

doi: 10.3788/gzxb20124110.1230

一种鲁棒的目标跟踪方法

贾伟¹, 孙伟², 李大健¹

(1 西北工业大学 第 365 研究所, 西安 710065)

(2 西安电子科技大学 机电工程学院智能控制与图像工程研究所, 西安 710071)

摘 要:针对传统特征光流场跟踪方法中由于误差积累和错误匹配而导致的特征点丢失问题, 基于一种新的 Harris-SIFT 特征点表示方法, 提出基于预测帧与关键帧的算法框架, 实现了光流场运动估计与局部特征识别相结合的目标跟踪方法. 预测帧利用塔式分解和递归算法计算特征点的光流场运动矢量, 使用运动矢量直方图获取目标的运动矢量, 并剔除误匹配点; 当特征点数量小于 5 个时, 关键帧使用 Harris-SIFT 特征点进行局部特征匹配, 利用仿射模型对目标精确定位及姿态修正. 实验结果表明, 本方法对视频序列中的纹理特征目标跟踪的鲁棒性较好, 在背景复杂、目标遮挡或暂时丢失情况下, 仍可以继续完成目标的可靠跟踪.

关键词:图像处理; 目标跟踪; 鲁棒性; 光流场; 特征检测

中图分类号: TP391

文献标识码: A

文章编号: 1004-4213(2012)10-1230-6

0 引言

基于图像序列的运动目标自动跟踪是图像处理、模式识别和精确制导领域的重要课题, 在国防、工业和交通等领域有着广泛的应用. 国内外很多文献对目标的表述和跟踪方法进行了研究. 尽管 Mean shift, 粒子滤波和卡尔曼滤波等算法已广泛使用, 目标的直方图、轮廓、模板、Gabor 等特征或特征的组合被大量使用^[1-2]. 但在实际的跟踪中, 由于模型的相对固定不能适应光照背景变化和物体的大范围运动等^[3], 很难完成长序列视频目标的跟踪. 因此, 一些研究者已经注意到目标局部特征点的重要性, 并且在跟踪过程中加以运用.

局部特征方法是通过检测图像的局部特征, 如边缘、角点、轮廓等, 形成易于匹配、稳定性好的特征向量, 把目标跟踪问题转化为特征空间特征向量的匹配问题, 具有很强的鲁棒性. 因此, 局部不变特征作为目标的有效表示方法, 在目标识别跟踪领域获得了广泛的应用. 性能较好的尺度不变特征 (Scale-Invariant Feature Transform, SIFT) 表示方法, 对旋转、尺度缩放、亮度变化保持不变性, 对视角变化、仿射变换、噪音也保持一定程度的稳定性. 但是 SIFT 算法获取特征点多, 且在 3D 空间寻找局部极值, 较高的运算复杂度使其在视频中的应用受到了限制. 在视频序列中, 运动信息是目标的重要信息.

传统的光流特征跟踪方法^[4], 将局部特征与运动信息相结合, 是一种有效的目标跟踪方法. 该算法首先提取局部特征点, 利用光流场约束方程和特征点区域的运动一致性, 可以有效获得目标的运动矢量场. 但由于视频序列中的目标的变化而导致的误差积累, 及局部区域块匹配结果的不唯一性, 常常造成特征点的丢失和错误匹配.

针对以上两个问题, 文献[5]使用了 SIFT 局部特征完成目标的鲁棒跟踪, 但实时性有待提高. 文献[6]给出了基于粒子滤波的目标区域的置信度的判定方法, 取得了良好的效果, 但是其方法基于颜色直方图来获得目标相似度的测量, 对光线变化敏感. 而文献[7]使用了仿射模型来对特征点监测, 计算复杂度较高. 本文使用预测帧和关键帧的机制, 实现了光流场运动估计与目标识别的有机结合. 提出一种新的简化的局部特征表示方法 (Harris-SIFT), 用于提高目标识别的实时性能. 预测帧使用光流特征跟踪法来获取运动矢量, 通过运动矢量的统计特性获取目标的运动信息, 并剔除误匹配点. 关键帧利用 Harris-SIFT 特征点对目标特征描述, 使用仿射变换模型来锁定目标的姿态及置信区域, 同时完成目标特征更新. 研究表明, 这种弹性的表示方法, 符合人类视觉特性^[8-10], 可以灵活地处理目标遮挡、背景混杂问题.

基金项目: 教育部直属高校基本科研业务费 (No. K50511040008) 资助

第一作者: 贾伟 (1980-), 男, 工程师, 硕士, 主要研究方向为无人机任务规划与控制, 侦察信息处理等. Email: hello.jiawei@163.com

导师 (通讯作者): 孙伟 (1980-), 男, 副教授, 博士, 主要研究方向为视觉计算与嵌入式系统等. Email: sunweitom@yahoo.com.cn

收稿日期: 2012-02-27; **修回日期:** 2012-06-12

1 传统光流特征跟踪方法

假设图像 A 中的特征点 $u=[u_x \ u_y]^T$, 运动估计就是在图像 B 中找到点

$$v=u+d=[u_x+d_x \ u_y+d_y]^T$$

式中 $d=[d_x \ d_y]^T$ 为特征点 u 的位移矢量, 定义为 w_x, w_y 窗口内使残差 ϵ 函数最小的取值.

$$\epsilon(d) = \sum_{x=u_x-w_x}^{u_x+w_x} \sum_{y=u_y-w_y}^{u_y+w_y} (A(x,y) - B(x+d_x, y+d_y))^2 \tag{1}$$

在图像 A 和 B 定义下, 找平移矢量

$\bar{v}=[v_x \ v_y]^T$, 使匹配残差函数 $\epsilon(\bar{v})$ 最小, 即满足 ϵ 的一阶差分为 0 的 \bar{v} .

$$\epsilon(\bar{v}) = \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} (A(x,y) - B(x+v_x, y+v_y))^2 \frac{\partial \epsilon(\bar{v})}{\partial \bar{v}} \Big|_{\bar{v}=\bar{v}_{opt}} = [0 \ 0]$$

则 $\bar{v}_{opt}=G^{-1}\bar{b}$

其中 $G^{-1}\bar{b}$ 的定义为

$$G \cong \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$
$$b \cong \sum_{x=p_x-w_x}^{p_x+w_x} \sum_{y=p_y-w_y}^{p_y+w_y} \begin{bmatrix} \delta I & I_x \\ \delta I & I_y \end{bmatrix}$$

传统的光流特征方法中, 块匹配残差函数 ϵ 大于一定阈值可以认为这个特征点丢失. 而这个阈值难以确定. 当使用固定的参考特征块对序列的特征点进行跟踪, 特征点会很快丢失. 如果对相邻的图像对进行跟踪, 则作为参考的图像块在每次跟踪的时候进行初始化, 即使图像间的差别很小, 这个点将会在图像序列中漂移. 另一种方法是利用仿射模型来消除目标变化带来的影响^[8-9]. 这种方法也难以处理目标区域变化带来的影响, 问题的根源就在于块匹配方法无法唯一的确定匹配区域, 而导致了错误的目标匹配和跟踪结果, 在实际计算中表现为多个相似的峰值, 如图 1, 横轴方向表示以像素为单位的搜索窗口大小, 纵轴为相似度量结果. 本文对相邻图像

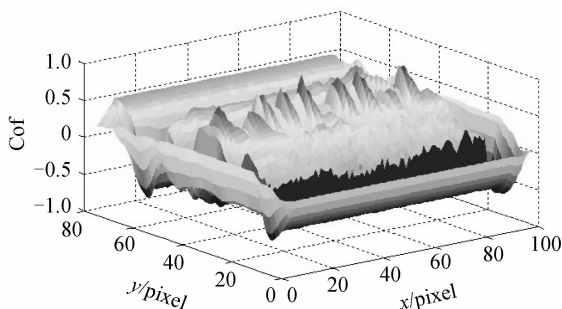


图 1 特征邻域内多个相似的峰值
Fig. 1 Multiple similar peaks of the neighborhood area

进行跟踪, 并将局部特征识别引入算法框架, 为了保证目标特征的仿射不变性及算法的实时性, 本文提出一种 Harris-sift 特征提取算法及 PKT 目标跟踪方法, 用于对目标区域的准确定位, 以消除误匹配带来的影响.

2 Harris-SIFT 局部特征表示

SIFT 算法由 D. G. Lowe 1999 年提出, 2004 年完善总结. SIFT 通过拟和三维二次函数以精确确定关键点的位置和尺度, 同时去除低对比度的关键点和不稳定的边缘响应点, 以增强匹配稳定性、提高抗噪能力. 算法步骤为: 1) 检测尺度空间极值点; 2) 精确定位极值点; 3) 为每个关键点指定方向参量; 4) 关键点描述子的生成.

局部特征在 3D 尺度空间搜索局部极值, 这个思想最早来自于 Crowley1981^[11]. 在 SIFT 方法中^[12], 使用了 DoG 滤波的塔式分解, 大于某阈值的 3D 局部极值作为特征点. 此方法可以获得大量的局部极值点, 但是计算量大, 信息冗余. 角点是二维图像亮度变化剧烈的点, 有利于图像的可靠匹配. Harris^[13]方法就是在水平、竖直两个方向上变化均较大的点. Harris 特征点具有计算简单、角点特征均匀合理、可以定量提取以及算子稳定的特点.

为了使 Harris 角点具有尺度及仿射不变性, 本文在不同尺度空间和分辨率下检测角点, 寻找局部极值点, 最后使用方向直方图进行特征点描述.

尺度自适应的 Harris 算子表述如下

$$M_{Harris}(x, \sigma, \sigma_D) = \mu(x, \sigma, \sigma_D) = \sigma_D^2 G(\sigma) * \begin{bmatrix} I_x^2(x, \sigma_D) & I_x(x, \sigma_D) I_x^L(x, \sigma_D) \\ I_x^L(x, \sigma_D) I_x^L(x, \sigma_D) & I_y^2(x, \sigma_D) \end{bmatrix} \tag{2}$$

$$C_{Harris} = \det M - \alpha \text{trace}^2 M$$

C_{Harris} 的局部极值决定了所选的 Harris 角点. σ 为积分尺度, L 为 octave 坐标, σ_D 为差分尺度, $\sigma(o) = \sigma_0 2^o, o \in [0, \dots, L-1]$. I_x 为 x 方向的梯度, I_y 为 y 方向的梯度, $G(\tilde{s})$ 为高斯模板, \det 为矩阵的行列式, trace 为矩阵直迹, α 为默认常量.

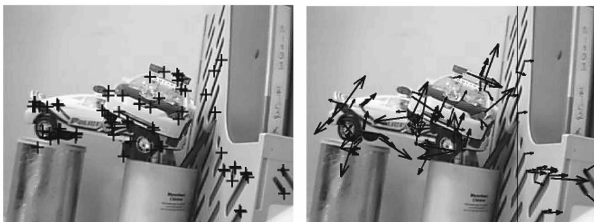
在 SIFT 算法中, 特征点主曲率通过一个 2×2 的 Hessian 矩阵 H_D 求出, 且满足

$$\text{trace}^2 H_D / \det H_D < \text{th} \tag{3}$$

由式(1)、(2)可以看出, Harris 角点满足式(1)条件. 因此 Harris-sift 算子通过如下方法构造: 1) 检测尺度空间 Harris 角点; 2) 检测为局部极值点的角点; 3) 为每个关键点指定方向参量; 4) 关键点描述子的生成.

算法首先提取 Harris 角点再进行极值的获取,

所获得的极值点也满足式(1)条件,故降低了计算量.利用特征点邻域像素的梯度方向分布特性为每个特征点指定方向参量,使算子具备旋转不变性.这种邻域方向性信息联合的思想增强了算法抗噪音的能力,同时对于含有定位误差的特征匹配也提供了较好的容错性.在实际计算时,用直方图统计邻域像素的梯度方向.一个特征点指定一个主辅方向,增强匹配的鲁棒性^[8].目标图像中的特征点,利用 Kd-tree 结构,找出其与待匹配图像中欧氏距离小于某个比例阈值的前两个特征点为匹配点.图2给出 Harris-sift 特征点及位置、所处尺度、方向.表1中,



(a) Key points that remain following an additional threshold on ratio of principal curvatures (b) Key points are displayed as vectors indicating scale, orientation and location

图2 Harris-sift 特征点及所处位置、尺度、方向
Fig.2 Location scale and orientation of Harris-sift

将相对旋转 45°.表1 实验数据表明,Harris-sift 获得了更有效的目标表示.

表1 Harris-SIFT 和 SIFT 比较

Table 1 Comparison of Harris-SIFT and SIFT

Sample	Method	SIFT	HARRIS-SIFT
I1	Feature number	321	48
	Matched ratio/(%)	69.3	48.6
	Time/ms	57.22	9.56
I2	Feature number	400	80
	Matched ratio/(%)	70.1	46.5
	Time/ms	75.70	14.72
Result	Angle/(°)	44.978 5	45.080 6
	Scale	1.000 1	1.000 1

3 预测-关键帧(PKT)跟踪方法

3.1 目标跟踪框架

当目标出现遮挡时,传统光流估计就会因为无法完成特征匹配丢失目标.而局部特征向量,对背景混杂以及视角的变换具有鲁棒性,特别适合于匹配和识别应用.可以认为,Harris-sift 获得了更“稀疏”的目标特征表示.预测-关键帧跟踪(Predictive Key Tracking, PKT)算法如图3.

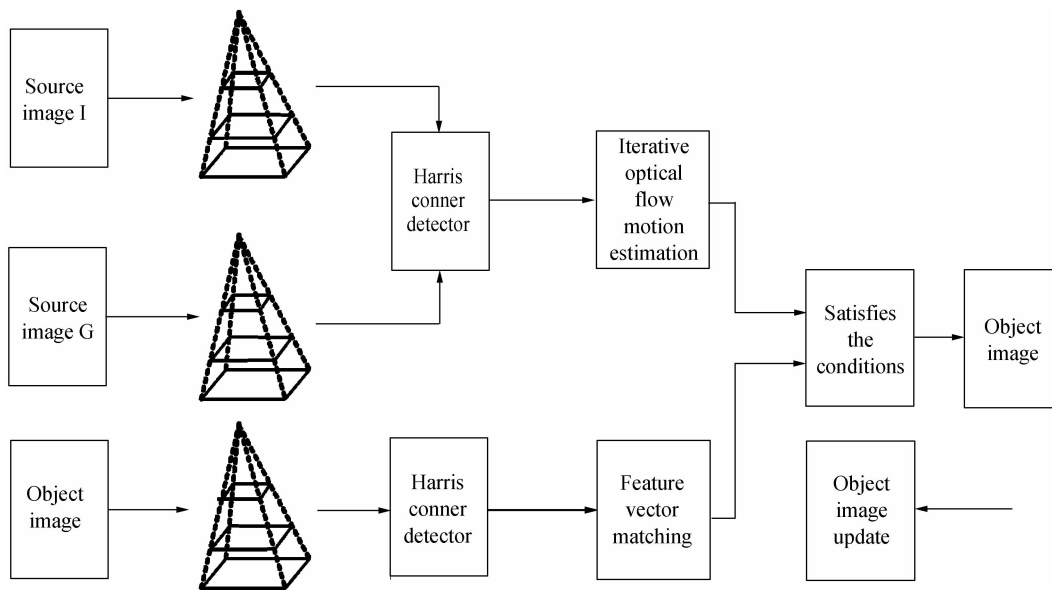


图3 目标跟踪框架
Fig.3 Proposed object tracking algorithm; PKT

算法实现步骤如下:

- 1) 首先对目标图像进行 L 级塔式分解;
- 2) 不同尺度提取特征点,获得尺度不变的 Harris-sift 特征点, $\sigma=0$ 用于特征跟踪;
- 3) 对于图像的 L 级塔式分解,通过迭代光流场 Lucas-Kanade^[4] 方法,获得运动矢量;
- 4) 通过运动矢量直方图对目标区域更新;
- 5) 目标区域特征点数量正常则返回步骤 3). 当特征点数量小于 5 时,则启动关键帧,在目标区域 1.5 倍范围内进行目标识别方法;
- 6) 根据几何或光度的约束

信息去除候选匹配点中的错配,根据仿射模型锁定目标位置.对特征点匹配区域进行跟踪特征重提取,返回步骤 3).

3.2 迭代光流场运动估计

通过迭代方法,获得运动矢量.本文首先对目标图像进行 L 级塔式分解,图像金字塔共 L 组,下一组的图像由上一组图像降采样得到.对塔式分解图像在不同尺度空间提取特征点,获得尺度不变特征点, $\sigma=0$ 用于光流运动估计,不同尺度下特征点用

于特征匹配。

取 $\sigma_0=0.5$ 为基准层尺度. 当 octave 设为 -1 的时候, 图像在计算高斯尺度空间前先使用双线性插值扩大一倍, 运动向量可以达到子像素准确度. 如果 (M_0, N_0) 是基础组 $o=0$ 的分辨率, 则其他组的分辨率由降采样获得. 图 4 给出塔式分解及 $o=0$ 的 Harris 特征角点检测结果.

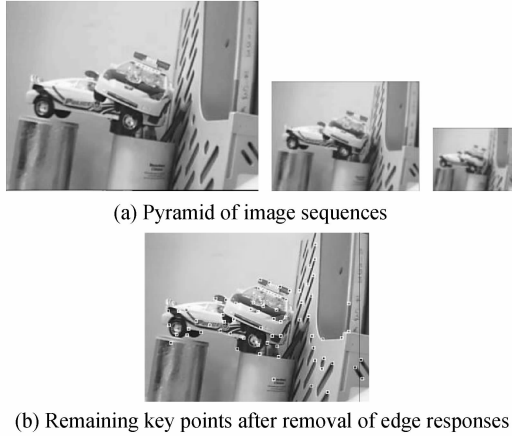


图 4 塔式分解及光流跟踪特征点提取

Fig. 4 Pyramid and remaining key points of the image

算法对第 L 级分解图像进行 K 次迭代或 L 级光流矢量 $\|\bar{\eta}^k\| < \text{threshold}$, 所得 L 级运动矢量为

$$\bar{v} = d^L = \bar{v}^K = \sum_{k=1}^K \bar{\eta}^k \quad (4)$$

进入第 $L-1$ 级分解图像, 初始运动矢量 $g^{L-1} = [g_x^{L-1} \ g_y^{L-1}]^T = 2(g^L + d^L)$, 以此类推, 最终的运动矢量 d 及目标点位置 v 为: $d = g^0 + d^0, v = u + d$.

3.3 特征点有效性判定

长序列视频跟踪中, 由于背景混杂和出现遮挡, 常造成特征点丢失, 如何准确的判定有效特征, 是目标跟踪领域的一个重要问题^[14-15].

对于预测帧, 由于帧间运动矢量较小, 根据刚性物体表面特征运动一致性原理可知, 特征点运动矢量服从高斯分布. 通过运动矢量直方图可以获得目标的运动矢量, 同时可以剔除不可靠特征点. 根据特征点相对距离统计规律, 可以获取运动目标的尺度信息. 与文献[6]使用了区域仿射相似度来剔除坏点相比较, 本算法对疑似坏点直接丢弃, 使用关键帧来保证目标跟踪的准确性, 消除了累积误差. 当目标特征点数 $f_i < N$ 时, 则启动用于目标识别的关键帧, 一般 $N=5$, 在 1.5 倍目标区域进行目标匹配.

3.4 目标识别

在关键帧的目标识别中, 特征点可以被分为内点 (inliers) 和外点 (outliers) 两类. 根据几何或光度

的约束信息去除候选匹配点中的错配. RANSAC 是在计算机视觉领域内应用最广泛的估计算法之一, 它能够在失匹配率超过 50% 时依然保持有效, 该算法在保证计算结果置信概率不变的前提下, 过滤掉大量偏差较大的模型参量. 这里定义一个误差函数

$$F = \sum_i (d(x_i, H^{-1}x'_i)^2 + d(x'_i, Hx_i)^2) \quad (5)$$

$$\text{corners} = \begin{cases} \text{inliers, if } F > \text{thresh} \\ \text{outliers, else} \end{cases} \quad (6)$$

thresh 为阈值, H 为一致性矩阵.

匹配点对点 $[x \ y]^T$ 到点 $[u \ v]^T$ 的仿射变换表示为

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (7)$$

式中 $[t_x \ t_y]^T$ 为平移参量, m_i 为旋转和尺度.

为了求解变换参量, 式(7)可以将未知变量归结为一个列矢量, 至少 3 个匹配点对可以求解.

$$\begin{bmatrix} x & y & 0 & 0 & 1 & 0 \\ 0 & 0 & x & y & 0 & 1 \\ \cdots & & & & & \\ \cdots & & & & & \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ t_x \\ t_y \end{bmatrix} = \begin{bmatrix} u \\ v \\ \vdots \end{bmatrix} \Leftrightarrow Ax = b \Rightarrow x = [A^T A]^{-1} A^T b \quad (8)$$

仿射矩阵最小均方误差解可以通过求解的正则化方程获得, 最终匹配结果如图 5.

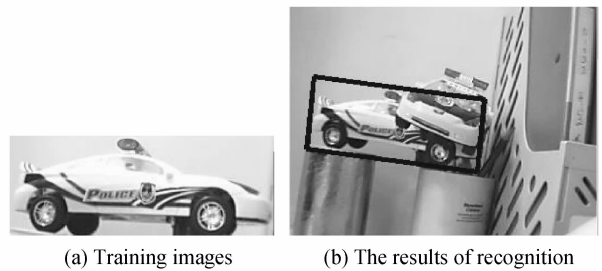


图 5 目标及 Harris-sift 特征匹配结果

Fig. 5 The training images for two objects and the results of recognition

4 实验结果

在实验中, 参量 $\sigma(o) = \sigma_0 2^o, o \in [0, \dots, L-1]$, $\sigma_{\min} = -1, L=3, w_x = w_y = 5$. 算法将运动信息和目标特征的结合, 在目标遮挡及光线变化的情况下, 仍能获得良好的跟踪结果, 图 6 为 PKT 算法对 2CAR 图像序列的跟踪结果, 实验结果表明, PKT 算法在发生遮挡后仍可以继续完成目标的跟踪.

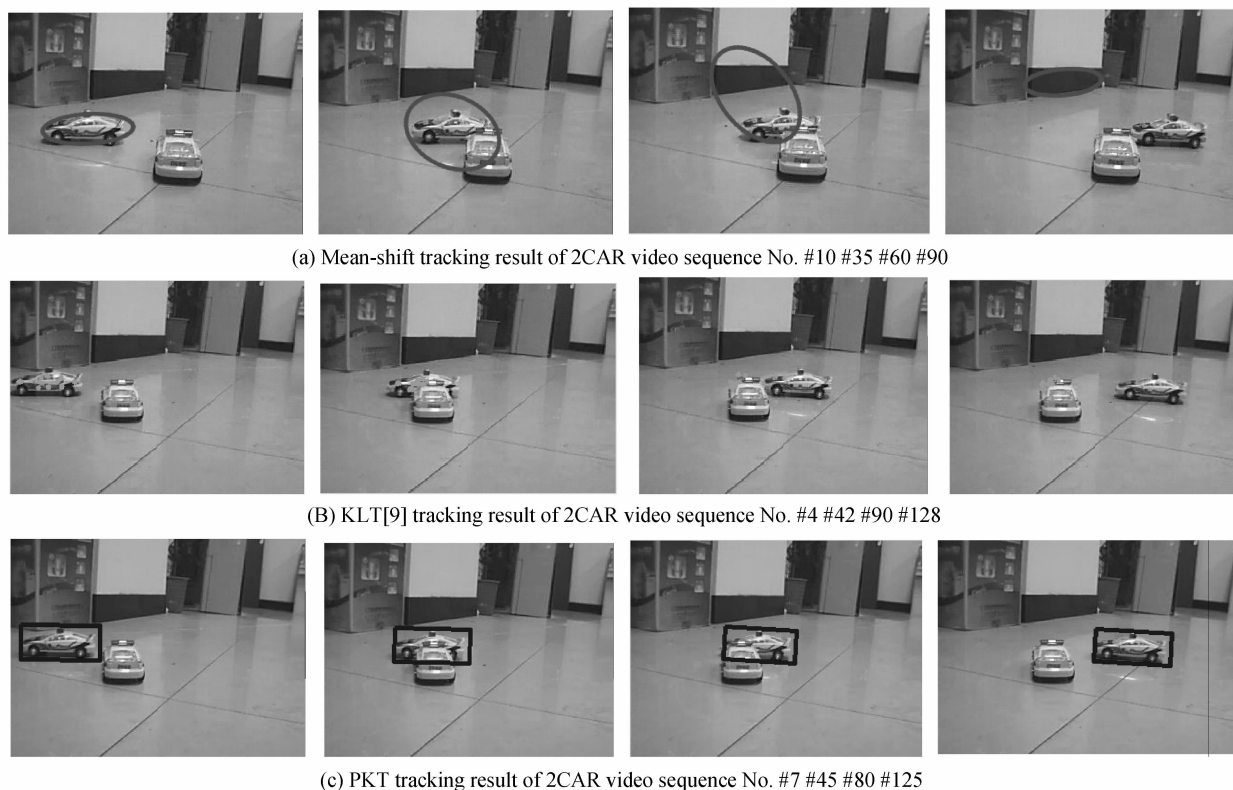


图6 PKT算法对2CAR序列的跟踪结果及Mean-shift和的KLT算法结果对比
Fig. 6 Object tracking results of 2CAR video sequence using mean-shift, KLT and PKT

5 结论

研究表明,人的视觉信息可以通过运动感知目标,然后进行目标识别.动物学家发现水母利用一个相当复杂的眼睛产生一个模糊的影像,对避开静止的大目标、同时又不必关注小的漂浮目标和浮游生物非常有利. PKT 算法模拟了生物的初级视觉功能.下一步将系统研究表征视觉体系结构和特性的新概念和新方法,并与视觉神经系统建模研究相结合,寻求自然图像的“稀疏”表示方法;研究视觉多通道处理机制、图景区分机理及生物仿生,通过人工视网膜建模,构造更高效的分自然图像理解与跟踪算法.

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] LIN Yu-chi, CUI Yan-ping, HUANG Yin-guo. Study on edge detection and target recognition in complex background [J]. *Optics and Precision Engineering*, 2006, **14**(3): 509-514. 林玉池, 崔彦平, 黄银国. 复杂背景下边缘提取与目标识别方法研究 [J]. *光学精密工程*, 2006, **14**(3): 509-514.
- [3] HARIHARAKRISHNAN K, SCHONFELD D. Fast object tracking using adaptive block matching [J]. *IEEE Transactions on Multimedia*, 2005, **7**(5): 853-859.
- [4] SHI Jian-bo, TOMASI C. Good features to track [C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 1994, 593-600.
- [5] LI Yan-jun, YANG Jin-feng, WU Ren-biao, et al. Efficient OBJECT TRACKING BASED ON LOCAL INVARIANT FEATUres [C]. *International Symposium on Communications and Information Technologies*, 2006, 697-700.
- [6] CHEN D, YANG J. Robust object tracking via online dynamic spatial bias appearance models [J]. *PAMI*, 2007, **29**(12): 2157-2169.
- [7] LUCAS B D, KANADE T. An iterative image registration technique with an application to stereo vision [C]. *International Joint Conference on Artificial Intelligence*, 1981, 674-679.
- [8] TANG F, TAO H. Object tracking with dynamic feature graph [C]. In *VS-PETS'05*, 2005, 25-32.
- [9] QU W, SCHONFELD D, MOHAMED M. Real-time distributed multi-object tracking using multiple interactive trackers and a magnetic-inertia potential model [J]. *IEEE Transactions on Multimedia*, 2007, **9**(3): 511-519.
- [10] GUO Bao-long, GUO Lei. A new approach to visual motion computation [J]. *Journal of Xidian University*, 1994, **21**(4): 457-463. 郭宝龙, 郭雷. 视觉运动计算的新方法 [J]. *西安电子科技大学学报*, 1994, **21**(4): 457-463.
- [11] CROWLEY J. A representation for visual information [D]. Pittsburgh: Carnegie Mellon Universtiy, 1981.
- [12] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, **60**(2): 91-110.
- [13] HARRIS C, STEPHENS M. A combined corner and edge detector [C]. In: *Proceedings of The Fourth Alvey Vision Conference*, 1988, 147-151.
- [14] SUN Wei, GUO Bao-long. Multiple objects detecting and tracking with the pseudo particle filter [J]. *Journal of Xidian University*, 2008, **35**(2): 248-253.
- [15] SUN Wei, GUO Bao-long. Robust object tracking via hierarchical particle filter [J]. *Acta Photonica Sinica*, 2010, **39**(5): 945-951. 孙伟, 郭宝龙. 一种新的层次粒子滤波的目标跟踪方法 [J]. *光子学报*, 2010, **39**(5): 945-951.

A Robust Object Detecting and Tracking Method

JIA Wei¹, SUN Wei², LI Da-jian¹

(1 No. 365 Research Institute, Northwest Polytechnical University, Xi'an 710065, China)

(2 School of Mechano-electronic Engineering, Xidian University, Xi'an 710071, China)

Abstract: According to the problem of error accumulation and matched feature points loss in the optical flow feature tracking method, a predictive frame and key frame algorithm framework is proposed based on the new Harris-SIFT feature representation method. The proposed target tracking algorithm was realized by combining optical flow motion estimation and local feature recognition. Predictive frame uses pyramid decomposition and recursive algorithm to compute the motion vectors from optical flow field characteristics. The proposed algorithm gets motion vector of the target and eliminates false matching point from motion vector histogram; when the number of matched point is less than 5, the key frames uses the Harris-SIFT feature point for local feature matching, and affine model was used for accurate target positioning and attitude correction. The experiment results show that the proposed algorithm still can continue to achieve reliable tracking in complex background, target occlusion or temporarily lost case.

Key words: Image processing; Target tracking; Robustness; Optical flow; Feature detection