

文章编号:1001-893X(2012)08-1291-07

基于粒子滤波框架联合仿射和外貌模型的目标跟踪^{*}

向志炎¹, 曹铁勇², 潘竞峰²

(1. 解放军理工大学 通信工程学院, 南京 210007; 2. 解放军理工大学 指挥自动化学院, 南京 210007)

摘 要: 针对视频序列中目标的跟踪问题, 提出了一种基于粒子滤波框架的联合仿射和外貌模型的目标跟踪算法。该算法首先提取图像帧之间的相关特征点, 通过求解 Sylvester 方程得到仿射参数, 然后将仿射参数嵌入到基于仿射群的粒子滤波框架中进行平滑估计。利用基于仿射群的一阶自回归过程模拟状态的变化, 联合仿射特征点模型和外貌模型进行似然估计, 得到粒子的最佳平均状态, 进而对目标实施跟踪。实验结果表明, 在目标经历姿势和尺度变化、遮挡以及复杂背景等情况下, 提出的算法能够有效地跟踪目标, 较之其他相关算法具有很强的鲁棒性。

关键词: 目标跟踪; Sylvester 方程; 粒子滤波; 仿射群; 仿射模型

中图分类号: TN911.73 **文献标志码:** A doi:10.3969/j.issn.1001-893x.2012.08.014

Object Tracking Based on Particle Filtering Framework Joint Affine Model and Appearance Model

XIANG Zhi-yan¹, CAO Tie-yong², PAN Jing-feng²

(1. Institute of Communications Engineering, PLA University of Science and Technology, Nanjing 210007, China;

2. Institute of Command Automation, PLA University of Science and Technology, Nanjing 210007, China)

Abstract: An algorithm based on particle filtering framework joint affine model and appearance model is introduced in this paper for object tracking in video sequences. The affine parameters for pose estimation from the corresponding feature points, which are extracted between two successive images, can be formed as a solution to Sylvester's equation. Then, the affine parameters can be smoothly estimated within the particle filtering framework based on the affine group. The state dynamic is modeled via the first order autoregressive (AR) process on the affine group. And the optimal mean state of particles is estimated through the total likelihood function which is a combination of affine feature model and appearance model. Thus, object tracking can be actualized via particle filtering based on the affine group. Experimental results demonstrate the proposed method is more effective and robust compared with other algorithms when the tracked object undergoes pose and scale changes, occlusion and complex background.

Key words: object tracking; sylvester's equation; particle filtering; affine group; affine model

1 引 言

近几十年来, 目标跟踪引起越来越广泛的关注, 已成为视觉领域的一个研究热点^[1]。在许多诸如视频监控、交互式会议、人机交互、机器定位等视觉应

用领域中, 目标的跟踪问题都是亟待解决的主要任务之一, 具有重要的研究价值和应用前景。

跟踪主要是监测视频序列中目标的位置变化, 通常包括目标的旋转及其平移^[2]。一般是在矢量空间, 通过局部坐标点的确定与匹配, 得到目标的旋转

^{*} 收稿日期: 2012-02-20; 修回日期: 2012-03-22

和平移等仿射参数,进而对目标实施跟踪。然而,常规的仿射变换并不是一个矢量空间,而是一个处理李群(仿射群)结构的曲线空间^[3]。选择一组局域坐标点并且利用矢量空间的方法,是很难得到理想的结果(依赖于坐标点的选择)。基于这些考虑,文献[4]利用目标的几何结构定义了一个基于仿射群的状态方程,得到的跟踪结果更加健壮,鲁棒性更好。

在 Isard 和 Blake^[5]的开创性研究之后,粒子滤波应用于大量的视觉跟踪问题^[6-8]之中。这是由于在现实世界里,目标物体和其所在的环境都是随时间的变化而变化的,而使用蒙特卡罗的方法^[9]能够处理这种非线性、非高斯以及非静态的滤波问题。研究表明,这种基于粒子滤波的方法是非常成功的。近年来,基于粒子滤波框架的目标跟踪方法层出不穷,根据其利用目标的特性不同,这些方法主要分为两大类:一类是利用目标的特征点^[2,10],另一类是利用目标的外貌模型^[11]。基于特征点的方法主要依赖于不同图像间的相关特征点,进行跟踪参数的估计,这种方法能对特征比较明显的目标进行比较好的跟踪,一般是处理环境比较简单的跟踪问题;基于外貌模型的方法虽然能很好地适应环境的变化,但是却忽视了目标本身的许多特性(如角点、轮廓等)。

针对彩色视频序列的目标跟踪问题,虽然彩色直方图对于跟踪非刚体目标有诸多优势^[12],如尺度和旋转的不变性、对部分遮挡的健壮性以及计算的有效性等,但是这些还不足以处理更为复杂的问题,如大范围的遮挡、外貌的剧烈变化、复杂环境等。

基于以上考虑,为了对目标实施有效的跟踪,提出了基于粒子滤波框架的联合仿射和外貌模型的目标跟踪算法。该算法首先通过目标的特征点进行仿射参数的求解,对目标的姿势进行估计;然后将其嵌入到基于仿射群的粒子滤波框架之中,联合仿射模型和外貌模型进行似然估计,确定目标的最佳运动状态。

2 基于 Sylvester 方程的仿射参数的求解

为了对目标的姿势进行估计,首先从一个已知目标姿势开始,并用一个矩形框进行标注。然后,利用文献[13]中 SIFT(尺度不变特征转换)算法提取前后帧的相关特征点,也即匹配特征点,对于提取的特征点,由于可能存在误匹配,通过 RANSAC 方法^[14]

来消除误匹配点。由于 SIFT 算法提取的是整个图像的特征点,但是对于姿势估计,只需要保留矩形区域内的特征点,用于基于 Sylvester 方程的仿射参数的求解^[2]。

姿势估计的问题关键在于从前后两帧图像之间的匹配点求解出旋转和平移参数。首先将三维目标投影成二维目标,得到二维投影模型;然后假设 (u_t^i, v_t^i) 和 (u_{t+1}^i, v_{t+1}^i) 是第 t 帧和第 $t+1$ 帧图像中相对应的投影点,则它们运动前后的位置有如下映射关系(即仿射变换):

$$\begin{pmatrix} u_{t+1}^i \\ v_{t+1}^i \end{pmatrix} = \mathbf{R}_{2 \times 2} \begin{pmatrix} u_t^i \\ v_t^i \end{pmatrix} + \begin{pmatrix} l_x \\ l_y \end{pmatrix} \tag{1}$$

式中, $i(i=1,2,\cdots,M)$ 表示第 i 个匹配特征点, M 为矩形框内特征点的总数; (l_x, l_y) 表示的是平移矢量; $\mathbf{R}_{2 \times 2}$ 是一个 2×2 的旋转矩阵。

式(1)只是利用了一对特征点来估计旋转和平移参数的,而考虑到矩形框内所有的 SIFT 匹配特征点,可以利用最小二乘估计的方法计算仿射参数,仿射参数的估计就变成了一个约束最优化问题,利用拉格朗日乘数法对其进行求解。最终得到一个矩阵方程:

$$-\mathbf{B} + \mathbf{R}_{2 \times 2} \mathbf{A} + \mathbf{A} \mathbf{R}_{2 \times 2} = \mathbf{0} \tag{2}$$

式中, $\mathbf{B} = (\mathbf{P}_{t+1} - \bar{\mathbf{P}}_{t+1}) \mathbf{P}_t^T$, $\mathbf{A} = (\mathbf{P}_t - \bar{\mathbf{P}}_t) \mathbf{P}_t^T$, 其中

$$\mathbf{P}_t = \begin{pmatrix} U_t \\ V_t \end{pmatrix} = \begin{pmatrix} u_t^1 & u_t^2 & \cdots & u_t^M \\ v_t^1 & v_t^2 & \cdots & v_t^M \end{pmatrix}$$

为第 t 帧图像的

SIFT 特征点矩阵, $\begin{pmatrix} \bar{u}_t \\ \bar{v}_t \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^M u_t^i / M \\ \sum_{i=1}^M v_t^i / M \end{pmatrix}$ 为第 t 帧图

像特征点的均值, $\bar{\mathbf{P}}_t = \begin{pmatrix} \bar{U}_t \\ \bar{V}_t \end{pmatrix}_{2 \times M} = \begin{pmatrix} \bar{u}_t & \bar{u}_t & \cdots \\ \bar{v}_t & \bar{v}_t & \cdots \end{pmatrix}$ 为

一个 $2 \times M$ 维的均值矩阵, $\mathbf{A} = \begin{pmatrix} \lambda_1 & \lambda_3 \\ \lambda_3 & \lambda_2 \end{pmatrix}$ 是拉格朗

日乘数矩阵。由于矩阵 $\mathbf{A}$ 是未知的,在实验中,首先设定一个近似的旋转矩阵 $\mathbf{R}_{2 \times 2}$,通过迭代计算满足约束条件的 $\mathbf{A}$ 。式(2)即为著名的 Sylvester 方程。利用克氏积的方法求解 Sylvester 方程,得到旋转参数如下:

$$\text{vec}(\mathbf{R}_{2 \times 2}) = (\mathbf{I} \otimes \mathbf{A} + \mathbf{A} \otimes \mathbf{I})^{-1} \text{vec}(\mathbf{B}) \tag{3}$$

一旦通过解 Sylvester 方程确定了前后两帧图像间的旋转参数,就可以通过下式计算平移矢量:

$$\begin{pmatrix} l_x \\ l_y \end{pmatrix} = \begin{pmatrix} \bar{u}_{t+1} \\ \bar{v}_{t+1} \end{pmatrix} - \mathbf{R}_{2 \times 2} \begin{pmatrix} \bar{u}_t \\ \bar{v}_t \end{pmatrix} \tag{4}$$

3 基于仿射群的目标跟踪

利用 SIFT 特征点,通过 Sylvester 方程计算仿射参数,估计出前后两帧图像之间的目标姿势的变化。而当前目标的状态是对过去所有状态积累更新得到的,这就可能会在状态更新的过程中产生误差累积的问题。因此,通过采用粒子滤波对仿射参数进行平滑估计,以避免误差及其产生的一些大的偏移。

3.1 基于仿射群的粒子滤波

在粒子滤波框架里,跟踪问题是利用概率密度函数对视频图像中的目标进行特征化。其概率密度函数是通过贝叶斯准则进行传递的:

$$p(X_t | O_{1:t}) \propto p(o_t | X_t) p(X_t | O_{1:t-1})$$
$$p(X_t | O_{1:t-1}) = \int p(X_t | X_{t-1}) p(X_{t-1} | O_{1:t-1}) dX_{t-1}$$

(5)

式中, X_t 表示在时刻 t 被跟踪目标的状态变量, $O_{1:t} = \{o_1, o_2, \dots, o_t\}$ 表示的是一组观测值。跟踪过程是由观测模型 $p(o_t | X_t)$ 和相邻时刻两状态之间的动态模型 $p(X_t | X_{t-1})$ 来模拟。在这里, X 是一个二维仿射变换矩阵:

$$X = \begin{pmatrix} R & L \\ 0 & 1 \end{pmatrix}$$

(6)

式中, R 是一个非奇异的 2×2 的矩阵, $L \in \mathbb{R}^2$ 。所有的仿射变换构成仿射群,即李群 Lie group(G)。

仿射群 G 是具有群结构的可微流形^[4],具有群的乘积和逆运算,并且该群的单位元素 I 的切空间形成一个李代数 Lie algebra($\mathfrak{g}$),具有如下矩阵结构:

$$X = \begin{pmatrix} x_1 & x_3 & x_5 \\ x_2 & x_4 & x_6 \\ 0 & 0 & 0 \end{pmatrix}$$

(7)

式中, $x_i \in \mathbb{R}$, 且李代数等价于六维矢量空间。

仿射群之间的距离可以通过测地线(测地线是两点之间的最小长度曲线)的长度来测量。对于单位元素点 I , 给定任一矢量 $x \in \mathfrak{g}$, 唯一的测地线将被确定。李群 G 及其李代数 $\mathfrak{g}$ 之间是通过指数映射相对应的, $\exp: \mathfrak{g} \rightarrow G$ 。其逆映射为, $\lg: G \rightarrow \mathfrak{g}$ 。假定 $x \in \mathfrak{g}$ 和 $X \in G$, 相应的计算公式为

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

和

$$\lg X = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} (I - X)^k$$

那么,两个群元素 $X_1, X_2 \in G$ 之间的测地线距离可被定义为

$$\rho(X_1, X_2) = \|\lg(X_1^{-1}X_2)\|$$

(8)

为了计算基于仿射群的采样均值 $\bar{X}_t$, 假设给定一组随机采样点 $X_{t-1}^j (j = 1, 2, \dots, N)$, 通过似然估计得到与权值 w_t^j 相对应的 X_t^j , 并在每个时间步骤 t 时刻对采样粒子 X_t^j 进行重采样。对于旋转参数 $\{R_t^1, R_t^2, \dots, R_t^N\}$, 由于重采样的粒子都比较相似, 因此, 基于仿射群的旋转参数的采样均值可以近似地计算得到:

$$\bar{R}_t = R_{t, \max} \cdot \exp\left(\frac{1}{N} \sum_{j=1}^N \lg(R_{t, \max}^{-1} R_t^j)\right)$$

(9)

式中, $R_{t, \max}$ 是对应于粒子重采样之前的最大权值的旋转矩阵, N 为采样粒子数。而平移矢量的采样均值 $\bar{L}_t$ 可以简单地通过 L_t^j 的算术平均得到。因此, 采样粒子的平均状态为: $\bar{X}_t = \begin{bmatrix} \bar{R}_t & \bar{L}_t \\ 0 & 1 \end{bmatrix}$ 。

3.2 基于仿射群的动态方程

假定状态传递方程模型是基于仿射群的一阶自回归(AR)过程^[3], 则状态的动态方程可表示为

$$X_t = X_{t-1} \cdot \exp(A_{t-1} + dW_t)$$

(10)

$$A_{t-1} = a \cdot \lg(X_{t-2}^{-1} X_{t-1})$$

(11)

式中, a 是 AR 过程的参数; $dW_t = \sum_{i=0}^6 w_{t,i} E_i$ 是基于李代数 $\mathfrak{g}$ 的高斯噪声, w_t 是从高斯分布 $N_p(0, \Sigma_p)$ 中采样得到, 并且 $w_t = \{w_{t,1}, w_{t,2}, \dots, w_{t,6}\}$, $\Sigma_p = \text{diag}(\sigma_s^2, \sigma_a^2, \sigma_\theta^2, \sigma_\phi^2, \sigma_x^2, \sigma_y^2)$ 。 E_i 为李代数 $\mathfrak{g}$ 的基本元素:

$$E_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$E_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, E_4 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$E_5 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, E_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

(12)

与每个 E_i 相对应的几何变换模式分别是尺度、面比、旋转、倾斜角、 x 方向和 y 方向的平移。

因此, 状态的测量方程可表示为

$$p(X_t | X_{t-1}) = N(X_t, \Sigma_q)$$

(13)

式中, Σ_q 是一个六维的零均值的高斯白噪声的对

角协方差矩阵。

3.3 观测似然函数

观测似然概率 $p(o_t | X_t^j)$ 描述的是一个粒子接近其真实状态的程度。通过使用彩色特性,基于概率主成分分析的外貌模型和特征模型这 3 个单元进行似然估计。记 $p_{\text{color}}^{(j)}$ 、 $p_{\text{PPCA}}^{(j)}$ 和 $p_{\text{feature}}^{(j)}$ 分别是基于彩色、外貌模型和仿射模型(特征点模型)的第 j 个采样的概率。为了简便,假设这 3 个观测似然概率值是相互独立的,那么,可以通过将这 3 个观测似然概率的相乘得到总体似然函数:

$$p_{\text{total}}^{(j)} = p(o_t | X_t^j) = p_{\text{color}}^{(j)} \times p_{\text{PPCA}}^{(j)} \times p_{\text{feature}}^{(j)} \tag{14}$$

3.3.1 彩色似然单元

彩色分布对非刚体、旋转以及部分遮挡具有鲁棒性,因而可以被用作目标模型进行似然估计^[12]。假设彩色分布被离散地分成 m bin。然后,用函数 $H(l_j)$ 生成彩色直方图($H(l_j)$ 函数是将在位置 l_j 的色彩分配到相应的 bin 中)。本文的彩色直方图是基于 HSV 颜色空间使用 $6 \times 6 \times 6$ bin 生成的。记 $p_t = \{p_t^i\}_{i=1,2,\dots,N}$ 为在时间步骤 t 时刻的归一化的模型直方图和 $q_t = \{q_t^i\}_{i=1,2,\dots,N}$ 为在时间步骤 t 时刻的归一化的粒子直方图。直方图分布 p 和 q 之间的一种比较流行的测量方法就是采用 Bhattacharyya 系数:

$$\rho[p, q] = \sum_{j=1}^N \sqrt{p_j q_j} \tag{15}$$

测量值 ρ 越大,表示这两个直方图分布就越相似。如果这两个归一化的直方图相同,那么它们之间的 ρ 值就等于 1,并且表明它们完全匹配。因而,两分布之间的相似度就可以定义一个距离来进行测量:

$$d = \sqrt{1 - \rho[p, q]} \tag{16}$$

因此,距离越小,表示分布越相似,对应地有更大的似然概率:

$$p_{\text{color}}^{(j)} = \frac{1}{\sqrt{2\pi}\sigma_c} e^{-\frac{d^2}{2\sigma_c^2}} = \frac{1}{\sqrt{2\pi}\sigma_c} e^{-\frac{(1-\rho[p,q])}{2\sigma_c^2}} \tag{17}$$

式中, $p_{\text{color}}^{(j)}$ 是方差为 σ_c^2 的高斯函数。

3.3.2 基于 PPCA 的外貌模型似然单元

目标的外貌由于许多外在的因素(如光照变化、照相机移动、照相机视角不同以及遮挡)和内在的因素(如姿势变化和形状畸变)都可能会发生剧烈的变化。因此,为了应对这些变化并实现有效跟踪,使用基于概率主成分分析的外貌模型进行目标跟踪^[11]。给定一个新的由状态 X_t 预测的观测值 o_t ,那么这个

o_t 可以看成是从目标外貌模型特征向量 U_s 张成的一个子空间所得,且该子空间的中心即为均值 μ_s ,因而,外貌子空间与观测值之间的距离可以定义为 $\text{Distance}(o_t) = \|o_t - \mu_s - U_s U_s^T(o_t - \mu_s)\|^2$ (18)

因此,由外貌子空间产生的采样点 j 的似然概率为

$$p_{\text{PPCA}}^{(j)} = \exp(-\text{Distance}(o_t)) \tag{19}$$

同时,为了降低噪声的影响,在计算 $p_{\text{PPCA}}^{(j)}$ 时,采样归一化的误差范数 $\rho(x, \sigma) = \frac{x^2}{\sigma^2 + x^2}$ 代替欧式范数 $\|x\|^2$ 去除异常值。

3.3.3 仿射模型似然单元

通过基于仿射群的自回归过程对目标的状态进行预测,仿射参数也会随着状态的变化而变化。设 $R_{2 \times 2E}^{(j)} (j = 1, 2, \dots)$ 为第 j 个粒子变换矩阵, $l_E^{(j)}$ 是根据式(4)计算得到的平移参数。为了计算仿射特征点模型的似然概率 $p_{\text{feature}}^{(j)}$,首先将第 t 帧图像的特征点投影到第 $t + 1$ 帧图像上,然后计算投影点 $(u_{t+1E}^{(j)i}, v_{t+1E}^{(j)i})$:

$$\begin{pmatrix} u_{t+1E}^{(j)i} \\ v_{t+1E}^{(j)i} \end{pmatrix} = R_{2 \times 2E}^{(j)} \begin{pmatrix} u_t^{(j)i} \\ v_t^{(j)i} \end{pmatrix} + \begin{pmatrix} l_{xE}^{(j)} \\ l_{yE}^{(j)} \end{pmatrix} \tag{20}$$

式中, i 表示的是特征点的指针。由于 SIFT 算法提取的特征点通常很多,这就很难保证所有相关的特征点之间相匹配,也即是,对于每一个 i ,都有 $u_{t+1E}^{(j)i} = u_{t+1}^{(j)i}$ 和 $v_{t+1E}^{(j)i} = v_{t+1}^{(j)i}$ 。所以,可以清晰地观察到,在第 $t + 1$ 帧图像里,投影点与相应的匹配点之间的距离越小,相应的权值越大,就越接近真实状态。因此,仿射模型的似然概率为

$$p_{\text{feature}}^{(j)} = \sum_{i=1}^M \frac{1}{\sqrt{(u_{t+1E}^{(j)i} - u_{t+1}^{(j)i})^2 + (v_{t+1E}^{(j)i} - v_{t+1}^{(j)i})^2}} \tag{21}$$

4 实验结果

为了评估所提出的算法的有效性和鲁棒性,选取了几段视频分别进行验证。首先手动地初始化目标跟踪的矩形框。在仿真实验中,考虑到跟踪区域像素会随着矩形窗的变化而变化的问题,同时为了特征表示,每个目标区域都被重新调整为 32×32 大小的图像块。对于基于概率主成分分析的外貌模型,选取的特征向量的个数为 16,考虑到效率与效果的平衡,其遗忘因子设为 0.95,用于特征向量更

新的图像块大小设为 5。每帧图像随机采样的粒子数设为 400(除特别说明外),跟踪的目标用矩形框表示。

图 1 是 PetsD1TeC1 序列中白色汽车的跟踪结果,共选取了 200 帧,其图像分辨率为 384×288 像素,帧率为 30 frame/s。这是固定摄像机拍摄的视频,从图中可以观察到,目标由远及近,尺度逐渐变

大;在第 133 帧,灯杆轻微地遮挡了目标;存在与目标外貌及其相似的背景(第 8、72、98、133 帧)。从矩形跟踪框可以看到,虽然目标的外貌没有发生大的改变并且背景中存在大量与之类似的目标物体,但是所提出的算法能够利用目标的特征点变化这一特性,很好地处理了这些变化,达到了预期的实验效果。



图 1 PetsD1TeC1 序列的跟踪结果
Fig.1 The tracking results of PetsD1TeC1 sequence

图 2 是直升机序列的跟踪结果,选取了 780 帧,其图像分辨率为 320×240 像素,帧率为 15 frame/s。图 2 中,直升机模型以变化的速度随机移动,并且经历较大的尺度和姿势的变化、遮挡以及复杂背景,如跟踪目标的姿势和尺度的变化(第 16、142、283、756 帧),目标经历了严重遮挡(从第 280 到 283 帧,其中,第 281、282 帧目标几乎完全被人头遮住)和复杂背景(第 280、283、498、552、593 帧,其中,在第 552、

593 帧图像中,由于目标与背景的颜色非常接近,导致跟踪的目标很难被发现),这些情况都使得我们的跟踪任务具有相当的挑战性。然而,从图 2 的实验结果可以看出,本文的跟踪方法能很好地处理这些问题,取得了非常好的实验效果。这是因为在粒子滤波框架内基于彩色分布和 PPCA 的思想,能很好地适应环境和外貌的变化,甚至能克服先前文献中提到的目标丢失和暂时偏移等问题。



图 2 直升机序列的跟踪结果
Fig.2 The tracking results of helicopter sequence

4.1 算法有效性验证

为了验证在粒子滤波框架内插入 Sylvester 方程的有效性,如图 3 所示,首先将直升机序列转化成灰

度图像,通过比较可以看出,插入 Sylvester 方程的 PPCA 的跟踪方法能有效地克服遮挡(如从第 280 到 283 帧),而仅仅基于文献[11]的 PPCA 方法在经历严重遮挡之后,却丢失了目标。



图 3 基于文献[11]的跟踪结果(虚线框)和基于仿射及外貌模型的跟踪结果(实线框)
Fig.3 The tracking results of the method based on Reference[11] (dashed) and our proposed approach (solid)

同时,为了验证本文嵌入 Sylvester 方程到基于仿射群的粒子滤波框架的跟踪算法的有效性,与文献[3]中的基于仿射群的粒子滤波框架的跟踪算法进行比较。选取一段灰度视频 David 序列,其图像分辨率为 320×240 像素,帧率为 15 frame/s。在这次实验中,我们选取的采样粒子数为 200。图 4 是两种跟踪方法比较的结果,从跟踪结果可以看出,当跟

踪目标的姿势发生大的变化时,文献[3]的跟踪精度下降了(如第 192、382、418 帧),以至于发生了较大的偏移;相比较而言,本文提出的方法提供了一个更为精确的跟踪结果,这是由于通过 SIFT 特征点, Sylvester 方程能更有效地对仿射参数进行估计,使得跟踪更加稳定、准确。



图 4 基于文献[3]的跟踪结果(虚线框)和基于仿射及外貌模型的跟踪结果(实线框)
Fig.4 The tracking results of the method based on Reference[3] (dashed) and our proposed approach (solid)

4.2 算法鲁棒性验证

为了验证所提出的算法的健壮性,也与文献[12]中基于彩色粒子滤波算法的目标跟踪结果作了比较。如图 5 所示,第一行为文献[12]中算法的跟踪结果,第二行为本文所提算法的目标跟踪。两者

比较可以看出:基于彩色的粒子滤波有较大偏移(如第 18 帧与第 166 帧椭圆的位置变化),在复杂背景下丢失了所跟踪的目标(第 552 帧),并且在实验中发现算法的适应性和稳定性都不是很好;而本文提出的算法克服了偏移并对目标物体成功地实施了跟踪,具有良好的鲁棒性。



(a)基于文献[12]算法



(b)基于本文算法

图 5 基于文献[12]中的算法与本文提出的算法的跟踪比较
Fig.5 Comparison between the tracker based on Reference [12] and the proposed tracker

5 结 论

本文提出的基于粒子滤波框架的联合仿射和外貌模型的目标跟踪算法,有效地解决了复杂情况下的目标跟踪问题。首先利用 Sylvester 方程求解仿射参数,对目标的姿势进行估计,这样就可以在后面的粒子滤波过程中使得随机采样更趋于正确的状态。通过将仿射参数嵌入到粒子滤波框架,利用基于仿射群的粒子滤波对仿射运动参数进行平滑估计。充分利用外貌模型和仿射特征点模型的优势进行似然估计,得到了最佳的采样均值,提高了滤波的效果。通过实验分析,并与其他相关文献算法对比,本文算法的有效性和鲁棒性得到了很好的验证。而对于跟踪目标丢失或进出视频画面的情况,本文算法由于是手动标定目标,难于继续对目标实施跟踪,同时,为了适应实际应用的需要,在后续的研究工作中,打算引入对目标的自动检测,以提高目标跟踪算法的适应性和实用性。

参考文献:

- [1] Yilmaz A, Javed O, Shah M. Object tracking: a survey[J]. ACM Computing Surveys, 2006, 38(4): 1 – 45.
- [2] Chen Chong, Schonfeld D. A particle filtering framework for joint video tracking and pose estimation [J]. IEEE Transactions on Image Processing, 2010, 19(6): 1625 – 1634.
- [3] Kwon J, Park F C. Visual tracking via particle filtering on the affine group [J]. The International Journal of Robotics Research, 2010, 29(2 – 3): 198 – 217.
- [4] Li Min, Chen Wei, Huang Kaiqi, et al. Visual tracking via incremental self – tuning particle filtering on the affine group [C]//Proceedings of 2010 IEEE conference on Computer Vision and Pattern Recognition. San Francisco: IEEE, 2010: 1315 – 1322.
- [5] Isard M, Blake A. Condensation – conditional density propagation for visual tracking[J]. International Journal of Computer Vision, 1998, 29(1): 5 – 28.
- [6] Ferrari V, Tuytelaars T, Van Gool L. Real – time Affine Region Tracking and Coplanar Grouping [C]// Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Hawaii: IEEE, 2001: 226 – 233.
- [7] Pan Pan, Schonfeld D. Video Tracking Based on Sequential Particle Filtering on Graphs [J]. IEEE Transactions on Image Processing, 2011, 20(6): 1641 – 1651.
- [8] Djuric P M, Vemula M, Bugallo M F. Target Tracking by Particle Filtering in Binary Sensor Networks[J]. IEEE Transactions on Signal Processing, 2008, 56(6): 2229 – 2238.
- [9] Kitagawa G. Monte carlo filter and smoother for non – Gaussian nonlinear state space models [J]. Journal of Computational and Graphical Statistics, 1996, 5(1): 1 – 25.
- [10] Singh Manku G, Jain P, Aggarwal A, et al. Object tracking using affine structure for point correspondences [C]// Proceedings of 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Puerto Rico: IEEE, 1997: 704 – 709.
- [11] Ross D, Lim J, Lin R S, et al. Incremental learning for robust visual tracking[J]. International Journal of Computer Vision, 2008, 77(1): 125 – 141.
- [12] Nummiaro K, Meierb E K, Van Gool L. An adaptive color – based particle filter[J]. Image and Vision Computing, 2002, 21(11): 99 – 110.
- [13] Lowe D G. Distinctive image features from scale – invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91 – 110.
- [14] Fischler M A, Bolles R C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography [J]. Communications of the ACM, 1981, 24(6): 381 – 395.

作者简介:

向志炎(1983—),男,江西九江人,2007年于南京解放军理工大学获电子工程专业学士学位,现为硕士研究生,主要研究方向为语音与图像处理;

XIANG Zhi-yan was born in Jiujiang, Jiangxi Province, in 1983. He received the B.S. degree from PLA University of Science and Technology in 2007. He is now a graduate student. His research concerns voice and image processing.

Email: xzy831018@163.com

曹铁勇(1971—),男,江苏南京人,2004年于南京解放军理工大学获信号与信息处理专业博士学位,现为教授,主要研究方向有视频编码、语音与图像处理和指纹识别等。

CAO Tie-yong was born in Nanjing, Jiangsu Province, in 1971. He received the Ph.D. degree from PLA University of Science and Technology in 2004. He is now a professor. His research concerns video coding, voice and image processing and fingerprint identification.