

基于增量式子空间学习的视觉跟踪系统

周仲夷, 朱远毅

(复旦大学计算机科学技术学院, 上海 200433)

摘要: 为了提高视觉跟踪方法在物体外观发生变化时的性能, 提出一种基于增量式子空间学习的视觉跟踪系统。该系统利用基于增量式主成分分析的粒子滤波方法增量式地学习一个表示跟踪结果的低维特征空间, 以反映目标物体的外观变化。实验结果表明, 当目标物体在复杂环境中承受姿态和光照变化时, 该视觉跟踪系统具有更好的性能。

关键词: 增量式子空间学习; 增量式主成分分析; 粒子滤波

Visual Tracking System Based on Incremental Subspace Learning

ZHOU Zhong-yi, ZHU Yuan-yi

(School of Computer Science and Technology, Fudan University, Shanghai 200433)

【Abstract】 In order to enhance the performance of visual tracking methods with object appearance variation, this paper proposes an visual tracking system based on incremental subspace learning. By using particle filter method based on Incremental Principal Component Analysis(IPCA), this system incrementally learns a low dimensional eigenspace representation of the tracking results to reflect appearance variation of the target object. Experiments demonstrate that the system has better performance when the target objects undergo large pose and illumination changes in some complex environments.

【Key words】 incremental subspace learning; Incremental Principal Component Analysis(IPCA); particle filtering

1 概述

视觉跟踪的主要难点来源于目标物体的外观变化。引起外观变化的内在因素包括物体位置和形状的变化, 而外在的光照条件、相机运动、相机视点以及物体遮挡等因素不可避免会造成巨大的物体外观变化。很多现有的视觉跟踪方法在跟踪开始前构造一个用于表示目标物体的模型, 并利用其不变的特征应对由光照、位置和视角等因素引起的物体外观变化。例如, 文献[1]提出的方法使用预先训练好的一组特征基来表征目标物体。这种方法假设某些特定因素(如光照条件)不会对目标物体的特征产生本质的影响, 所以, 这组预先训练好的特征基不会在跟踪过程中被更新。事实证明, 不变的特征模型不能很好地解决跟踪过程中内在和外在因素引起的外观变化问题。

本文提出了一种基于增量式子空间学习的视觉跟踪系统。利用一种增量式主成分分析(Incremental Principal Component Analysis, IPCA)方法^[2]对特征基进行持续且高效的更新。利用 IPCA 方法的好处在于: (1)低维的特征子空间可以紧凑且简洁地表征目标物体, 而并不是将目标物体看作一组互不相关像素的集合^[3]; (2)对特征基进行持续的更新反映出在跟踪过程中内在和外在因素引起的外观变化, 而这种更新的高效性保证了整个跟踪方法的效率; (3)与大多数增量式子空间学习方法^[4]不同, IPCA 方法不必假设输入样本为零均值, 可以动态地更新均值。

与文献[1]相比, 本文系统采用的方法不必在跟踪开始之前训练一组特征基, 而是在跟踪过程中进行在线的学习(和大多数算法一样, 假定第 1 帧图像中目标物体的区域是事先给定的)。另外, 相比于某些基于梯度下降的方法^[5], 本文利用一种基于粒子滤波(particle filter)的方法根据当前帧的跟踪结

果在下一帧中选取一定量的候选点, 这样可以使样本的概率分布随时间的推移而得以传递。最后, 结合一种高效的相似度度量方法, 在这些候选点中选取一个作为下一帧的跟踪结果。实验证明, 本文系统能很好地解决由目标物体的移动、旋转以及外界光照变化引起的物体外观变化问题。

2 基于增量式子空间学习的视觉跟踪系统

2.1 IPCA

IPCA 方法^[2]不必假设输入样本为零均值并可以动态地更新均值, 其描述如下:

给定一个 $d \times n$ 的矩阵 $X_n = \{x_1, x_2, \dots, x_n\}$, 其中, x_i 表示第 i 个输入样本(本文中为表示跟踪结果图像的 d 维向量)。记前 n 个输入样本的均值为 μ_n , 那么它们的协方差矩阵为

$$C_n = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_n)(x_i - \mu_n)^T \quad (1)$$

对于前 n 个输入样本, 已知按从大到小顺序排列后 C_n 的前 k 个特征值分别为 $\lambda_1^n, \lambda_2^n, \dots, \lambda_k^n$, 以及与之对应的特征向量分别为 $u_1^n, u_2^n, \dots, u_k^n$ 。当给定新输入样本 x_{n+1} 后, 需要计算新样本集 $X_{n+1} = \{X_n | x_{n+1}\}$ 协方差矩阵 C_{n+1} 的特征值 $\lambda_1^{n+1}, \lambda_2^{n+1}, \dots, \lambda_k^{n+1}$ 和特征向量 $u_1^{n+1}, u_2^{n+1}, \dots, u_k^{n+1}$ 。设 X_{n+1} 的均值为 μ_{n+1} , 则

$$\mu_{n+1} = \frac{n}{n+1} \mu_n + \frac{1}{n+1} x_{n+1} \quad (2)$$

基金项目: 国家科技支撑计划基金资助项目(2007BAH09B03); 上海市科委基金资助项目(07dz15008, 08dz1500109)

作者简介: 周仲夷(1984 -), 男, 硕士研究生, 主研方向: 计算机视觉; 朱远毅, 硕士研究生

收稿日期: 2009-06-29 **E-mail:** zhouzy@fudan.edu.cn

对于 C_{n+1} , 有

$$\begin{aligned}
C_{n+1} &= \frac{1}{n+1} \sum_{i=1}^{n+1} (x_i - \mu_{n+1})(x_i - \mu_{n+1})^T = \\
&\quad \frac{1}{n+1} \sum_{i=1}^n (x_i - \mu_n + \mu_n - \mu_{n+1})(x_i - \mu_n + \mu_n - \mu_{n+1})^T + \\
&\quad \frac{1}{n+1} (x_{n+1} - \mu_{n+1})(x_{n+1} - \mu_{n+1})^T = \\
&\quad \frac{1}{n+1} \sum_{i=1}^n (x_i - \mu_n)(x_i - \mu_n)^T + \frac{n}{n+1} (\mu_n - \mu_{n+1})(\mu_n - \mu_{n+1})^T + \\
&\quad \frac{1}{n+1} (x_{n+1} - \mu_{n+1})(x_{n+1} - \mu_{n+1})^T = \\
&\quad \frac{1}{n+1} \sum_{i=1}^n (x_i - \mu_n)(x_i - \mu_n)^T + \frac{n}{(n+1)^3} (x_{n+1} - \mu_n)(x_{n+1} - \mu_n)^T + \\
&\quad \frac{n^2}{(n+1)^3} (x_{n+1} - \mu_n)(x_{n+1} - \mu_n)^T = \\
&\quad \frac{n}{n+1} C_n + \frac{n}{(n+1)^2} (x_{n+1} - \mu_n)(x_{n+1} - \mu_n)^T \quad (3)
\end{aligned}$$

令 $U_n = \{u_1^n, u_2^n, \dots, u_k^n\}$, $A_n = \text{diag}\{\lambda_1^n, \lambda_2^n, \dots, \lambda_k^n\}$, 那么 C_n

可以近似地表示为

$$C_n \approx U_n A_n U_n^T \quad (4)$$

将式(4)代入式(3), 得到

$$\begin{aligned}
C_{n+1} &\approx \frac{n}{n+1} U_n A_n U_n^T + \frac{n}{(n+1)^2} (x_{n+1} - \mu_n)(x_{n+1} - \mu_n)^T \\
&= \frac{n}{n+1} \sum_{i=1}^k \lambda_i^n u_i^n u_i^{nT} + \frac{n}{(n+1)^2} (x_{n+1} - \mu_n)(x_{n+1} - \mu_n)^T \quad (5)
\end{aligned}$$

令

$$y_i^n = \sqrt{\frac{n\lambda_i^n}{n+1}} u_i^n, \quad i=1, 2, \dots, k \quad (6)$$

$$y_{k+1}^n = \frac{\sqrt{n}}{n+1} (x_{n+1} - \mu_n) \quad (7)$$

$$A_n = \{y_1^n, y_2^n, \dots, y_{k+1}^n\} \quad (8)$$

结合式(5)可以得到

$$C_{n+1} \approx A_n A_n^T \quad (9)$$

即问题转换成了求 $A_n A_n^T$ 的特征值和特征向量。注意到 $A_n A_n^T$ 是一个 $d \times d$ 的矩阵, 为了加快特征分解的速度, 这里对规模较小的 $(k+1) \times (k+1)$ 矩阵(在实际应用中, k 的值一般很小)

$$B_{n+1} = A_n^T A_n \quad (10)$$

进行特征分解, 得到 B_{n+1} 的 k 个特征值 $\lambda_1^{n+1}, \lambda_2^{n+1}, \dots, \lambda_k^{n+1}$ 和特征向量 $v_1^{n+1}, v_2^{n+1}, \dots, v_k^{n+1}$, 满足

$$B_{n+1} v_i^{n+1} = \lambda_i^{n+1} v_i^{n+1}, \quad i=1, 2, \dots, k \quad (11)$$

在式(11)的两边同时左乘 A_n , 得到

$$A_n A_n^T A_n v_i^{n+1} = \lambda_i^{n+1} A_n v_i^{n+1}, \quad i=1, 2, \dots, k \quad (12)$$

这样, 可以得到

$$u_i^{n+1} = A_n v_i^{n+1}, \quad i=1, 2, \dots, k \quad (13)$$

综上所述, 对于每一个新输入样本 x_{n+1} , IPCA 方法的步骤如下:

- (1)更新均值(式(2))。
- (2)计算 $y_1^n, y_2^n, \dots, y_k^n$ (式(6))和 y_{k+1}^n (式(7))。
- (3)构造矩阵 A_n (式(8))。
- (4)计算矩阵 B_{n+1} (式(10))。
- (5)对 B_{n+1} 进行特征分解, 得到特征值 $\lambda_1^{n+1}, \lambda_2^{n+1}, \dots, \lambda_k^{n+1}$ 和特征向量 $v_1^{n+1}, v_2^{n+1}, \dots, v_k^{n+1}$ 。

(6)计算新的特征向量 $u_1^{n+1}, u_2^{n+1}, \dots, u_k^{n+1}$ (式(13))。

2.2 动态模型和观察模型

视觉跟踪可以归结为一种隐马尔可夫模型的推导问题。假设 S_t 表示在 t 时刻目标物体的几何运动参数, 给定一组图像的观察值 $O_t = \{I_1, I_2, \dots, I_t\}$, 那么目标是要求出隐态变量 S_t 的值。根据贝叶斯定理, 可以得到

$$p(S_t | O_t) = p(I_t | S_t) \int p(S_t | S_{t-1}) p(S_{t-1} | O_{t-1}) dS_{t-1} \quad (14)$$

视觉跟踪问题由表示 2 个状态之间关系的动态模型 $p(S_t | S_{t-1})$ 以及用于估计在 S_t 状态观察到 I_t 可能性的观察模型 $p(I_t | S_t)$ 共同决定。本文系统将利用一种类似于 Condensation^[6]的方法对目标物体位置的概率分布进行建模。

目标物体在连续帧内的运动可以近似地用图像几何变换表示。这里用 4 个几何变换参数表示目标物体从状态 S_{t-1} 到状态 S_t 的变化。令 $S_t = (x_t, y_t, s_t, \alpha_t)$, 其中, x_t, y_t, s_t, α_t 分别表示在 t 时刻的 x, y 平移量、与标准大小的比例以及宽高比。 S_t 中的每一个参数都符合一个由 S_{t-1} 中对应量决定的独立高斯分布, 即

$$p(S_t | S_{t-1}) = N(S_t; S_{t-1}, \Psi) \quad (15)$$

其中, 对角阵 Ψ 的对角线元素 $\sigma_x^2, \sigma_y^2, \sigma_s^2, \sigma_\alpha^2$ 分别表示各个参数的方差。

本文使用主成分分析的一种概率描述^[7]对目标物体的图像观察值进行建模。假设观察图像 I_t 可以用由 U 构成并以 μ 为中心的子空间生成, 那么某个样本由这个子空间生成的概率反比于该样本到该子空间中心点的距离 d 。距离 d 可以分解为样本到子空间的距离 d_i 和样本在子空间的投影样本到子空间中心点的距离 d_w 。这种在正交子空间及其补空间上的距离表示和文献[8]十分相似。状态 S_t 对应的样本由该子空间生成的概率可以表示为

$$p_{d_i}(I_t | S_t) = N(I_t; \mu, U U^T + \varepsilon I) \quad (16)$$

其中, I 为单位阵; εI 表示在高斯分布上所加的噪声。在子空间内, 此样本的投影到子空间中心点的距离可以由到均值的马氏距离表示, 那么投影样本在子空间内的相似度为

$$p_{d_w}(I_t | S_t) = N(I_t; \mu, U A^{-1} U^T) \quad (17)$$

其中, A 为对角阵, 其对角线上元素为 U 各列特征向量对应的特征值。综合两者得到

$$p(I_t | S_t) = N(I_t; \mu, U U^T + \varepsilon I) N(I_t; \mu, U A^{-1} U^T) \quad (18)$$

2.3 系统流程

下面给出完整的基于增量式子空间学习的视觉跟踪系统的流程。假设跟踪序列前 n 帧图像的跟踪结果为 S_1, S_2, \dots, S_n , 这些结果对应的输入样本 x_1, x_2, \dots, x_n 的协方差矩阵的前 k 个特征值和特征向量分别为 $\lambda_1^n, \lambda_2^n, \dots, \lambda_k^n$ 和 $u_1^n, u_2^n, \dots, u_k^n$, 对于第 $n+1$ 帧图像:

- (1)根据第 n 帧图像的跟踪结果 S_n , 在第 $n+1$ 帧图像上以一定的概率分布随机选择 m 个候选状态 $S_1^{n+1}, S_2^{n+1}, \dots, S_m^{n+1}$, 并计算它们出现的概率(式(15))。
- (2)计算各个候选状态对应的子图像成为目标物体的概率(式(18))。
- (3)计算各个候选状态为跟踪结果的概率(式(14)), 选取具有最大概率值的候选状态为最终跟踪结果 S_{n+1} 。
- (4)将跟踪结果 S_{n+1} 对应的子图像作为新的输入样本

x_{n+1} 。若 $n+1 = k$, 利用一般的特征分解方法更新特征值和特征向量; 反之, 利用 IPCA 方法更新特征值和特征向量。

图 1 描述了整个跟踪系统的流程。

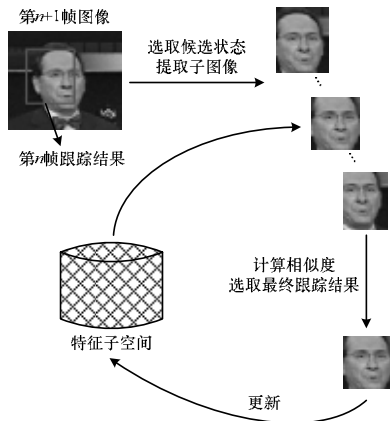


图 1 基于增量式子空间学习的视觉跟踪系统流程

3 实验

为了检验本文系统的性能,对一系列包含目标物体移动、旋转以及外界光照变化的视频数据进行跟踪实验。这些视频数据的分辨率是 320×240 , 帧率为每秒 15 帧。为了方便特征子空间的表示,将每帧图像跟踪结果对应的子图像转化为 48×48 标准大小的灰度图像。出于跟踪精度与速度的考虑,选用 16 个特征向量表示特征子空间,并在下一帧中随机选取 200 个候选样本点。

在实际应用中,由于目标物体的外观随着时间的变化而变化,新输入样本往往最能够代表目标物体当前的外观特征,因此希望新输入样本能够在跟踪过程中对整个特征子空间产生较大的影响。可以通过增加新输入样本的权重或减小旧输入样本的权重达到这个目的。假设遗忘因子为 f , 那么当给定新输入样本 x_{n+1} 之后,新样本集变为 $X_{n+1} = \{fX_n | x_{n+1}\}$ 。此时,更新特征子空间时应做一些改动。新的均值为

$$\mu_{n+1} = \frac{fn}{n+1} \mu_n + \frac{1}{n+1} x_{n+1} \quad (19)$$

同时,应将 fx_1, fx_2, \dots, fx_n 代替原来的作为旧输入样本。

另外,虽然引入遗忘因子将导致旧输入样本的变化,但是其协方差矩阵的特征值和特征向量维持不变。

3.1 实验结果

本文对 2 段人脸视频进行了跟踪实验。图 2 展示了对含有人脸平移和水平旋转运动的视频的跟踪结果。结果说明本文系统能有效地应对目标物体的运动。



图 2 含有人脸平移和水平旋转运动的人脸视频跟踪结果

如图 3 所示,第 2 个视频在含有人脸平移和水平旋转运动的同时,还包含外界光照的变化。从跟踪结果可以看到,本文系统还能很好地解决环境光照的问题。



图 3 含有外界光照变化的人脸视频跟踪结果

3.2 结果分析

本文系统的成功归结于多方面的原因。众所周知,具有姿态变化的目标物体的外观能够用基于外观的表示方法很好地建模^[9]。同时,对于某个固定的姿态,不同光照条件下目标物体也能用低维特征子空间很好地近似^[10]。实验证明本文系统能很好地用在线的学习方法解决由姿态和光照变化引起的外观变化,而不需要任何在跟踪任务开始之前的训练过程。另外,由于加入了遗忘因子,因此较早的跟踪结果对整个特征子空间的影响将随着时间的推移慢慢减弱,这也是本文系统能很好解决外观变化问题的另一个重要原因。

4 结束语

本文提出了一种基于外观的视觉跟踪系统,它增量式地对一个用于表示目标物体的特征子空间进行学习且不需要任何事先的学习过程。它利用对特征子空间的学习来反映目标物体在跟踪过程中因姿态和光照条件变化引起的外观变化,从而促进整个跟踪任务的进行。而大多数现有的视觉跟踪方法假设在跟踪过程中目标物体的姿态和外界环境的光照条件不会发生变化。另外,不同于现有的大多数增量式子空间学习方法均假设样本具有零均值的特性,本文系统采用的增量式 PCA 方法可以准确有效地更新样本的均值和特征基。最后,从目标物体承受姿态和光照条件变化时的实验结果来看,本文系统确实能很有效地解决这些外观变化问题。

本文中用于选取候选样本点的动态模型是基于高斯分布的,当然动态模型也可以用更有效的参数估计方法来训练。本文系统还可以扩展为对多个特征子空间进行学习,从而使跟踪结果更加精确。

参考文献

- [1] Black M J, Jepson A D. Eigentracking: Robust Matching and Tracking of Articulated Objects Using View-based Representation[C]//Proceedings of European Conference on Computer Vision. Cambridge, UK: Springer, 1996: 329-342.
- [2] Li Yongmin, Xu Liqun, Jason M, et al. On Incremental and Robust Subspace Learning[J]. Pattern Recognition, 2004, 37(7): 1509-1518.
- [3] Adelson E H, Bergen J R. The Plenoptic Function and the Elements of Early Vision[M]//Landy M, Movshon J A. Computational Models of Visual Processing. [S. l.]: MIT Press, 1991.
- [4] Golub G H, van Loan C F. Matrix Computations[M]. [S. l.]: Johns Hopkins University Press, 1996.
- [5] Comaniciu D, Ramesh V, Meer P. Real-time Tracking of Non-rigid Objects Using Mean Shift[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Hilton Head, SC, USA: IEEE Computer Society, 2000: 142-149.

(下转第 200 页)