

文章编号 1004-924X(2016)02-0448-12

## 采用核相关滤波器的自适应尺度目标跟踪

张 雷<sup>1,2\*</sup>, 王延杰<sup>1</sup>, 孙宏海<sup>1</sup>, 姚志军<sup>1</sup>, 吴 培<sup>1,2</sup>

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033;

2. 中国科学院大学, 北京 100049)

**摘要:** 由于现存的大多数基于检测的跟踪器都没有解决尺度变化问题, 本文在传统的基于检测的目标跟踪框架下设计了一种尺度估计策略, 并给出了基于核相关滤波器的自适应尺度目标跟踪算法。该算法利用核函数对正则化最小二乘分类器求解获得核相关滤波器, 通过对核相关滤波器的在线学习完成目标位置和尺度的检测, 并在线更新核相关滤波器。为了验证本文算法的有效性, 选取了 10 组场景复杂的视频序列进行测试, 并与其它 5 种优秀跟踪方法进行了对比。结果表明, 本文提出的方法比上述 5 种优秀跟踪方法中的最优者的平均距离精度提高了 6.9%, 且在目标发生尺度变化、光照变化、部分遮挡、姿态变化、旋转、快速运动等复杂场景下有较强的鲁棒性。

**关 键 词:** 核相关滤波器; 目标跟踪; 自适应尺度; 正则化最小二乘分类器;

**中图分类号:** TP391 **文献标识码:** A **doi:** 10.3788/OPE.20162402.0448

## Adaptive scale object tracking with kernelized correlation filters

ZHANG Lei<sup>1,2\*</sup>, WANG Yan-jie<sup>1</sup>, SUN Hong-hai<sup>1</sup>, YAO Zhi-jun<sup>1</sup>, WU Pei<sup>1,2</sup>

(1. *Changchun Institute of Optics, Fine Mechanics and Physics,*  
*Chinese Academy of Sciences, Changchun 130033, China;*

2. *University of Chinese Academy of Sciences, Beijing 100049, China)*

*\* Corresponding author, E-mail: zhanglei8080@126.com*

**Abstract:** As most of tracking-by-detection methods have not dealt with the scale estimation problem in target tracking, this paper proposes a scale estimation strategy based on the tracking-by-detection framework. Meanwhile, it designs an adaptive scale tracking algorithm based on kernelized correlation filters. The algorithm uses a kernel function to solve the regularized least square classifier to obtain the kernelized correlation filters. Then it completes the target position and scale detection by online learning the kernelized correlation filters, and updates the filters in online. To verify the feasibility of the proposed algorithm, ten groups of benchmark video sequences are tested and obtained results are compared with those of five kinds of tracking algorithms. The experimental results show that the proposed approach improves the performance by 6.9% in the average distance precision as compared to the best one of the other five excellent existing tracking algorithms. It is robust to scale changing, illumination variation, partial occlusion, pose variation, rotation, fast motion and other complex

收稿日期: 2015-02-09; 修订日期: 2015-04-07.

基金项目: 国家 863 高技术研究发展计划资助项目 (No. 2013AA7031010B)

scenes.

**Key words:** kernelized correlation filter; object tracking; adaptive scale; regularized least square classifier

## 1 引言

目标跟踪在计算机视觉中具有非常重要的地位,其在许多领域均有重要应用,如成像制导、视频监控、运动分析及人机交互等<sup>[1]</sup>。目标跟踪本质上就是通过处理非平稳的,随时间变化的目标和背景图像流来获得感兴趣目标在图像中的位置等信息<sup>[2]</sup>。经过多年的研究,人们提出许多跟踪方法,但是对于光照变化、遮挡、姿态及尺度变化等情况下的跟踪问题仍然很困难<sup>[3-5]</sup>。有效的目标外观模型在目标跟踪中至关重要<sup>[6]</sup>,根据目标外观模型所使用的表达策略可以将现有的大多数跟踪方法分为两类:生成式算法<sup>[6-8]</sup>及判别式算法<sup>[9-15]</sup>。生成式算法首先在线或离线建立目标外观模型,然后搜索图像中与该模型最相似的区域作为目标;判别式算法是将跟踪问题看作在线分类,一类属于目标,一类属于背景,其任务就是将目标和背景区分开,并将机器学习的各种在线学习方法应用到目标模型的表示与更新中,以适应目标的内在变化和背景的外在变化,这一类方法也被称为基于检测的跟踪算法<sup>[16-17]</sup>。

近年,学者们开始把相关滤波器引入到基于检测的目标跟踪框架中<sup>[9,11,12,15]</sup>,并取得了很好的跟踪效果。文献[9]中将设计的最小平方误差和输出(Minimum Output Sum of Squared Error, MOSSE)自适应相关滤波器用于目标跟踪,用设计的相关滤波器对目标外观进行建模,其对目标外观变化有较强的适应性,且对目标和背景有很强的区分能力。文献[11]中提出了一种密集采样策略,通过学习正则化最小二乘(Regularized Least Squares, RLS)分类器来判别目标,并利用循环矩阵将核函数与相关滤波器联系起来,从而提出了核相关滤波器,并设计了使用核函数的基于检测的跟踪循环结构(Circulant Structure of Tracking-by-detection with Kernels, CSK)跟踪器,文献[18]对现存的10种优秀跟踪器进行了评

测,结果显示CSK的处理速度最快并具有良好的跟踪性能。文献[12]中采用径向基函数(Radial Basis Function, RBF)核将多维颜色特征加入到CSK跟踪器中,并提出了一种自适应可变的低维颜色特征。文献[15]对CSK进行了扩展,即将单通道的核相关滤波器扩展到多通道,以便使用更复杂的多维特征,并使用梯度方向直方图(Histogram of Oriented Gradients, HOG)特征代替了简单的自然像素值,提出了核相关滤波器(Kernelized Correlation Filter, KCF)及双相关滤波器(Dual Correlation Filter, DCF)跟踪器。

但是上述列举的跟踪器及现存的大多数基于检测的跟踪器都没有解决尺度变化问题,在目标尺度发生较大变化的情况下跟踪性能受到很大的限制。为此本文在CSK跟踪器的基础上,提出了一种基于核相关滤波器的尺度估计方法,以实现尺度自适应的目标跟踪。为了验证本文方法的有效性,从文献[18]中选取了10组场景复杂的视频序列进行测试,并与CSK及其它4种优秀跟踪方法进行了对比。

## 2 CSK跟踪器

本文算法是基于CSK跟踪器提出的,CSK跟踪器是通过一个单独的图像块来学习目标的RLS分类器,并利用循环矩阵及核函数对其求解,得到核相关滤波器,然后通过计算核相关滤波器响应的最大值确定目标位置,下面简单介绍一下CSK的相关内容。

### 2.1 循环矩阵

假设 $C(r)$ 是一个 $n \times n$ 的循环矩阵,则它可以通过一个 $1 \times n$ 向量 $r$ 的循环移位获得,则有:

$$C(r) = \begin{bmatrix} r_0 & r_{n-1} & \cdots & r_1 \\ r_1 & r_0 & \cdots & r_2 \\ \vdots & \vdots & \ddots & \vdots \\ r_{n-1} & r_{n-2} & \cdots & r_0 \end{bmatrix}, \quad (1)$$

$C(r)$ 的第一列是向量 $r$ 的转置,第二列是向

量  $r$  向右循环移动一位元素转置后的结果,依此类推可以得到  $C(r)$  的所有列。对于一个  $n \times 1$  的向量  $v$ ,则  $C(r)$  与  $v$  的乘积就是向量  $r$  和  $v$  的卷积<sup>[19]</sup>,那么在频域下就有:

$$C(r)v = F^{-1}(F(r)F(v)), \quad (2)$$

其中  $F^{-1}$  和  $F$  分别表示傅里叶反变换和傅里叶正变换。

## 2.2 正则化最小二乘分类器的求解

在许多实际问题中,RLS 分类器在性能上完全可以与支持向量机相媲美,而且实现简单,只要求解一个线性方程<sup>[20]</sup>。分类器  $f(x)$  使用一个包含目标大小为  $M \times N$  的灰度图像块进行训练,通过样本  $x_i$  找到一个函数  $f(x) = w^T x$ ,使正则化风险泛函最小:

$$\min_w \sum_i (f(x_i) - y_i)^2 + \lambda \|w\|^2, \quad (3)$$

其中  $y_i$  为样本  $x_i$  的期望输出,  $\lambda$  为一个固定的正则化参数。

采用核函数将输入  $x$  映射到特征空间  $\varphi(x)$  上,定义  $\kappa(x, x') = \varphi^T(x) \varphi(x')$ ,  $\kappa$  为核函数,则可以把式(3)的解表示为输入的线性组合<sup>[21]</sup>,系数为  $b$ :

$$w = \sum_i b_i \varphi(x_i), \quad (4)$$

根据文献[20]可以得到  $b$  的表达式:

$$b = (K + \lambda I)^{-1} y, \quad (5)$$

其中  $K$  为核矩阵,元素  $K_{ij} = \kappa(x_i, x_j)$ ,  $I$  为单位矩阵,向量  $y$  的元素为  $y_i$ ,向量  $b$  的元素为  $b_i$ 。

文献[11]中采用密集采样方法,也就是在选定的图像块中用所有的子窗口(大小固定)图像来训练分类器,对于表示成  $n \times 1$  向量的某个图像块  $x$ ,那么采集的单通道的自然像素值特征样本就可以表示为:

$$x_i = P^i x, \quad (6)$$

其中  $P$  为式(7)所示的  $n \times n$  矩阵,  $i=0, 1, \dots, n-1$ 。

$$P = \begin{bmatrix} 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad (7)$$

文献[11]证明了具有酉不变性的核函数的核

矩阵具有循环矩阵结构,则对于  $\kappa$  是具有酉不变性的核函数(如径向基核函数、线性及多项式核函数等),则有  $K$  为循环矩阵,表示成  $K = C(k)$ ,  $k$  的元素  $k_i = \kappa(x, x_i)$ ,结合 2.1 节中循环矩阵的性质经推导可得:

$$B = F(b) = \frac{F(y)}{F(k) + \lambda}. \quad (8)$$

## 2.3 核相关滤波器的响应

在新的一帧中通过获取感兴趣图像块  $z$  来完成位置检测,则分类器的响应  $\hat{y} = w^T z = \sum_i b_i \kappa(z_i, x)$ ,利用前面的知识有:

$$\hat{y} = F^{-1}(BH), \quad (9)$$

其中:  $H = F(h)$ ,  $h$  的元素  $h_i = \kappa(x, z_i)$ ,  $z_i$  为新一帧中获取的训练样本,  $x$  为上一帧学习获得的目标模型,使  $\hat{y}$  为最大的位置就是新一帧中目标的位置,并更新  $B$  和  $x$  如下:

$$\hat{B}_t = (1 - \alpha) \hat{B}_{t-1} + \alpha B_t, \quad (10)$$

$$\hat{x}_t = (1 - \alpha) \hat{x}_{t-1} + \alpha x_t, \quad (11)$$

其中  $\alpha$  为学习因子,  $\hat{B}_t$ 、 $\hat{B}_{t-1}$  分别表示当前帧和上一帧更新获得的  $B$ ,  $\hat{x}_t$ 、 $\hat{x}_{t-1}$  分别表示当前帧和上一帧更新获得的  $x$ 。

文献[11]中并没有明确说明什么是核相关滤波器,而文献[15]中给出了一些解释和分析,从前面可以知道有  $\hat{y} = \sum_i b_i h_i$ ,则  $\hat{y}$  就是  $h$  乘以学习系数  $b$  的线性组合,就是一个滤波操作,  $h$  就是核相关滤波器,则分类器的响应  $\hat{y}$  就相当于  $h$  的输出响应。

文献[11]中只针对单通道特征进行了讨论,为了使其具有通用性,文献[15]对于具有  $C$  个通道的特征向量  $x = [x_1, \dots, x_C]$  进行了分析,下面仅给出对于高斯核  $\kappa(x, x') = \exp(-\frac{1}{\sigma^2} \|x - x'\|^2)$  的计算形式,其中“ $*$ ”表示共轭运算。

$$h^{\text{gauss}} = \exp\left(-\frac{1}{\sigma^2} (\|x\|^2 + \|x'\|^2 - 2F^{-1}(\sum_{c=1}^C F(x_c) F^*(x'_c)))\right). \quad (12)$$

### 3 核相关滤波器自适应尺度目标跟踪

文献[12]及文献[15]对 CSK 跟踪器进行了改进,但是都没有解决尺度问题。最近文献[22]提出一种基于 MOSSE 滤波器的尺度估计方法,本文受其启发在采用核相关滤波器的 CSK 跟踪器<sup>[11]</sup>的基础上,增加了采用核相关滤波器对目标尺度进行估计的环节,以实现自适应尺度目标跟踪。

通过学习一个独立的一维核相关滤波器(尺度核相关滤波器)去检测目标尺度变化,从而能够依据目标尺度变化合理地限定搜索区域。在目标附近采用一系列尺寸变化的图像块,将其通过双线性插值变换到同初始目标图像块一样大小,然后对尺度变化后的图像块提取特征,并用提取的特征来训练 RLS 分类器和学习尺度核相关滤波器。假设当前帧的目标大小为  $M \times N$ ,尺度滤波器的大小为  $S \times 1$ ,提取目标周围大小为  $m \times n$  的图像块获得的样本为  $x_{s_i}$ ,  $m = r^a M$ ,  $n = r^a N$ ,  $a \in \left\lfloor -\frac{S-1}{2} \right\rfloor, \dots, \left\lfloor \frac{S-1}{2} \right\rfloor$ ,  $r$  为尺度因子。依然按照第二节中目标位置的检测方法训练分类器和学习核相关滤波器,则学习系数  $B_s$  (频域)可由式(13)获得,检测尺度滤波器的响应  $\hat{y}_s$  可由式(14)获得,找到使  $\hat{y}_s$  最大的尺度作为新一帧中目标的尺度,并按照式(15)、(16)更新相应参数。

$$B_s = \frac{F(y_s)}{F(k) + \lambda}, \quad (13)$$

其中  $k$  的元素  $k_i = \kappa(x_s, x_{s_i})$ ,  $y_s$  为分类器的期望输出。

$$\hat{y}_s = F^{-1}(B_s H_s), \quad (14)$$

其中  $H_s = F(h_s)$ ,  $h_s$  的元素  $h_{s_i} = \kappa(x_s, z_{s_i})$ , 输入  $x_s$  为上一帧学习获得的尺度模型,  $z_{s_i}$  为新一帧中获取的样本。

$$\hat{B}_{s_i} = (1 - \beta) \hat{B}_{s_{i-1}} + \beta B_{s_i}, \quad (15)$$

$$\hat{x}_{s_i} = (1 - \beta) \hat{x}_{s_{i-1}} + \beta x_{s_i}, \quad (16)$$

其中:  $\beta$  为尺度学习因子,  $\hat{B}_{s_i}$ 、 $\hat{B}_{s_{i-1}}$  分别表示当前帧和上一帧更新后的  $B_s$ ,  $\hat{x}_{s_i}$ 、 $\hat{x}_{s_{i-1}}$  分别表示当前帧和上一帧更新后的  $x_s$ 。

在目标跟踪中,由于相邻两帧的目标尺度变

化很小,为此先利用位置核相关滤波器  $H$  检测出目标位置,然后在检测出的目标位置周围采集样本,利用尺度核相关滤波器  $H_s$  检测出目标尺度,这样就完成了目标位置和尺度的检测。整个跟踪算法框架如表 1 所示。

表 1 本文算法的框架

Tab. 1 Framework of proposed algorithm

初始化: 根据第一帧视频选取要跟踪的目标, 确定相关参数;
输入: 第 $i(i \geq 2)$ 帧视频 $V_i$ ; 上一帧的目标位置 $P_{i-1}$ 及尺度 $s_{i-1}$ ; 位置检测中的 $\hat{B}_{i-1}$ , $\hat{x}_{i-1}$ 及尺度检测中的 $\hat{B}_{i-1}$ , $\hat{x}_{i-1}$ 。
输出: 检测出当前帧目标的位置 $P_i$ 及尺度 $s_i$ ;
位置检测:
① 根据 $P_{i-1}$ 及 $s_{i-1}$ 从 $V_i$ 中提取样本 $x$ ;
② 使用 $x$ , $\hat{B}_{i-1}$ 及 $\hat{x}_{i-1}$ 利用式(9)计算响应 $\hat{y}$ ;
③ 找到使 $\hat{y}$ 取得最大值时的目标位置记为 $P_i$ 。
尺度检测:
④ 根据 $P_i$ 及 $s_{i-1}$ 从 $V_i$ 中提取样本 $x_s$ ;
⑤ 使用 $x_s$ , $\hat{B}_{i-1}$ 及 $\hat{x}_{i-1}$ 利用式(14)计算响应 $\hat{y}_s$ ;
⑥ 找到使 $\hat{y}_s$ 取得最大值时的尺度记为 $s_i$ 。
更新:
⑦ 根据 $P_i$ 及 $s_i$ 从 $V_i$ 中提取样本 $x^{\text{new}}$ , $x_s^{\text{new}}$ ;
⑧ 利用式(10)、(11)更新 $B_i$ , $x_i$ ;
⑨ 利用式(15)、(16)更新 $B_{s_i}$ , $x_{s_i}$ 。

其中位置检测和尺度检测均用高斯核函数, 尺度检测采用多通道的 HOG 特征, 分类器的期望输出  $y$  及  $y_s$  所使用的高斯函数分别如式(17)、式(18)所示:

$$y = \exp \left( - \left( \frac{p - p^*}{\sigma} \right)^2 \right), \quad (17)$$

$$y_s = \exp \left( - \left( \frac{s - s^*}{\sigma_s} \right)^2 \right), \quad (18)$$

其中:  $p$  为目标位置,  $p^*$  为目标中心位置,  $s$  为目标可取的所有尺度( $s$  的元素  $s_i$  为整数,  $1 \leq s_i \leq S$ ),  $s^*$  取  $\left[ \frac{S}{2} \right]$ ,  $\sigma, \sigma_s$  为标准差。

### 4 实验结果与分析

为了验证本文方法的有效性,从文献[18]中

选取了 10 组有挑战性的视频进行测试,这些视频涵盖了目标尺度变化、光照变化、姿态变化、遮挡、快速运动、旋转等复杂情况(如表 2 所示),并对实验结果进行了定量和定性的分析。为了保证公平性、准确性,实验中用于对比的跟踪方法的代码均从其作者提供的网站上获得。实验分为两部分:第一部分是与 CSK 跟踪器进行了简单比较;第二部分是将本文方法与近年来出现的几种基于检测的优秀跟踪方法进行比较,这些方法包括 MOSSE<sup>[9]</sup>、加权多示例学习跟踪(Weighted Multiple Instance Learning Tracker, WMILT)<sup>[23]</sup>、压缩跟踪(Compressive Tracker, CT)<sup>[24]</sup>、KCF<sup>[15]</sup>(使用 HOG 特征的 KCF)以及 CSK<sup>[11]</sup>。

表 2 实验中的测试视频  
Tab. 2 Test sequences in our experiments

视频	帧数	存在的主要挑战
Car4	659	尺度、光照变化,背景干扰
CarScale	252	尺度变化、遮挡
Dog1	1 350	尺度、姿态变化
Girl	500	尺度、姿态变化,360 度旋转、遮挡
Trellis	569	光照、姿态、尺度变化
Singer1	351	光照、尺度、姿态变化
David	462	姿态、尺度、光照变化
Faceoccl	892	遮挡
Tiger1	354	快速运动、姿态变化、遮挡
CarDark	393	光照变化、背景干扰

#### 4.1 实验环境及参数

本文算法采用 Matlab 和 C 语言混合编写,实验均在 Intel Core i 3-2130 CPU,主频为 3.4GHz,2GB 内存配置的电脑上进行。实验中对于所有测试视频,算法中的参数保持不变。目标大小为  $m \times n$ ,尺度滤波器的尺度为  $S \times 1$ , $y$  的标准差取  $\sigma = \sqrt{mn}$ , $y_s$  的标准差取  $\sigma_s = \sqrt{S}$ ,高斯核的标准差  $\sigma$  取 0.2,学习因子  $\alpha$  取 0.075,正则化参数  $\lambda$  取 0.01,尺度滤波器的  $S$  取 32,尺度因子  $r$  取 1.1,尺度学习因子  $\beta$  取 0.025。

#### 4.2 评价标准

为了定量分析跟踪方法的性能,实验中选用中心位置误差(Center Location Error, CLE)、成功率<sup>[25]</sup>(Success Rate, SR)及距离精度<sup>[26]</sup>(Dis-

tance Precision, DP)作为评测标准,这 3 种标准的计算如式(19)~式(22)所示:

$$CLE = \sqrt{(x_i - x_{i\_gt})^2 + (y_i - y_{i\_gt})^2}, \quad (19)$$

式中: $CLE$  为中心位置误差,  $(x_i, y_i)$  为第  $i$  帧视频通过跟踪算法获得的目标中心位置,  $(x_{i\_gt}, y_{i\_gt})$  为第  $i$  帧视频通过手动标记所获得的目标中心位置。

$$SR = \frac{sn}{n}, \quad (20)$$

$$score = \frac{area(R_t \cap R_g)}{area(R_t \cup R_g)}, \quad (21)$$

式中: $sn$  为跟踪成功次数,  $n$  为视频帧数,通过式(21)来统计  $sn$ 。 $score$  称为重叠率,如果  $score > 0.5$ ,则  $sn$  增加一次,  $R_t$  为实验获得的目标矩形窗口,  $R_g$  为手动标记获得的目标矩形窗口,  $area$  表示区域面积。

$$DP = \frac{m}{n}, \quad (22)$$

式中: $m$  为  $CLE$  小于某个固定阈值(实验中均取 20 pixel)的视频帧数,  $n$  为测试视频总帧数。

6 种跟踪方法对 10 组视频进行测试获得的平均 CLE (Median Center Location Error, MCLE)、SR 及 DP 分别如表 3、表 4、表 5 所示,其中最佳值均被加粗。

表 3 平均中心位置误差

Tab. 3 Average center location error (pixel)

视频	MOSSE	WMILT	CT	KCF	CSK	Ours
Car4	105.176	85.703	77.135	9.472	19.072	<b>4.394</b>
CarScale	69.999	70.415	<b>15.972</b>	16.142	90.527	19.499
Dog1	<b>4.209</b>	6.680	9.054	4.418	4.863	5.304
Girl	52.393	51.018	40.495	32.483	35.968	<b>17.223</b>
Trellis	77.523	43.042	44.784	8.210	19.070	<b>7.819</b>
Singer1	12.153	16.784	19.148	12.824	13.930	<b>4.528</b>
David	129.093	24.420	<b>8.973</b>	9.529	17.213	9.468
Faceoccl	6.064	31.750	29.620	43.683	<b>5.465</b>	6.685
Tiger1	32.553	<b>9.500</b>	22.189	18.031	26.060	11.693
CarDark	3.242	60.841	46.955	5.764	<b>2.585</b>	2.616
平均值	49.241	40.015	31.433	16.056	23.475	<b>8.923</b>

表4 成功率

Tab. 4 Success rate (%)

视 频	MOSSEWMILT	CT	KCF	CSK	Ours	
Car4	27.5	24.6	27.6	36.7	27.6	<b>100</b>
CarScale	44.8	44.8	44.8	44.4	44.8	<b>81.0</b>
Dog1	65.3	62.7	59.4	65.3	65.1	<b>99.9</b>
Girl	35.6	27.7	30.7	<b>84.2</b>	51.5	76.2
Trellis	30.6	33.6	26.0	84.0	55.0	<b>89.8</b>
Singer1	29.6	27.6	27.6	29.6	29.6	<b>100</b>
David	14.0	51.6	<b>100</b>	<b>100</b>	57.0	25.8
Faceoccl	<b>100</b>	58.4	74.7	65.7	99.4	99.4
Tiger1	39.4	<b>70.4</b>	56.3	69.0	50.7	57.7
CarDark	96.4	0.30	12.2	72.3	<b>100</b>	<b>100</b>
平均值	48.3	40.2	45.9	65.1	58.1	<b>83.0</b>

表5 距离精度

Tab. 5 Distance precision (%)

视频	MOSSEWMILT	CT	KCF	CSK	Ours	
Car4	28.1	24.1	35.4	95.3	35.5	<b>100</b>
CarScale	65.1	63.1	65.1	<b>80.6</b>	65.5	76.2
Dog1	<b>100</b>	94.3	94.9	<b>100</b>	99.9	99.6
Girl	34.7	21.8	24.8	60.4	39.6	<b>65.3</b>
Trellis	34.1	45.0	25.3	<b>100</b>	75.6	<b>100</b>
Singer1	84.9	63.5	32.2	81.5	67.5	<b>100</b>
David	14.0	40.9	<b>100</b>	<b>100</b>	50.5	97.8
Faceoccl	<b>99.4</b>	19.1	32.0	64.6	<b>99.4</b>	98.9
Tiger1	52.1	<b>93.0</b>	67.6	73.2	63.4	87.3
CarDark	<b>100</b>	10.4	21.4	<b>100</b>	<b>100</b>	<b>100</b>
平均值	57.8	47.5	49.9	85.6	69.7	<b>92.5</b>

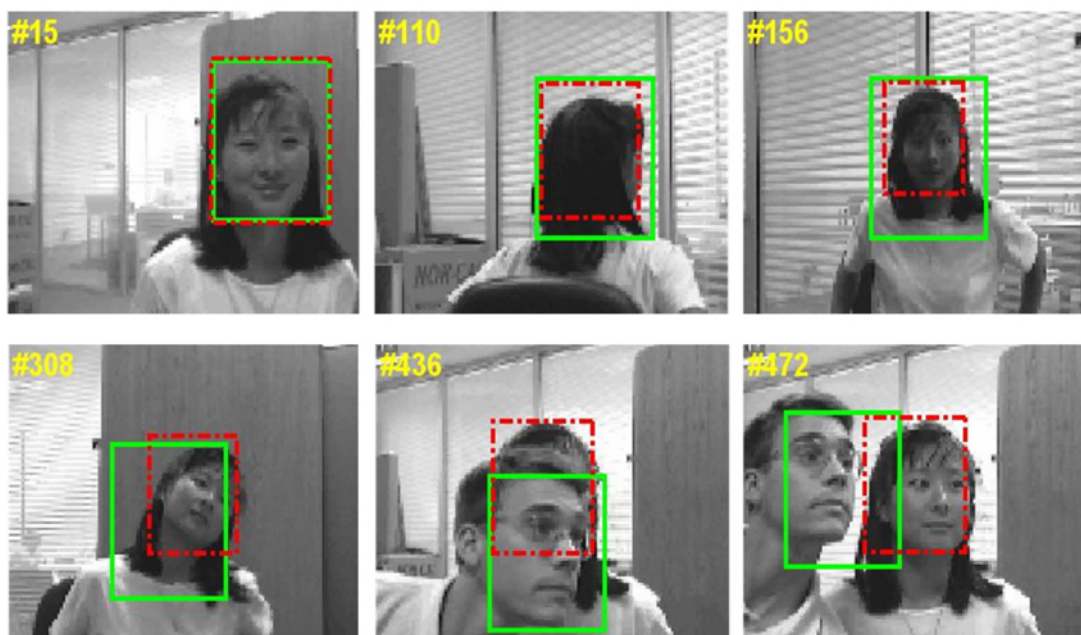


图1 实验一部分跟踪结果(第15,110,156,308,436和472帧)

Fig. 1 Partial tracking results of experiment 1(Frames 15,110,156,308,436 and 472)

#### 4.3 实验一:与CSK跟踪器的比较实验

从表3、表4、表5中可以看到,平均中心位置误差的均值由原来的23.475 pixel减少到8.923 pixel,成功率均值由原来的58.1%提高到83%,距离精度的均值由原来的69.7%提高到92.5%,这3个评价标准的定量分析可以证明本文方法的跟踪性能优于CSK跟踪方法。

为了更直观地说明对比结果,下面以Girl视频的测试结果为例进行实验,图1、图2分别给出了两种方法的部分跟踪结果及3个评价标准的曲

线,其中图1中点画线跟踪框为本文跟踪结果,实线框为CSK跟踪结果;图2(a)中标注的数值为MCLE,(b)图中标注的数值为成功率,(c)中标注的数值是阈值为20 pixel时的距离精度。

Girl视频中的目标分别发生了尺度变化、旋转、姿态变化及遮挡等情况,从图1可以看到:第110帧目标发生旋转时,CSK跟踪开始出现偏差;第156帧目标变小时,本文方法的跟踪框也相应变小并仍能准确跟踪;随着目标姿态的剧烈变化,CSK的跟踪误差逐渐积累,从第308帧开始出现



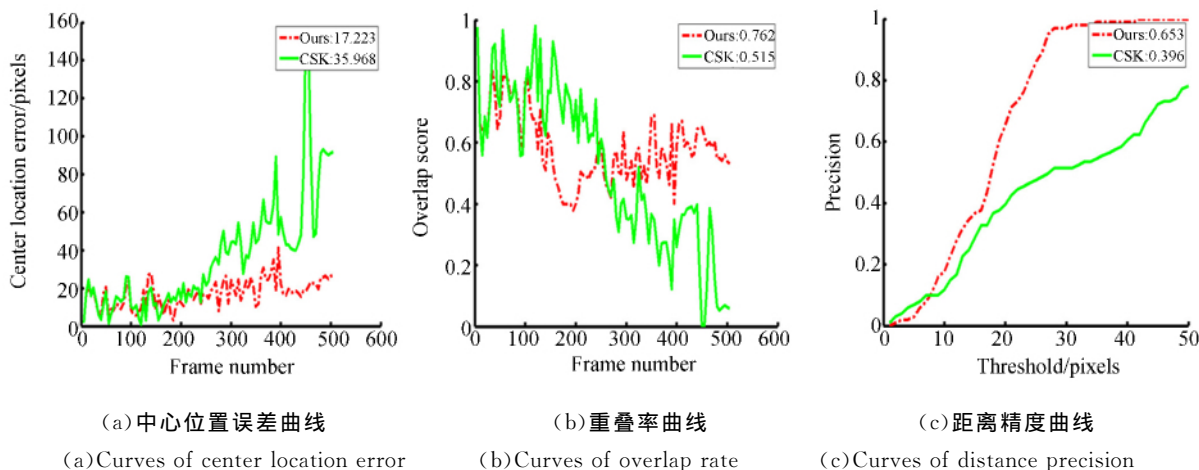


图 2 实验一曲线

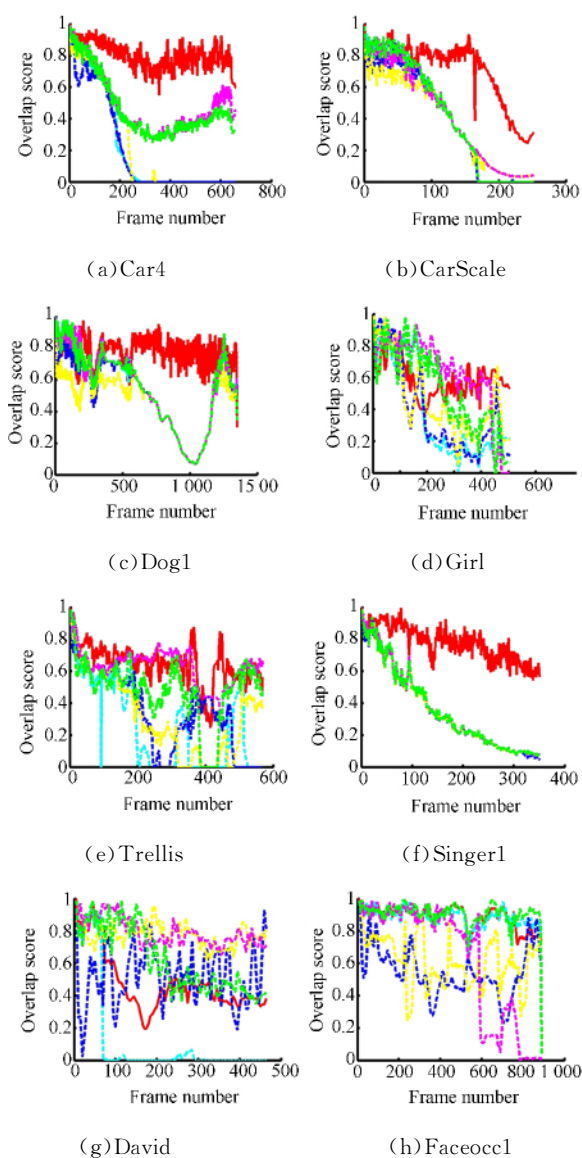
Fig. 2 The curves of experiment 1

很大的偏差;而对于第 436 帧,目标发生严重遮挡时,CSK 方法错把遮挡物当作目标,在第 472 帧遮挡消失后已丢失目标,而本文从发生遮挡到遮挡消失后一直能较准确地跟踪目标。从图 2 可以看到本文算法的 3 种评价标准曲线都优于 CSK。从以上分析可知本文方法较好。

#### 4.4 实验二:与几种优秀跟踪方法的比较实验

通过分析表 3、表 4 及表 5 中的数据可知:本文方法的 MCLE、SR 及 DP 3 个评测标准在大多数测试视频中均达到最佳或为次最佳,对于 10 组测试视频,本文方法对于 3 个标准的均值都为最佳值,而且,本文的 MCLE 较次最佳值减少了 44.4%,平均 SR 均值提高了 17.9%,平均 DP 提高了 6.9%。通过 3 个标准数据的比较可见,相对于其他几种优秀跟踪方法,本文方法在跟踪性能上有了很大的提高。实验中绘制了 6 种算法对 10 组视频的 3 种标准曲线,但为了节省篇幅,文中仅给出 10 组重叠率曲线,如图 3 所示(彩图见期刊电子版)。图中红色实线代表本文方法,浅蓝色虚线、蓝色虚线、黄色虚线、紫色虚线及绿色虚线分别代表 MOSSE、WMILT、CT、KCF 及 CSK 曲线。

图 4(彩图见期刊电子版)和图 5(彩图见期刊电子版)展示了 6 种跟踪方法在 10 组视频上的部分实验结果,其中红色点画线框代表本文算法的跟踪结果,浅蓝色点线框、蓝色点线框、黄色虚线框、白色虚线框及绿色实线框分别代表 MOSSE、WMILT、CT、KCF 及 CSK 的跟踪结果。



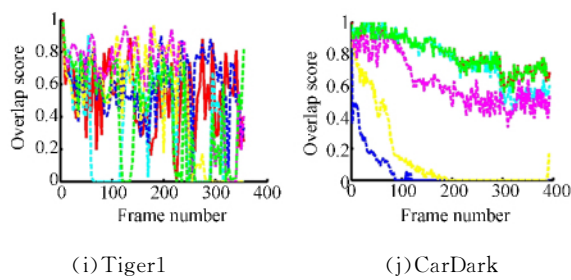


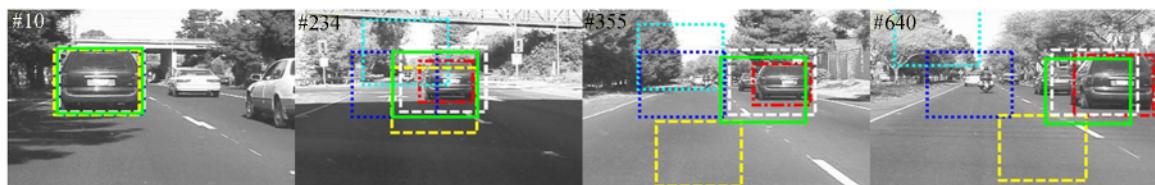
图3 实验二重叠率曲线

Fig. 3 Overlap rate curves of experiment 2

#### 4.4.1 尺度、光照、姿态变化及背景干扰情况的实验结果

图4(a)中的Car4视频中存在目标尺度变化、光照变化及背景干扰等因素,整个过程中只有KCF和本文方法的跟踪效果较好,但由于目标尺度的不断变化,KCF跟踪窗口引入的背景干扰信息逐渐积累,最终也出现了较大的跟踪偏差(如第640帧),而本文方法由于能够根据目标尺度变

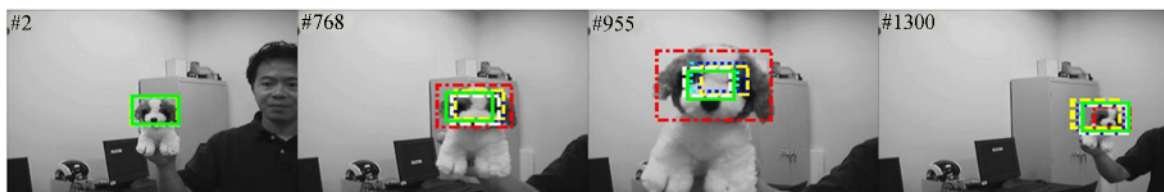
化自动调整跟踪窗口大小,从而减少了背景干扰信息,所以能够一直对目标位置和尺度进行估计; Singer1视频中的目标经历了光照、尺度剧烈变化等情况(如图4(b)中第65,110,351帧),整个过程中只有本文方法能够完成整个视频的跟踪,其他方法都不能适应光照和尺度的剧烈变化; Dog1视频中目标发生了一定的尺度、姿态变化,对于该视频, MOSSE、KCF及本文方法的跟踪表现均较好,但本文方法的SR最高; David和Trellis视频中存在光照、尺度及姿态变化等问题(如图4(d)、(e)所示),对于David视频只有CT、KCF及本文方法能够正确跟踪大部分视频序列,对于Trellis视频,由于目标发生了快速的姿态变化,此时,只有KCF及本文方法表现较好,而本文方法的CLE最小且SR最高; CarDark视频存在光照变化和背景干扰问题(如图4(f)所示),只有MOSSE、CSK及本文方法的跟踪效果较好,其中CSK和本文方法的SR最高。



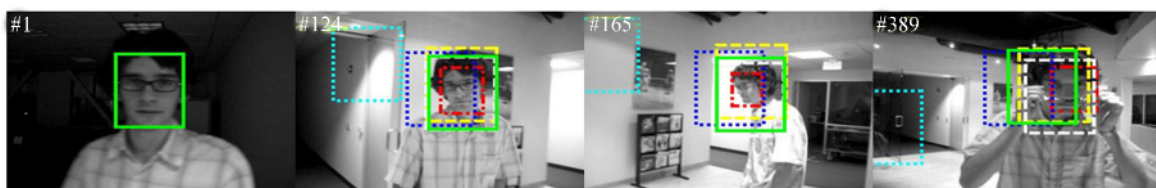
(a) Car4



(b) Singer1



(c) Dog1

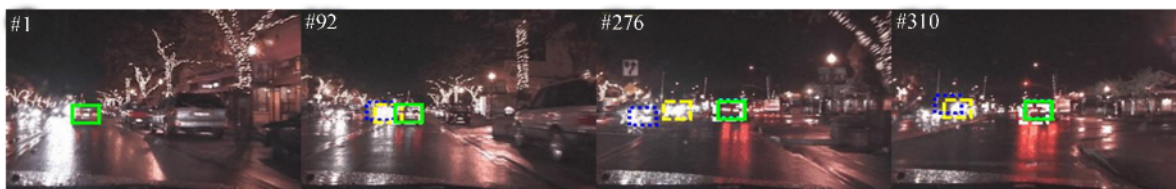


(d) David





(e) Trellis



(f) CarDark

图 4 实验二部分跟踪结果

Fig. 4 Partial tracking results of experiment 2

#### 4.4.2 遮挡、尺度变化、旋转及快速运动的实验结果

CarScale、Girl、Faceoccl 及 Tiger1 视频中目标均出现部分遮挡的情况(如图 5(a)、(b)、(c)及(d))。CarScale 中的目标还存在尺度变化,目标由远及近逐渐变大,由图 5(a)可知,从目标被树遮挡到遮挡结束后(第 159~182 帧),只有 KCF 和本文方法实现了准确跟踪,而且本文方法的 SR 较高,两者的 CLE 和 DP 相差不大;Girl 视频中目标还存在尺度和姿态变化,目标发生 360° 旋转(如第 174, 307, 437 帧),使跟踪更加困难。由图 5(b)可知,只有本文方法能够较好地完成该视频

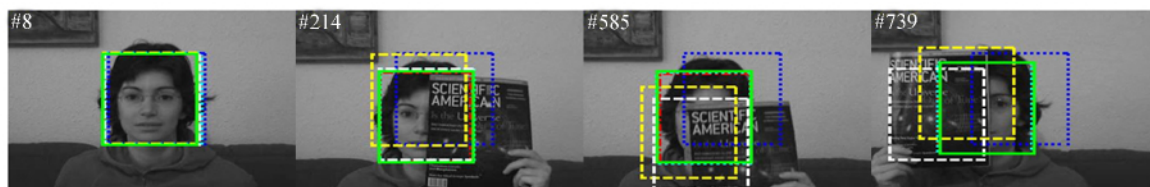
的跟踪任务,而其他方法偏差较大,甚至丢失目标;Faceoccl 视频主要存在部分遮挡问题。由图 5(c)可以看到, MOSSE、CSK 及本文方法都具有较强的抗遮挡能力,其他方法每次发生遮挡时都比较敏感(如第 214, 585, 739 帧);Tiger1 视频中目标存在快速运动,同时存在姿态变化,有时还发生遮挡,由图 5(d)可知,整个过程中只有 WM-ILT 和本文方法能够适应这些情况,前者的 CLE、SR 及 DP 略好些,但两者相差不大,而其他方法不是在遮挡时丢失目标(如第 60 帧)就是在快速运动和姿态变化时丢失目标(如第 247, 327 帧)。



(a) CarScale



(b) Girl



(c)Faceoccl



(d)Tiger1

图5 实验二部分跟踪结果

Fig.5 Partial tracking results of experiment 2

## 5 结 论

本文在传统的基于检测的目标跟踪框架下,在核相关滤波器目标跟踪方法的基础上,提出了一种核相关滤波器的自适应尺度的目标跟踪方法。该方法利用核相关滤波器检测目标的位置和尺度,在跟踪过程中能够随目标尺度变化而自适应调整,因此可以获得更多的有效目标信息并减

少背景干扰信息的干扰,从而在复杂场景下具有更强的适应性。对10组测试视频进行实验,结果显示本文方法的平均中心位置误差均值为8.923 pixel,成功率的均值为83%,距离精度的均值可达92.5%。实验数据表明,在目标发生尺度变化、光照变化、姿态变化、旋转、遮挡以及快速运动等复杂场景下,与近年出现的几种优秀的基于检测的跟踪方法相比,本文方法具有更强的鲁棒性。

## 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] 李静宇,王延杰. 基于子空间的目标跟踪算法研究[J]. *液晶与显示*, 2014, 29(4):617-622.  
LI J Y, WANG Y J. Subspace based target tracking algorithm [J]. *Chinese Journal of Liquid Crystals and Displays*, 2014, 29(4):617-622. (in Chinese)
- [3] 陈东成,朱明,高文,等. 在线加权多示例学习实时目标跟踪[J]. *光学 精密工程*, 2014, 22(6): 1661-1667.  
CHEN D CH, ZHU M, GAO W, *et al.*. Real-time object tracking via online weighted multiple instance learning [J]. *Opt. Precision Eng.*, 2014, 22(6): 1661-1667. (in Chinese)
- [4] 宋策,张葆,尹传历. 适于机载环境对地目标跟踪的粒子滤波设计[J]. *光学 精密工程*, 2014, 22(4): 1037-1047.  
SONG C, ZHANG B, YIN CH L. Particle filter design for tracking ground targets in airborne environment[J]. *Opt. Precision Eng.*, 2014, 22(4): 1037-1047. (in Chinese)
- [5] 张雷,王延杰,何舒文. 基于相位一致性的实时压缩跟踪方法[J]. *光子学报*, 2014, 43(8):0810003.  
ZHANG L, WANG Y J, HE SH W. Real-time compressive tracking method based on phase congruency[J]. *Acta Photonica Sinica*, 2014, 43(8): 0810003. (in Chinese)
- [6] JEPSON A, FLEET D, EL-MARAGHI T. Robust online appearance models for visual tracking [J].

- IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(10): 1296-1311.
- [7] ADAM A, RIVLIN E and SHIMSHONI I. Robust fragments-based tracking using the integral histogram[J]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2006:798-805.
- [8] ROSS D, LIM J, LIN R, *et al.*. Incremental learning for robust visual tracking [J]. *International Journal of Computer Vision*, 2008, 77(1-3): 125-141.
- [9] BOLME D S, BEVERIDGE J R, DRAPER B A, *et al.*. Visual object tracking using adaptive correlation filters [C]. *23rd IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, 13-18.
- [10] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking-learning-detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(7):1409-1422.
- [11] HENRIQUES J F, CASEIRO R, MARTINS P, *et al.*. Exploiting the circulant structure of tracking-by-detection with kernels[C]. *European Conference on Computer Vision*, 2012:702-715.
- [12] DANELLJAN M, KHAN F S, FELSBERG M. Adaptive color attributes for real-time visual tracking[C]. *27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014:23-28.
- [13] ZHANG K H, ZHANG L, YANG M H. Fast compressive tracking [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(10):2002-2015.
- [14] 郭敬明,何昕,魏仲慧. 基于在线支持向量机的 Mean Shift 彩色图像跟踪[J]. *液晶与显示*, 2014, 29(1):120-128.
- GUO J M, HE X, WEI ZH H. New Mean Shift tracking for color image based on online support vector machine [J]. *Chinese Journal of Liquid Crystals and Displays*, 2014, 29(1):120-128. (in Chinese)
- [15] HENRIQUES J F, CASEIRO R, MARTINS P, *et al.*. High speed tracking with kernelized correlation filters[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 583-596.
- [16] 孙晓燕,常发亮. 梯度特征稀疏表示目标跟踪[J]. *光学 精密工程*, 2013, 21(12):3191-3197.
- SUN X Y, CHANG F L. Object tracking based on sparse representation of gradient feature[J]. *Opt. Precision Eng.*, 2013, 21(12): 3191-3197. (in Chinese)
- [17] 朱秋平,颜佳,张虎,等. 基于压缩感知的多特征实时跟踪[J]. *光学 精密工程*, 2013, 21(2):437-444.
- ZHU Q P, YAN J, ZHANG H, *et al.*. Real-time tracking using multiple features based on compressive sensing [J]. *Opt. Precision Eng.*, 2013, 21(2): 437-444. (in Chinese)
- [18] WU Y, LIM J, YANG M-H. Online object tracking: a benchmark[J]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2013: 2411-2418.
- [19] GRAY R M. *Toeplitz and Circulant Matrices: A Review* [M]. Boston: Now Publishers Inc, 2006.
- [20] RIFKIN R, YEO G, POGGIO T. Regularized least-squares classification[J]. *Nato Science Series Sub Series III: Computer and Systems Sciences*, 2003, 190: 131-154.
- [21] SCHÖLKOPF B, SMOLA A J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* [M]. London: The MIT Press, 2002.
- [22] DANELL J M, HAGERG, KHAN F S, *et al.*. Accurate scale estimation for robust visual tracking [J]. *British Machine Vision Conference*, 2014.
- [23] ZHANG K H, SONG H H. Real-time visual tracking via online weighted multiple instance learning [J]. *Pattern Recognition*, 2013, 46(1): 397-411.
- [24] ZHANG K H, ZHANG L, YANG M H. Real-Time compressive tracking[J]. *European Conference on Computer Vision*, 2012, 7574:864-877.
- [25] EVERINGHAM M, GOOL L, WILLIAMS C, *et al.*. The Pascal Visual Object Classes (VOC) challenge[J]. *International Journal of Computer*

- Vision*, 2010, 88(2): 303-338.
- [26] BABENKO B, YANG M-H, BELONGIE S. Robust object tracking with online multiple instance learning [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(8): 1619-1632.

## 作者简介:



张 雷(1987—),男,吉林长春人,博士研究生,2011 年于吉林大学获得学士学位,主要从事数字图像处理、目标跟踪等方面的研究。E-mail : zhanglei8080@126.com

## 导师简介:



王延杰(1963—),男,吉林长春人,研究员,博士生导师,主要从事实时图像处理、电视跟踪和自动目标识别技术的研究。E-mail: wangyj@ciomp.ac.cn

(版权所有 未经许可 不得转载)