基于判别字典在线学习的视觉跟踪

doi:10.3969/j.issn.1674-7100.2019.02.013

司　元[1]　朱文球[1, 2]

1. 湖南工业大学
计算机学院
湖南　株洲　412007
2. 湖南工业大学
智能信息感知及处理技术
湖南省重点实验室
湖南　株洲　412007

摘　要：提出了一种基于判别字典在线学习的跟踪算法，通过将字典项与标签信息相结合，分类的字典既具有重构性，又具有鉴别性。为了增强模型判别能力，将分类器嵌入到目标表示模型中，依据重构误差和判别分类得分最终确定候选目标。字典学习阶段采用在线字典学习算法同时对字典和分类器进行更新，使模型能够适应目标外观和背景环境的动态变化。实验结果表明，该方法在大量遮挡、快速运动、强光和姿态变化的大部分测试中达到了比较满意的效果。

关键词：稀疏编码；标签信息；字典学习；判别分类
中图分类号：TP391.41　　**文献标志码**：A
文章编号：1674-7100(2019)02-0087-10
引文格式：司　元，朱文球 . 基于判别字典在线学习的视觉跟踪 [J]. 包装学报，2019，11(2)：87-96.

1　研究背景

目标跟踪在计算机视觉领域起着重要的作用，可作为人机交互、机器人导航和智能交通等许多应用场景的预处理步骤。到目前为止，研究者已经提出了众多的目标跟踪方法，其稳定性、实时性和鲁棒性取得了重大进展。但是由于受目标姿态变化、快速运动、光照变化、物体遮挡、摄像机运动以及背景杂波等因素的影响，目标跟踪仍是一个具有挑战性的任务。

目标外观模型的跟踪算法大致可以分为生成式、判别式和混合式方法。生成式方法将跟踪问题确定为搜索最符合目标方案的区域。它们通常使用特定的特性构造健壮的对象表示模型，其方法包括子空间表示 [1]、基于片段的表示 [2] 和局部描述符 [3]。生成式跟踪器有平均偏移跟踪器、VTD（visual tracking decomposition）跟踪器 [4]。在判别式方法中，跟踪被视为一个二分类问题，目标是找到一个最佳的目标与背景分离的决策边界。判别式方法与生成式方法的主要不同之处在于对与目标模型的重构，判别模型不仅具有模型生成能力，而且还考虑了模型的识别能力，通常会具有更好的跟踪性能。主要的判别式方法有在线多示例学习目标跟踪 [5]、联合训练跟踪 [6]、集成跟踪 [7]。混合式方法利用了前两种方法的优势。Zhong W. 等 [8] 提出了一种协同整体模板和局部块的稀疏跟踪算法。Zhou T. F. 等 [9] 开发了对象跟踪的混合模型，目标用不同的外观流形来表示。Dou J. F. 等 [10] 提出了将结构局部稀疏模型和支持向量机的判别分类器相结合的跟踪方法。

最早的字典学习算法是由 L. Matthews 等 [11] 提出的一种基于 Lucas-Kanade 算法的更新方法，该方法首先估计跟踪结果，然后在第一个帧框架中使用模板来确定目标位置。H. Grabner 等 [12] 除了监督判别对

收稿日期：2018-10-12
基金项目：湖南省重点研发计划基金资助项目（2016GK2017）
作者简介：司　元（1978-），男，河南荥阳人，湖南工业大学硕士生，主要研究方向为数字图像处理，目标跟踪，E-mail：siyuan1979@163.com
通信作者：朱文球（1969-），男，湖南工业大学教授，硕士生导师，主要从事图像处理，模式识别等方面的研究，E-mail：wenqiu_zhu@126.com

象外，还将所有的视觉信息在跟踪结果中作为未标记的数据进行处理，并在半监督学习框架内对分类器进行调整。B. Babenko 等 [13] 用多实例学习（multiple instance learning，MIL）处理在线获得的模糊标记的正负数据以减少视觉漂移。Z. Kalal 等 [14] 也将分类器的跟踪结果视为未标签数据，利用其底层结构来选择正负样本进行更新。J. Mairal 等 [15] 则在原有的字典学习模型中引入一个更加成熟的 softmax 损失函数来增强字典的判决性。

在 J. Mairal 等研究成果的基础上，借鉴重构判别模型，本文提出了一种基于粒子滤波框架的跟踪算法。在所提出的方法中，跟踪是一种分类任务，用来识别可能的目标样本，并在一个判别稀疏表示基础上使用在线训练的分类器来拒绝背景样本。与目标外观和背景信息的处理相似，分类器和稀疏表示字典是同时训练的。不同于批处理方法（K-SVD，K-Singular value decomposition）来获得训练字典，课题组使用在线字典学习策略来处理视觉跟踪任务中的动态外观变化，然后将判别分类表示模型与贪婪搜索模型相结合，形成一个鲁棒跟踪算法。

2 判别表示模型

给定一组训练样本 $\boldsymbol{X}=\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$，类标签 $y_i=\{-1, +1\}$，每个样本 $\boldsymbol{x}_i$ 包含目标样本和背景样本共 n 个训练样本。字典 $\boldsymbol{D}=\{\boldsymbol{d}_1, \boldsymbol{d}_2, \cdots, \boldsymbol{d}_k\}$，其中 $\boldsymbol{d}_i$ 为字典中的原子，k 为字典维数。定义 $\boldsymbol{c}_i$ 为 $\boldsymbol{x}_i$ 在 $\boldsymbol{D}$ 上的编码系数，$i=1, 2, \cdots, n$，因此 $\boldsymbol{x}_i \approx \boldsymbol{D}\boldsymbol{c}_i$ 表示测试样本。字典学习方程为

$$J(\boldsymbol{D}, \boldsymbol{c}_i)=\min_{\boldsymbol{c}_i}\|\boldsymbol{x}_i-\boldsymbol{D}\boldsymbol{c}_i\|_2^2+\lambda\|\boldsymbol{c}_i\|_1, \tag{1}$$

式中：λ 为正则化系数；$\|\cdot\|_1$ 表示 L1 正则化。

目标跟踪不但要求字典能很好地重建样本，而且还需其具有一定的判别能力，J. Marial 等提出在原始字典学习中加入 logistic 损失函数（当用于多类识别时利用 softmax 损失函数），其判决项为

$$\langle \boldsymbol{D}, \boldsymbol{w}\rangle=\min_{(\boldsymbol{D}, \boldsymbol{w})}\sum l(y_i, f(\boldsymbol{c}_i, \boldsymbol{w}))+\lambda_2\|\boldsymbol{w}\|_2^2, \tag{2}$$

式中：$l(y_i, f(\boldsymbol{c}_i, \boldsymbol{w}))$ 为 logistic 损失函数，$\boldsymbol{w} \in \mathbf{R}^{n\times k}$ 为分类器 f 的分类参数，$y_i \in \{-1, +1\}$ 是对应的标签，$l(x)=\lg(1+\mathrm{e}^{-x})$；$\lambda_2$ 为分类器参数的正则化系数；$\|\cdot\|_2$ 表示 L2 正则化。

将判决项加入原始的字典学习模型，即

$$S(\boldsymbol{c}_i, \boldsymbol{x}_i, \boldsymbol{D}, \boldsymbol{w}, y_i)=l(y_i, f(\boldsymbol{c}_i, \boldsymbol{w}))+\lambda_1\|\boldsymbol{x}_i-\boldsymbol{D}\boldsymbol{c}_i\|_2^2+\lambda_2\|\boldsymbol{c}_i\|_1。 \tag{3}$$

式中 λ_1 用于控制重构项的权重。模型中，损失函数重构系数和分类器参数两项均进行了正则化约束。将 $(\boldsymbol{x}_i, y_i)$ 相关的损失项定义为

$$S^*(\boldsymbol{c}_i, \boldsymbol{x}_i, \boldsymbol{D}, \boldsymbol{w}, y_i)=\min_{\boldsymbol{c}_i} l(y_i, f(\boldsymbol{c}_i, \boldsymbol{w}))+\lambda_1\|\boldsymbol{x}_i-\boldsymbol{D}\boldsymbol{c}_i\|_2^2+\lambda_2\|\boldsymbol{c}_i\|_1。 \tag{4}$$

在此基础上 J. Marial 又提出了具有更强判决能力的模型，即正确标签值时 $S^*(\boldsymbol{c}_i, \boldsymbol{x}_i, \boldsymbol{D}, \boldsymbol{w}, y_i)$ 很小，且 $S^*(\boldsymbol{c}_i, \boldsymbol{x}_i, \boldsymbol{D}, \boldsymbol{w}, y_i)$ 比 $S^*(\boldsymbol{c}_i, \boldsymbol{x}_i, \boldsymbol{D}, \boldsymbol{w}, -y_i)$ 更小，

$$\min_{(\boldsymbol{D}, \boldsymbol{w})}\Big(\sum l\big(S^*(\boldsymbol{c}_i, \boldsymbol{x}_i, \boldsymbol{D}, \boldsymbol{w}, -y_i)-S^*(\boldsymbol{c}_i, \boldsymbol{x}_i, \boldsymbol{D}, \boldsymbol{w}, y_i)\big)+\lambda_3\|\boldsymbol{w}\|_2^2\Big)。 \tag{5}$$

式（5）很难求解，J. Marial 等利用参数 μ 控制生成项和判决项之间的权重，从而得到：

$$\min_{(\boldsymbol{D}, \boldsymbol{w})}\Big(\sum \mu l(S^*(\boldsymbol{c}_i, \boldsymbol{x}_i, \boldsymbol{D}, \boldsymbol{w}, y_i)-S^*(\boldsymbol{c}_i, \boldsymbol{x}_i, \boldsymbol{D}, \boldsymbol{w}, -y_i))+(1-\mu)S^*(\boldsymbol{c}_i, \boldsymbol{x}_i, \boldsymbol{D}, \boldsymbol{w}, y_i)+\lambda_3\|\boldsymbol{w}\|_2^2\Big)。 \tag{6}$$

为了得到理想稀疏编码 $\boldsymbol{c}_{\mathrm{L}}$，字典 $\boldsymbol{D}$ 应很好地重构一种信号，而对另一类信号没有好的重构性。$\boldsymbol{c}_{\mathrm{L}}$ 可由损失函数式（7）求得：

$$\begin{cases} S^*(\boldsymbol{c}_i, \boldsymbol{x}_i, \boldsymbol{D}, \boldsymbol{w}, y_i)=\min_{\boldsymbol{c}} S(\boldsymbol{c}_i, \boldsymbol{x}_i, \boldsymbol{D}, \boldsymbol{w}, y_i), \\ S(\boldsymbol{c}_{\mathrm{L}}, \boldsymbol{x}_i, \boldsymbol{D}, \boldsymbol{w}, y_i)=\big(1-y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{c}_{\mathrm{L}}+b)\big)^2+\lambda_1\|\boldsymbol{x}_i-\boldsymbol{D}\boldsymbol{c}_{\mathrm{L}}\|_2^2+\lambda_2\|\boldsymbol{c}_{\mathrm{L}}\|_1。 \end{cases} \tag{7}$$

式（7）并不是一个凸问题，可利用随机梯度下降法来获得一个局部最优解。

3 在线字典学习算法

3.1 在线判决字典学习算法

在线判决字典学习算法包括 2 个阶段：稀疏编码阶段和字典更新阶段。稀疏编码阶段，在第一帧中的初始目标区域，对目标位置周围的前景模板以及离目标一定像素的环形区域内的背景模板进行采样，形成正负训练样本。然后在正负样本上分别运行 K-SVD，形成 2 个相同大小的字典。然后将它们结合在一起形成初始字典 $\boldsymbol{D}$，在字典 $\boldsymbol{D}$ 下用式（3）和（7）计算 $\boldsymbol{x}_i$ 的目标样本和背景样本系数及真实标签系数 $\boldsymbol{c}_i^+$，$\boldsymbol{c}_i^-$，$\boldsymbol{c}_i^{\mathrm{L}}$，根据得到的系数利用存储矩阵 $\boldsymbol{A}$、$\boldsymbol{B}$ 求出字典 $\boldsymbol{D}$。在随后的学习中，更新 d_i 值，每个字典项目的标签

保持不变。字典更新阶段，判决字典学习算法用块坐标下降算法同时更新字典和分类器。具体算法如下。

算法 1　在线字典学习算法

输入：训练样本 $\boldsymbol{X}\in\mathbf{R}^{m\times n}$（$m$ 为样本维数，n 为样本个数）；标签 $y_1, y_2, \cdots, y_n\in\mathbf{R}^n$；$\lambda_1$；$\lambda_2$。

输出：字典 $\boldsymbol{D}$ 和分类器参数 $\boldsymbol{w}$。

步骤：

1）用 K-SVD 方法获得初始字典 $\boldsymbol{D}_0\in\mathbf{R}^{n\times k}$，$\boldsymbol{A}\in\mathbf{R}^{n\times k}\leftarrow 0$，$\boldsymbol{B}\in\mathbf{R}^{k\times k}\leftarrow 0$（$\boldsymbol{A}$、$\boldsymbol{B}$ 为求字典 $\boldsymbol{D}$ 的存储矩阵）。

2）排列训练样本 $\boldsymbol{x}_i$，i=1, 2,…, n。

3）使用式（3）和（7）计算正负样本稀疏码 $\boldsymbol{c}_i^+$、$\boldsymbol{c}_i^-$ 和理想稀疏码 $\boldsymbol{c}_i^{\mathrm{L}}$，并得到 $\boldsymbol{A}$、$\boldsymbol{B}$。

$$\boldsymbol{c}_i^+=\min_{\boldsymbol{c}_i} S\left(\boldsymbol{c}_i,\boldsymbol{x}_i,\boldsymbol{D},\boldsymbol{w},1,\lambda_1,\lambda_2\right),$$

$$\boldsymbol{c}_i^-=\min_{\boldsymbol{c}_i} S\left(\boldsymbol{c}_i,\boldsymbol{x}_i,\boldsymbol{D},\boldsymbol{w},-1,\lambda_1,\lambda_2\right),$$

$$\boldsymbol{c}_i^{\mathrm{L}}=\min_{\boldsymbol{c}_i} S\left(\boldsymbol{c}_i,\boldsymbol{x}_i,\boldsymbol{D},\boldsymbol{w},y_i,\lambda_1,\lambda_2\right);$$

$$\boldsymbol{A}\leftarrow\boldsymbol{A}+\boldsymbol{c}_i\boldsymbol{x}_i^{\mathrm{T}},\ \boldsymbol{B}\leftarrow\boldsymbol{B}+\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}};$$

$$\boldsymbol{u}_j\leftarrow\frac{1}{A_{jj}}\left(\boldsymbol{b}_j-\boldsymbol{D}\boldsymbol{x}_j\right)+\boldsymbol{d}_j,\quad \boldsymbol{d}_j\leftarrow\frac{1}{\max\left(\|\boldsymbol{u}_j\|_2,1\right)}\boldsymbol{u}_j,$$

其中，A_{jj} 为矩阵新加入的样本 $\boldsymbol{x}_j$ 的模，$\boldsymbol{u}_j$ 为存储向量，$\boldsymbol{b}_j$、$\boldsymbol{d}_j$ 分别为矩阵 $\boldsymbol{B}_i$ 和字典 $\boldsymbol{D}$ 中的列向量；

$$\boldsymbol{w}\leftarrow\frac{\boldsymbol{c}_i\boldsymbol{x}_i^{\mathrm{T}}}{\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}+\lambda_2\boldsymbol{I}}。$$

用算法 2 对字典 $\boldsymbol{D}$ 和分类器参数 $\boldsymbol{w}$ 进行更新；

结束。

算法 2　字典和分类器参数更新算法

输入：字典 $\boldsymbol{D}$，存储矩阵 $\boldsymbol{A}$、$\boldsymbol{B}$；

其中，$\boldsymbol{D}=\left[\boldsymbol{d}_1,\boldsymbol{d}_2,\ldots,\boldsymbol{d}_k\right]\in\mathbf{R}^{n\times k}$；

$$\boldsymbol{A}=\left[\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_k\right]\in\mathbf{R}^{k\times k}=\frac{1}{2}\sum_{i=1}^{t}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}};$$

$$\boldsymbol{B}=\left[\boldsymbol{b}_1,\boldsymbol{b}_2,\ldots,\boldsymbol{b}_k\right]\in\mathbf{R}^{n\times k}=\sum_{i=1}^{t}\boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}}。$$

输出：字典 $\boldsymbol{D}$ 和分类器参数 $\boldsymbol{w}$。

for j=1 to k

利用下式更新字典 $\boldsymbol{D}$ 中的每一个原子 $\boldsymbol{d}_j$，

$$\boldsymbol{d}_j=\frac{1}{\boldsymbol{a}_j}\left(\boldsymbol{b}_j-\boldsymbol{D}\boldsymbol{a}_j\right)+\boldsymbol{d}_j$$

更新 $\boldsymbol{D}$ 和 $\boldsymbol{w}$：

$$\boldsymbol{D}=\frac{\boldsymbol{d}_j}{\left\|\boldsymbol{d}_j\right\|_2}=\left[\frac{\boldsymbol{d}_1}{\left\|\boldsymbol{d}_1\right\|_2},\frac{\boldsymbol{d}_2}{\left\|\boldsymbol{d}_2\right\|_2},\cdots,\frac{\boldsymbol{d}_k}{\left\|\boldsymbol{d}_k\right\|_2}\right];$$

$$\boldsymbol{w}=\frac{\boldsymbol{w}_j}{\left\|\boldsymbol{d}_j\right\|_2}=\left[\frac{\boldsymbol{w}_1}{\left\|\boldsymbol{d}_1\right\|_2},\frac{\boldsymbol{w}_2}{\left\|\boldsymbol{d}_2\right\|_2},\cdots,\frac{\boldsymbol{w}_k}{\left\|\boldsymbol{d}_k\right\|_2}\right]。$$

end for。

3.2　跟踪预测

学习字典和分类器之后就可以对一个测试样本 $\boldsymbol{x}$ 跟踪预测，根据 $\boldsymbol{x}$ 和训练样本的相似度和分类器的分类误差进行预测。测试期间的联合决策度量误差定义为

$$\varepsilon(\boldsymbol{x})=\left\|\boldsymbol{x}_{\mathrm{tr}}-\boldsymbol{D}\boldsymbol{c}\right\|^2+\rho\left\|f-\boldsymbol{w}\boldsymbol{c}\right\|^2。\tag{8}$$

式中：$\boldsymbol{x}_{\mathrm{tr}}$ 为一个集合中样本元素的加权平均值；$\left\|\boldsymbol{x}_{\mathrm{tr}}-\boldsymbol{D}\boldsymbol{c}\right\|^2$ 为重构样本 $\boldsymbol{D}\boldsymbol{c}$ 和 $\boldsymbol{x}_{\mathrm{tr}}$ 之间的重构误差；$\left\|f-\boldsymbol{w}\boldsymbol{c}\right\|^2$ 为分类误差的线性回归损失；f 代表分类器；ρ 为一个控制线性回归损失的权重的常数。具有最小联合分类误差的图像帧最可能被确定为跟踪结果。

4　基于粒子滤波的目标跟踪

在本研究中，视觉跟踪是在粒子滤波框架下进行的，它使用有限的加权样本来递归地逼近后验分布。给定一组观测图像向量 $z_{1:t}$=[z_1, z_2 …, z_{t-1}]，直到 t−1 帧，视觉跟踪可以通过最大后验概率（maximum a posteriori estimation，MAP）来估计目标状态（即运动参数）。

$$\boldsymbol{T}_t=\arg\min\ p\left(\boldsymbol{T}_t\mid z_{1:t}\right)\tag{9}$$

式中：$\boldsymbol{T}_t$ 为第 t 帧的目标状态，$\boldsymbol{T}$=(x; y; s; r; h; k)，其中，x 和 y 为图像坐标，s 和 r 为尺度和方向，h 为旋转角，k 为倾斜度。

后验概率可以通过贝叶斯定理推断：

$$P\left(\boldsymbol{T}_t|z_{1:t}\right)\propto p\left(z_t|\boldsymbol{T}_t\right)p\left(\boldsymbol{T}_t|z_{1:t}\right)。\tag{10}$$

式中：$P\left(\boldsymbol{T}_t|z_{1:t}\right)=\int p\left(\boldsymbol{T}_t|\boldsymbol{T}_{t-1}\right)p\left(\boldsymbol{T}_{t-1}|z_{1:t-1}\right)\mathrm{d}\boldsymbol{T}_{t-1}$ 为粒子相似性后验概率，其中 $p\left(\boldsymbol{T}_t\mid\boldsymbol{T}_{t-1}\right)$ 是描述连续帧中目标状态的时间相关性的动态模型，$p\left(z_t|\boldsymbol{T}_t\right)$ 是观测模型，$p\left(\boldsymbol{T}_t\mid\boldsymbol{T}_{t-1}\right)$ 估计一个观察状态的可能性。观测模型 $p\left(z_t|\boldsymbol{T}_t\right)$ 反映了一个候选目标和字典模板之间的相似性；$p\left(z_t|\boldsymbol{T}_t\right)$ 是联合度量比例；$p\left(\boldsymbol{T}_t|z_{1:t}\right)$ 可以用一组有限的粒子来近似。本算法使用仿射变换对两个连续帧之间的目标对象的运动进行建模。用两个连续帧之间的高斯分布来独立地模拟每个参数的变换。然后，动态模型可以用高斯分布 $p\left(\boldsymbol{T}_t|\boldsymbol{T}_{t-1}\right)=N\left(\boldsymbol{T}_t,\boldsymbol{T}_{t-1},\Sigma\right)$ 表示，其中 Σ 为对角线协方差矩阵，其元素是仿射参数的方差。观测模型由式（11）定义：

$$p(\boldsymbol{Z}|\boldsymbol{T})\propto\varepsilon(\boldsymbol{X})。\quad(11)$$

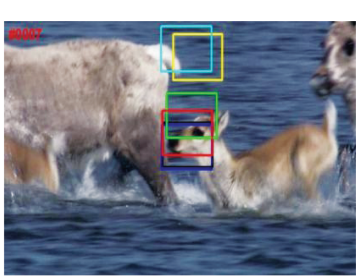
5 实验结果与分析

提出的跟踪算法使用 Matlab R2015b 实现。电脑硬件配置环境为 2.66 GHz 的 Intel Pentium 处理器和 8 G 内存。主要使用的软件环境为 Windows 系统。参数设置如下：字典大小固定为 200，其中包含 100 个正样本，100 个负样本。初始化和在线学习的迭代次数分别为 5 和 30；在线学习速率为 0.2；在第一帧中，正负样本数目都为 200，更新数都为 100。每个帧随机样本的候选数为 800。

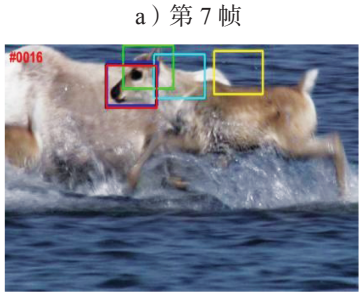
为了比较跟踪器和其他基于状态稀疏表示的跟踪器的性能，课题组将跟踪器与 4 个公共序列的 5 个先进的跟踪器进行比较，选择 SCM（robust object tracking via sparsity-based collaborative mode） 跟踪器、PCOM（visual tracking via probability continuous outlier model） 跟踪器、LSST（least soft-thresold squares tracking） 跟踪器、OTSP（online object tracking with sparse prototypes）跟踪器和增量视觉跟踪器（incremental learning for robust visual tracking，IVT）。测试序列包括视频跟踪中的常见场景，如快速运动、姿态变化、遮挡、尺度变化和模糊。通过这些实验可以验证跟踪器的有效性，定性跟踪结果如图 1~4 所示。

5.1 定性分析

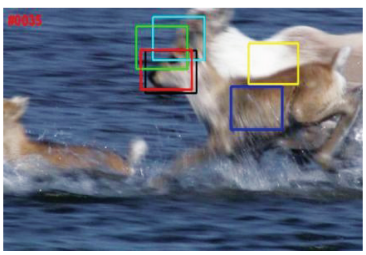
在 Animal 视频序列中，目标经历姿势变化、背景杂乱、视频模糊以及部分遮挡等干扰，如图 1 所示。



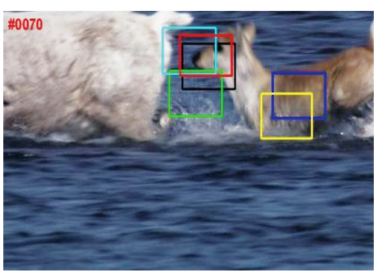
a）第 7 帧



b）第 16 帧



c）第 35 帧



d）第 70 帧

注：—SCM；—PCOM；—OTSP；—LSST；—ITV；—Ours；下同。

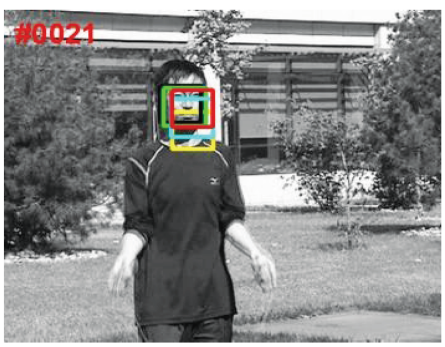
图 1 不同方法对 Animal 视频序列跟踪情况

Fig. 1 Different methods for tracking Animal video sequences

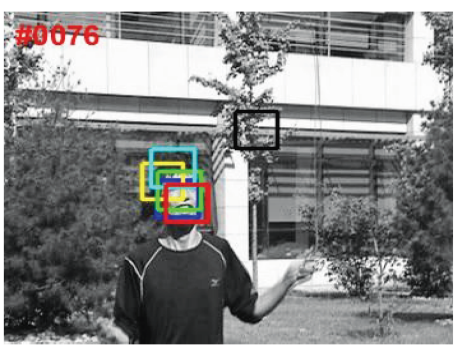
由图 1 可知，大多数的跟踪方法失败或漂移远离目标实体。ITV 和在线字典学习方法有更好的 OR（overlap rate），这表明跟踪结果更稳定。由于目标的快速运动，OTSP 和 LSST 在第 7 帧发生了目标漂移，在后续的视频序列中，除 PCOM 跟踪器返回目标，另外三种跟踪器则表现为目标漂移或跟踪失败。虽然 PCOM 在漂移后具有重新寻找目标机制，但它的跟踪精度和重叠率都低于在线字典学习跟踪器。相比之下，在线字典学习跟踪器在整个视频序列中能够提供更准确和一致的跟踪边界框。尤其在模糊视频帧中（如图 1c 所示），在线字典学习依旧表现良好。这归因于学习的判别字典，其可以编码目标和背景之间的细微视觉差异。

在 Jumping 视频序列中，展示了一个复杂的场景，包括快速运动、视频模糊等干扰，如图 2 所示。在第 21 帧前，目标处于缓慢运动状态，这几种比较算法基本都可以稳定跟踪目标。在第 76 帧时，由于目标发生了快速运动，除了 SCM 跟踪器，其他跟踪器包括在线学习字典学习跟踪器都产生了一定的偏离，在接下来视频帧中，其他算法都无法跟踪目标，但从第 205 帧开始，尽管序列出现了模糊和快速运动，在线字典学习跟踪器逐渐跟踪到目标实体。在该视频序列中，除 SCM 跟踪器，在线字典学习方法产生了次优的跟踪效果，这是由于在线字典学习算法需要

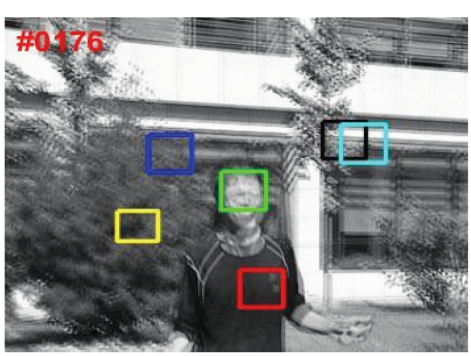
计算两次正负样本字典，增加了算法的时间复杂度，因此在目标快速运动与外观变化的视频序列中尚不能达到最优的效果。



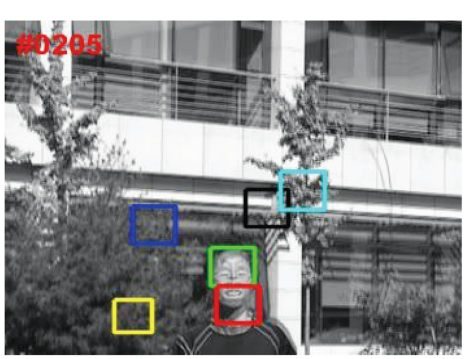
a）第 21 帧



b）第 76 帧



c）第 176 帧



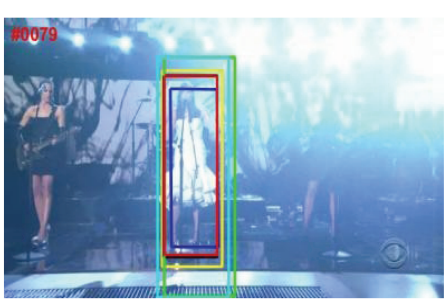
b）第 205 帧

图 2　不同方法对 Jumping 序列跟踪情况

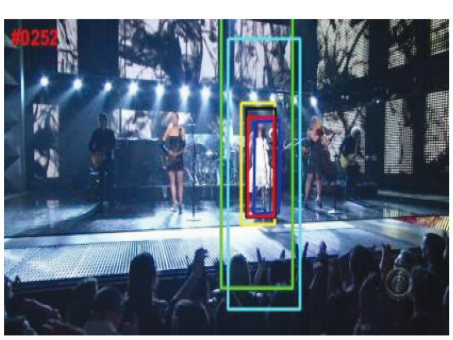
Fig. 2　Different methods for tracking Jumping sequences

在 Singer1 视频序列中，展示了一个光照发生明显变化的场景，如图 3 所示。各跟踪器仅在前 3 帧目标保持一致跟踪，从第 4 帧开始出现偏差，在第 79 帧时光照发生明显变化，SCM 和 OTSP 跟踪框都产生了大幅度尺度变化，在线字典学习跟踪器和另外 3 个跟踪器 LSST、ITV、PCOM 产生了可比的跟踪效果，相比之下，LSST 和在线字典学习的方法表现更好。在线字典学习跟踪器对照明变化的稳健性说明所学习的特征已能够对照明变化产生适应性。在 Singer1 序列中，除了光照变化之外，目标还经历姿势变化。当 Singer1 身体发生转动时（如图 3c 所示），SCM 和 OTSP 跟踪器产生尺度变化，另外 3 种跟踪器和在线字典学习跟踪器能够使用相对精确的边界框大小成功跟踪整个序列中的对象。相比较而言，在这 3 种跟踪器中 LSST 跟踪器表现较好，但成功率低于在线字典学习的方法。

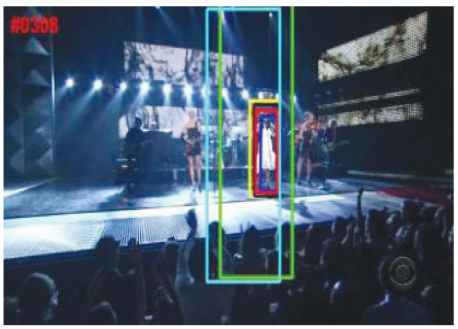
a）第 3 帧



b）第 79 帧



c）第 252 帧



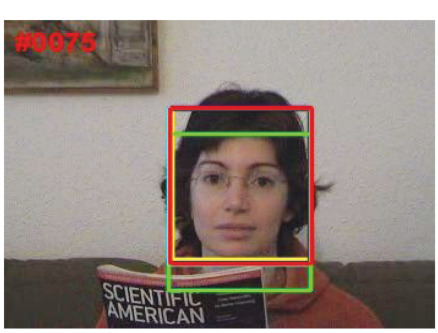
d）第 308 帧

图 3　不同方法对 Singer1 视频序列跟踪情况

Fig. 3　Different methods for tracking Singer1 video sequences

Occlusion 视频序列展示了一个遮挡的场景，目标经历平面内旋转和频繁遮挡，如图 4 所示。从第 30 帧开始，出现了局部遮挡的情况，在遮挡出现的前后帧中，其他跟踪器对目标都有所偏移，当部分遮挡发生时（如图 4a 所示），SCM 方法随着时间逐渐偏离目标。只有在线字典学习的方法能够将跟踪框保持在目标上（如图 4c 所示）。从第 568 帧开始目标被全部遮挡，LSST、PCOM、ITV、OTSP 4 种跟踪方法均对目表产生偏离，而在线字典学习方法始终能够将跟踪框保持在目标上。虽然这 4 种跟踪器也能够追踪目标，但在线字典学习方法实现了较低的跟踪误差和较高的重叠率。总的来说，在线字典学习方法能够很好地跟踪对象并提供更准确和一致的跟踪框。

a）第 30 帧



b）第 75 帧

c）第 194 帧

d）第 568 帧

图 4　不同方法对 Occlusion 视频序列跟踪情况

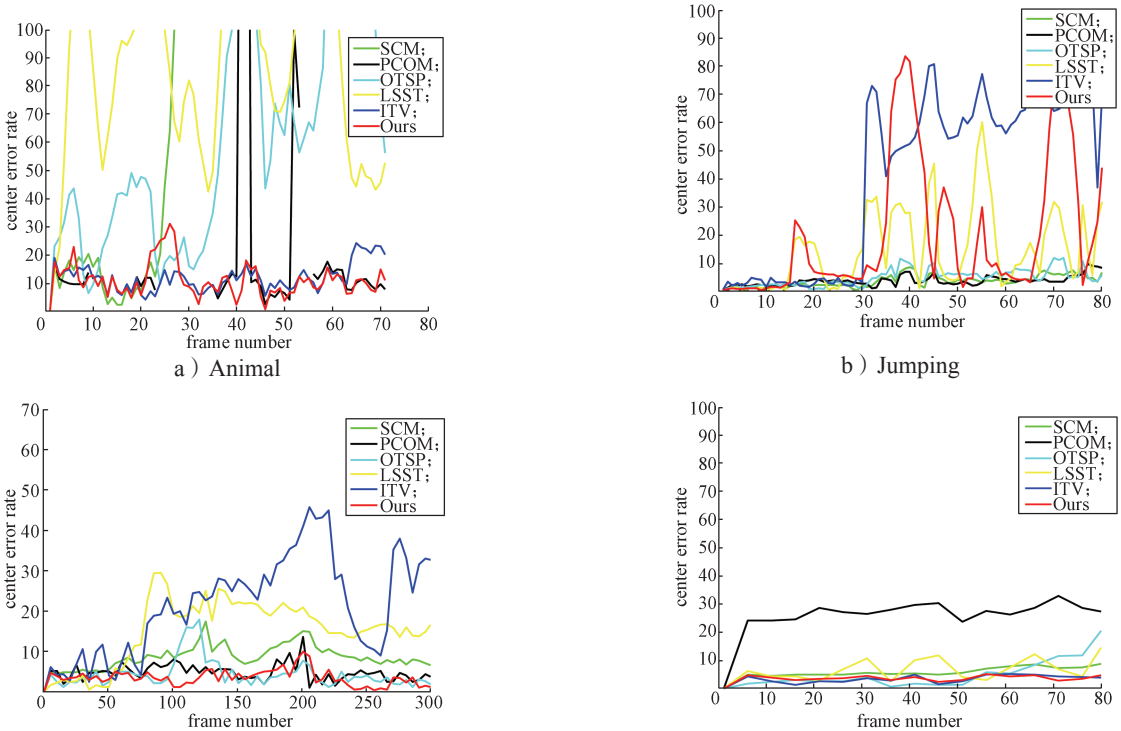
Fig. 4　Different methods for tracking Occlusion video sequences

5.2　定量比较

为了在 6 种方法之间进行定量比较，采用中心位置误差率（center error rate，CER）和重叠率（overlap rate，OR）（见图 5、6）两种评估标准。CER 为预测中心位置和实际中心位置之间的距离，以像素为单位。重叠率为预测目标跟踪框和实际跟踪框的交集与 2 个跟踪框并集之比。

由图 5 可知，本跟踪器表现良好；在线字典学习算法除了在 Jumping 视频序列中波动比较大，在 Animal、Singer1 和 Occlusion 视频序列中均具有最低的中心位置误差率。由此可以看出，在姿态变化、遮挡、光照变化和模糊等因素影响下，在线字典学习算法均能够产生鲁棒的跟踪结果。

从图 6 中可以看出，所提出的方法在 4 个视频序列中的 3 个中最精确地跟踪目标，并且在另一个序列中实现第二最佳性能。另一方面，在线字典学习跟踪器在更具挑战性的测试视频序列（例如，Animal、Singer1 和 Occlusion）中具有更好的性能，即基于生成稀疏表示的跟踪器通常具有大的跟踪位置误差和故障率。因为它利用了稀疏表示的灵活描述并且使分类考虑背景信息。

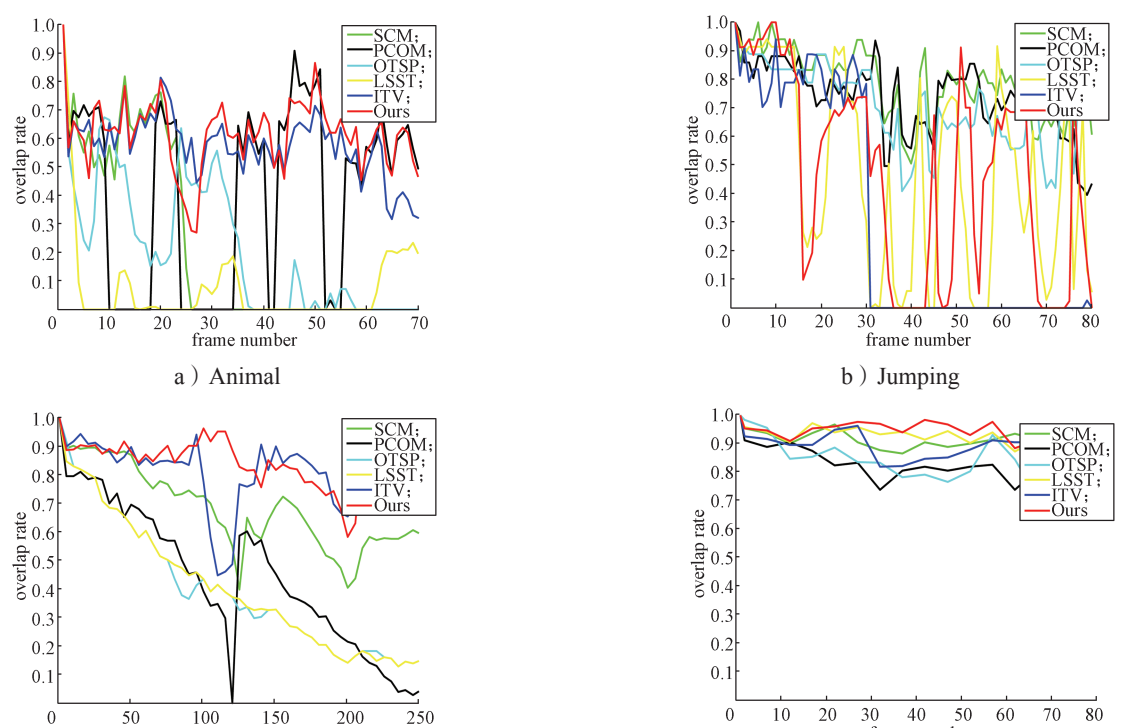


a）Animal　b）Jumping

c）Singer1　d）Occlusion

图 5　不同视频序列中心误差率曲线

Fig. 5　Different video sequences central error rate curves



a）Animal　b）Jumping

c）Singer1　d）Occlusion

图 6　不同视频序列重叠率曲线

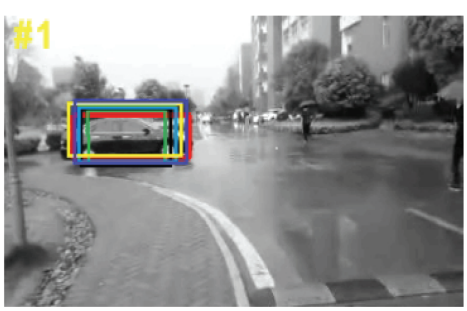
Fig. 6　Overlapping coefficient curves of different video sequences

5.3 论证

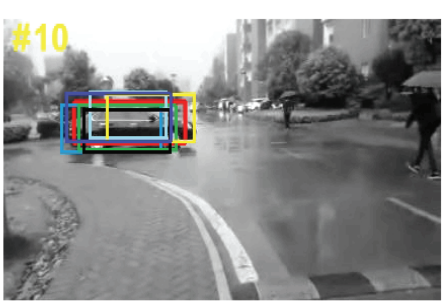
最后用自拍视频（本视频 2018 年 4 月取材于湖南工业大学崇真楼前，跟踪目标为一辆黑色轿车，视频图像共 138 帧）进行算法验证并且给出了比较算法的中心误差率曲线图和重叠率曲线图。

5.3.1 定性评估

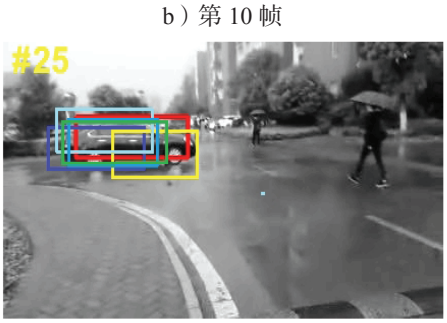
尺度变化、平面内（外）旋转以及快速运动都会影响跟踪算法的性能。自拍的视频序列中包含了尺度变化、平面外旋转和快速运动等复杂因素。这些属性以交互或并发的方式存在，因此，跟踪控制比较困难。跟踪比较结果如图 7 所示，在整个视频序列中，所提出的算法始终保持了精确的跟踪，且优于其他比较算法，这也证明了所提出算法的稳定跟踪性能，也证明了字典学习在跟踪方面的优良表现。所提出的算法需要学习目标及周围区域的正负模板以构建分类字典（即正负字典），这会显著提高算法的时间复杂度，因此所提出的算法实时性能并不乐观，这是算法还需要改进的地方。



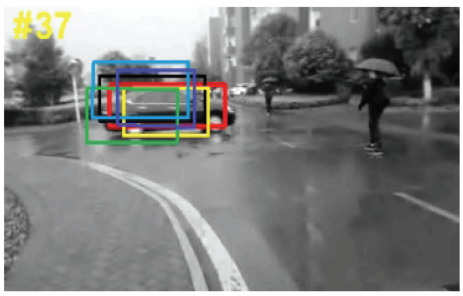
a）第 1 帧



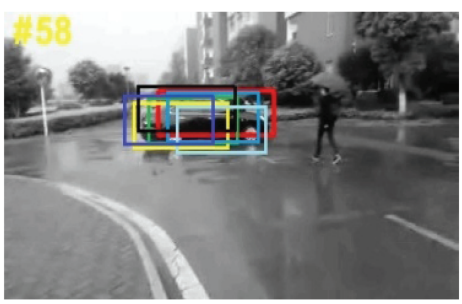
b）第 10 帧



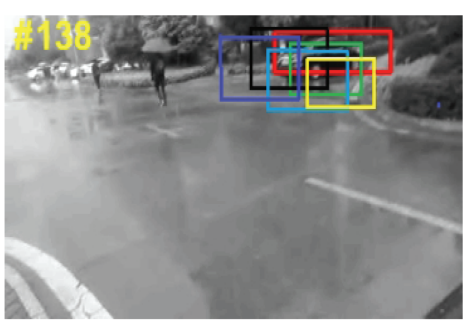
c）第 25 帧



d）第 37 帧



e）第 58 帧



f）第 138 帧

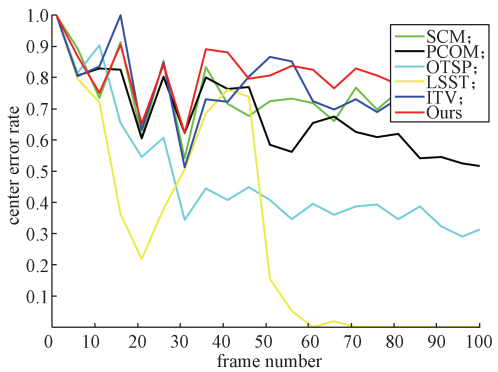
图 7 自拍视频跟踪比较结果

Fig. 7 Self-timer video tracking comparison results

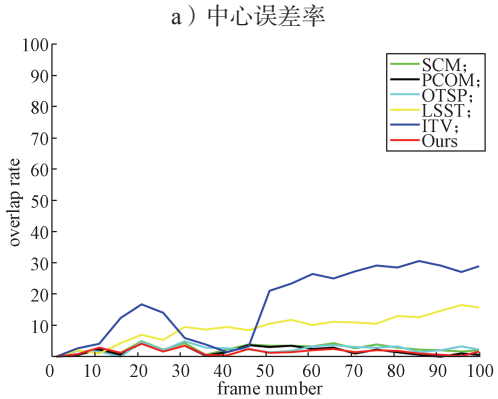
5.3.2 定量评估

将所提出的算法与上述基于稀疏表示的跟踪器在 OTB（ object visual tracking ）数据集上进行比较。通过 CER 和 OR 曲线图得到的结果如图 8 所示。从两类曲线图可以看出，所提出的算法较其他算法明显改进了跟踪效果。在精确率曲线图中，提出的算法在跟踪开始时精确率波动幅度较大，并且其他算法也有此中表现，这是因为算法在起初跟踪时段需要构造稀疏表示字典，对目标模型的表示尚不确切，这段时间内还不足以评价各种算法的优劣性，但随着字典趋于完备，算法性能差异逐渐显现。相比较而言，所提出的算法其精确度的波动幅度大大减小，并且精确率始终领先于所比较的算法。如图 8a）所示所提出的算法获得了优于 ITV 算法的最优效果。在重叠率（即成功率）方面，比较的结果与精确度并不一致，ITV 最高，

这也说明了 ITV 算法是以重叠率换取精确度的，而所提出的算法在保持精度的基础上取得了与 SCM、PCOM 与 OTSP 可比的重叠率，并且效果稍微优于这三种算法。



a）中心误差率



b）重叠率

图 8　自拍视频比较曲线

Fig. 8　Self-timer video comparison curves

6　结语

本文提出的基于稀疏表示的判别式在线字典学习跟踪算法在跟踪过程中，通过重构误差和分类误差的联合评价，选出最佳候选对象，并将可靠的跟踪结果作为训练样本用于字典更新。实验结果表明，课题组提出的方法能够应对各种场景因素，在线字典学习和联合决策都有助于提高跟踪性能。

参考文献：

[1] ROSS D A，LIM J，LIN R S，et al. Incremental Learning for Robust Visual Tracking[J]. International Journal of Computer Vision，2008，77(1/2/3)：125-141.

[2] ADAM A，RIVLIN E. Robust Fragments-Based Tracking Using the Integral Histogram[C]//2006 IEEE Computer Vision and Pattern Recognition (CVPR). New York：IEEE，2006：798-805.

[3] HE W，YAMASHITA T，LU H，et al. Surf Tracking [C]//2009 IEEE 12th International Conference on Computer Vision. Kyoto：IEEE，2009：1586-1592.

[4] KWON J，LEE K M. Visual Tracking Decomposition[C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco：IEEE，2010：1269-1276.

[5] BABENKO B，YANG M，BELONGIE S. Visual Tracking with Online Multiple Instance Learning[C]// 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami：IEEE，2009. DOI：10.1109/CVPR.2009.5206737.

[6] LIU R，CHENG J，LU H，A Robust Boosting Tracker with Minimum Error Bound in a Co-Training Framework[C]// 2009 IEEE 12th International Conference on Computer Vision. Kyoto：IEEE，2009：1459-1466.

[7] AVIDAN S. Ensemble Tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence 2007. Washington：IEEE：261-271.

[8] ZHONG W，LU H，YANG M H. Robust Object Tracking Via Sparse Collaborative Appearance Model[J]. IEEE Transactions on Image Processing，2014，23(5)：2356-2368.

[9] ZHOU T F，LU Y，DI H J. Locality-Constrained Collaborative Model for Robust Visual Tracking[J]. IEEE Transactions on Circuits & Systems for Video Technology，2015，27(2)：313-325.

[10] DOU J F，QIN Q，TU Z M. Robust Visual Tracking Based on Generative and Discriminative Model Collaboration[J]. Multimedia Tools Applications，2017，76(14)：15839-15866.

[11] MATTHEWS L，ISHIKAWA T，BAKER S. The Template Update Problem[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2004，26(6)：810-815.

[12] GRABNER H，LEISTNER C，BISCHOF H. Semi-Supervised On-Line Boosting for Robust Tracking[C]// 2008 European Conference on Computer Vision. Munich：ECCV，2008：234-247.

[13] BABENKO B，YANG M H，BELONGIE S. Visual Tracking with Online Multiple Instance Learning[C]// 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami：IEEE，2009：983-990.

[14] KALAL Z，MATAS J，MIKOLAJCZYK K. P-N Learning：Bootstrapping Binary Classifiers by Structural Constraints[C]//2010 IEEE Computer Society Conference

on Computer Vision and Pattern Recognition. San Francisco：IEEE，2010：49-56.
[15] MAIRAL J，BACH F，PONCE J，et a1. Online Learning for Matrix Factorization and Sparse Coding[J]. Journal of Machine Learning Research，2010，11：19-60.

（责任编辑：申　剑）

Visual Tracking Based on Online Learning of Discriminant Dictionaries

SI Yuan[1]，ZHU Wenqiu[1, 2]

（1. College of Computer，Hunan University of Technology，Zhuzhou Hunan 412007，China；2. Key Laboratory of Intelligent Information on Perception and Processing Techonology，Hunan University of Technology，Zhuzhou Hunan 412007，China）

Abstract：A tracking algorithm based on online learning of discriminant dictionary has been proposed, which combines dictionary items with label information, thus making the discriminant dictionary both reconstructive and discriminant. In order to enhance the discriminant ability of the model, the classifier is embedded in the target representation model, with its candidate target to be determined according to the reconstruction error and discriminant classification score. The online dictionary learning algorithm is used to update both the dictionary and the classifier in the process of learning, so that the model can adapt to the dynamic changes of the target appearance and background environment. The experimental results show that the proposed method achieves satisfactory results in most of such tests concerning large occlusion, fast motion, strong light and attitude change.

Keywords：sparse coding；label information；dictionary learning；discriminant classification