进行特征值分解, 得到:

$$\Lambda = U^T G_i^M U \quad (5)$$

式中, $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, $U = [X_1, X_2, \dots, X_n]$ 。保存 U 中的 d 个最大特征值对应的特征向量, 构造投影矩阵 $U_d = [X_1, X_2, \dots, X_d]$, 则 U_d 就是我们需要的能够描述场景背景的特征空间。

要说明的是, 即使训练集图像中包含了前景运动目标, 但由于运动目标不会在同一位置始终出现, 在二维主成分分析中, 这些运动目标的信息占据了非常少的部分, 因此通过如上过程得到的特征空间 U_d 主要描述了背景信息。

3.2 模型更新

虽然通过上述方法我们从训练集中得到了描述背景的特征空间 U_d , 但实际场景是在不断变化的, 仅仅凭借训练集图像离线训练得到的背景模型 U_d 并不能一直反映背景的诸多变化。因此, 我们有必要不断地在线更新模型来处理背景的不断变化。由于视频图像往往占用很多的存储空间, 不可能将所有的图像序列存储下来用于计算新的模型, 这里我们提出了一种增量方法来更新背景模型, 既避免了存储量的增加, 也降低了计算量。

设 $\{A_1, A_2, \dots, A_M\}$ 是已有的 M 帧图像, 这些图像的均值 $\bar{A}_M$ 和图像协方差矩阵 G_i^M 已经事先得到。设 $\{A_{M+1}, A_{M+2}, \dots, A_{M+N}\}$ 是新增的 N 帧图像, 这部分图像的均值为 $\bar{A}_N$, 图像协方差矩阵为 G_i^N , 则总的图像协方差矩阵按下式更新:

$$G_i^{new} = \frac{1}{M+N} (MG_i^M + NG_i^N) + \frac{MN}{(M+N)^2} (\bar{A}_M - \bar{A}_N)^T (\bar{A}_M - \bar{A}_N) \quad (6)$$

证明: 设所有 $M+N$ 个样本的总均值 $\bar{A}$ 为:

$$\bar{A} = \frac{M}{M+N} \bar{A}_M + \frac{N}{M+N} \bar{A}_N \quad (7)$$

则新图像协方差矩阵是

$$G_i^{new} = \frac{1}{M+N} \sum_{j=1}^{M+N} (A_j - \bar{A})^T (A_j - \bar{A}) + \frac{1}{M+N} \sum_{j=M+1}^{M+N} (A_j - \bar{A})^T (A_j - \bar{A}) \quad (8)$$

其中(8)式第一项中的求和项

$$\begin{aligned} & \sum_{j=1}^M (A_j - \bar{A})^T (A_j - \bar{A}) \\ &= \sum_{j=1}^M (A_j - \bar{A}_M + \bar{A}_M - \bar{A})^T (A_j - \bar{A}_M + \bar{A}_M - \bar{A}) \\ &= MG_i^M + M(\bar{A}_M - \bar{A})^T (\bar{A}_M - \bar{A}) + \\ & \quad (\sum_{j=1}^M A_j - M\bar{A}_M)^T (\bar{A}_M - \bar{A}) + \\ & \quad (\bar{A}_M - \bar{A})^T (M\bar{A}_M - \sum_{j=1}^M A_j) \\ &= MG_i^M + M(\bar{A}_M - \bar{A})^T (\bar{A}_M - \bar{A}) \end{aligned} \quad (9)$$

类似推导可得(8)式中的第二项

$$\frac{1}{M+N} \sum_{j=M+1}^{M+N} (A_j - \bar{A})^T (A_j - \bar{A}) = \frac{N}{M+N} G_i^N + \frac{N}{M+N} (\bar{A}_N - \bar{A})^T (\bar{A}_N - \bar{A}) \quad (10)$$

将(9)和(10)代入(8)式即可得到:

$$G_i^{new} = \frac{M}{M+N} [G_i^M + (\bar{A}_M - \bar{A})^T (\bar{A}_M - \bar{A})] + \frac{N}{M+N} [G_i^N + (\bar{A}_N - \bar{A})^T (\bar{A}_N - \bar{A})] \quad (11)$$

将(7)代入(11), 整理后即可得到(6)。

因此在更新过程中, 只要存储过往图像的均值 $\bar{A}_M$ 和图像协方差矩阵 G_i^M , 以及最新获取到的 N 帧图像即可。每新得到 N 帧的图像, 进行一次协方差矩阵 G_i 的更新, 然后对其进行特征值分解, 就可以得到更新后的特征向量矩阵。由于 G_i 的维数比较小, 新的特征值分解不会额外增加很多计算量。

3.3 前景检测

通过上面两节, 我们得到了一个能在线更新的背景模型, 在新视频帧图像到来的时候, 可以通过检测背景模型和新图像的差异来分辨出前景运动物体。

假设新到来的待检测图像为 A , 将其投影到背景特征空间 U_d 上, 得到一组投影向量 $Y_k = AX_k, k=1, 2, \dots, d$, 令 $V = \{Y_1, Y_2, \dots, Y_d\}$, 则:

$$V = AU_d \quad (12)$$

将图像 A 的特征图像 V 反投影到图像空间进行重建, 得到:

$$\tilde{A} = VU_d^T = \sum_{k=1}^d Y_k X_k^T \quad (13)$$

接着计算重建图像和原始图像各像素点之间的欧氏距离: $d(x, y) = \|A(x, y) - \tilde{A}(x, y)\|$ 。如果在某像素点 (x, y) 位置该距离大于某个设定的阈值, 即 $d(x, y) > T$ 成立, 则该图像中点 (x, y) 被认为属于前景目标, 否则就被认为是背景点。

3.4 算法流程

第 1 步 建立初始子空间

收集 M 幅场景图像, 计算它们的均值和图像协方差矩阵。对协方差矩阵进行特征值分解, 保存 d 个最大的特征值对应的特征向量, 构造投影矩阵 U_d 。

第 2 步 前景检测

(下转第 217 页)

- IEEE Transactions on Information Theory, 2004, 50(12): 3146-3169
- [8] Makinen V, Navarro G, Ukkonen E. Approximate matching of run-length compressed strings [J]. Algorithmica, 2003, 35(4): 347-369
- [9] Radson D, Boyd H A. Graphical representation of run length distributions [J]. Quality Engineering, 2005, 17(2): 301-308
- [10] Amir A, Landau M G, Sokol D. Inplace run-length 2d compressed search [J]. Theoretical Computer Science, 2003, 290(3): 1361-1383
- [11] Flusser J. Refined moment calculation using image block representation [J]. IEEE Transactions on Image Processing, 2000, 9(11): 1977-1978
- [12] Malo J, Epifanio I, Navarro R, et al. Nonlinear image representation for efficient perceptual coding [J]. IEEE Transactions on Image Processing, 2006, 15(1): 68-80

- [13] Kharinov V M. Representation of image information for machine computation [J]. Pattern Recognition and Image Analysis, 2005, 15(1): 212-214
- [14] Wundrich J I, Von der Malsburg C, Wurtz P R. Image representation by complex cell responses [J]. Neural Computation, 2004, 16(12): 2563-2575
- [15] Liu Y, Ranganath S. Wavelet in POCS for image segment representation [J]. Electronics Letters, 2003, 39(19): 1379-1380
- [16] Huang K, Wu Z, Wang Q. Image enhancement based on the statistics of visual representation [J]. Image and Vision Computing, 2005, 23(1): 51-57
- [17] Jinsung O. Structuring element representation of an image and its applications [J]. International Journal of Control, Automation, and Systems, 2004, 2(4): 509-515
- [18] Gargantini I. An Effective Way to Represent Quadrees [J]. Communications of the ACM, 1982, 25(12): 905-910

(上接第 207 页)

对于每一帧输入图像 A , 计算其在背景特征空间 U_d 上的投影, 然后在图像空间重建图像。在每个像素点 (x, y) 处, 计算重建图像和原始图像的欧氏距离 $d(x, y) = \|A(x, y) - \tilde{A}(x, y)\|$, 如果该距离大于某个阈值, 即 $d(x, y) > T$ 成立, 则该像素点 (x, y) 被认为是前景目标, 否则就被认为是背景点。

第 3 步 更新子空间

每累积 N 帧新图像, 按 (6) 式更新协方差矩阵并对其进行特征值分解而得到新的投影矩阵。

第 4 步 转到步骤 2。

4 实验结果与分析

为了验证算法的有效性, 我们选择了两段有较多场景变化的视频来进行测试。第一段视频是海边的场景, 包括海浪的起伏、草的摆动、光照的变化以及地面的阴影等。第二段视频是喷泉的场景, 包括水的运动、光线的变化和树叶的运动引起的场景变化。两段视频的格式为: 分辨率 320×240 , 每秒 30 帧, RGB 3 通道。

为了消除光照等因素对图像的影响, 克服不同图像之间的平均灰度差异, 这里首先将输入图像由初始的 RGB 色彩空间转换到一个归一化的色彩空间, 转换方法如下: $s=R+G+B$, $r=R/s$, $g=G/s$, 此外, 为了进一步降低计算量, 每帧图像都分成 20×20 大小相同的图像块, 每个图像块独立运行上面提出的算法。

4.1 实验结果

经过简单形态学处理后的前景检测结果如图 1 和图 2 所示, 从实验结果可以看到, 提出的方法可以在复杂变化的动态背景下较完整地检测出运动目标。算法采用 Matlab 编程实现, 在 Pentium 4 3.0 GHz 机器上运行。在未经任何优化的情况下, 平均处理速度达到约每秒 10 帧, 经过优化后有望接近实时处理。

4.2 与其他方法的比较

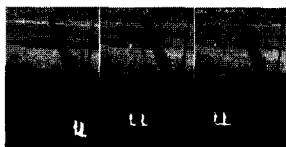


图 1 视频一结果截图



图 2 视频二结果截图

我们将提出的方法与两种主流运动目标检测方法: 混合

高斯法^[1,2]和核密度估计法^[3,4]进行了比较。实验中, 目标检测后, 三种算法都采用了简单的滤波处理。为了对三种算法的性能进行比较, 通过逐渐改变算法的阈值, 我们绘制了这三种算法的 ROC 曲线。如图 3 和图 4 所示, 曲线的横轴表示前景的误判率, 纵轴表示前景的检测率, 其中作为参照的真实前景区域事先由人工标定。从图 3 和图 4 中可以看出, 在相同的误判率下, 我们提出的算法的检测率要高于其他两种算法。

结束语 本文提出了一种基于二维主成分分析的自适应运动目标检测方法。和传统的像素点建模方法不同, 该方法采用二维主成分分析来建立背景模型, 加入了二维空间上像素点之间的相关性信息, 得到了更好的建模效果。为了能够适应不断变化的复杂环境, 提出的增量算法能够在线地更新背景模型, 并有效降低运行过程中的存储量和计算量。从实验结果来看, 在复杂变化的动态背景下, 本文提出的方法能够得到令人满意的检测性能。

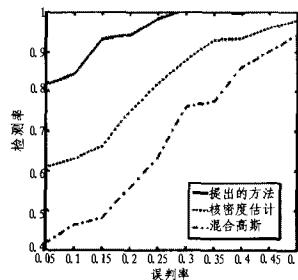


图 3 视频一算法比较 ROC 图

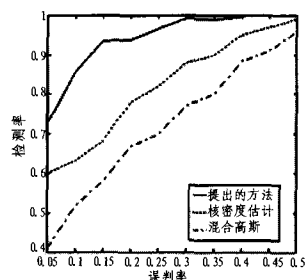


图 4 视频二算法比较 ROC 图

参考文献

- [1] Stauffer C, Grimson E. Adaptive background mixture models for real-time tracking // Proceeding of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 1999: 246-25
- [2] Friedman N, Russell S. Image segmentation in video sequences: A probabilistic approach // Proceeding of Conference on Uncertainty in Artificial Intelligence, 1997: 175-181
- [3] Mittal A, Paragios N. Motion-based background subtraction using adaptive kernel density estimation // Proceeding of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2004: 302-309
- [4] Elgammal A, Harwood D, Davis L. Non-parametric model for background subtraction // Proceeding of European Conference on Computer Vision. Springer
- [5] Yang Jian, Zhang David, Yang Jingyu. Two-Dimensional PCA: An approach to appearance based face representation and recognition. IEEE Trans Pattern Analysis and Machine Intelligence, 2004(1): 131-137