

基于目标分块多特征核稀疏表示的视觉跟踪

胡昭华^{1,2} 袁晓彤³ 李俊⁴ 何军¹
¹(南京信息工程大学电子与信息工程学院 南京 210044)
²(江苏省大气环境与装备技术协同创新中心 南京 210044)
³(南京信息工程大学信息与控制学院 南京 210044)
⁴(南京理工大学计算机科学与工程学院 南京 210094)
(zhaohua_hu@163.com)

Robust Fragments-Based Tracking with Multi-Feature Joint Kernel Sparse Representation

Hu Zhaohua^{1,2}, Yuan Xiaotong³, Li Jun⁴, and He Jun¹
¹(School of Electronic & Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044)
²(Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, Nanjing 210044)
³(School of Information and Control, Nanjing University of Information Science & Technology, Nanjing 210044)
⁴(School of Computer Science and Engineering, Nanjing University of Science & Technology, Nanjing 210094)

Abstract Most existing sparse representation based trackers only use a single feature to describe the objects of interest and tend to be unstable when processing challenging videos. To address this issue, we propose a particle filter tracker based on multiple feature joint sparse representation. The main idea of our algorithm is to partition each particle region into multiple overlapped image fragments. Every local fragment of candidates is sparsely represented as a linear combination of all the atoms of dictionary template that is updated dynamically and is merely reconstructed by the local fragments of dictionary template located at the same position. The weights of particles are determined by their reconstruction errors to realize the particle filter tracking. Our method simultaneously enforces the structural sparsity and considers the interactions among particles by using mixed norms regularization. We further extend the sparse representation module of our tracker to a multiple kernel joint sparse representation module which is efficiently solved by using a kernelizable accelerated proximal gradient (KAPG) method. Both qualitative and quantitative evaluations demonstrate that the proposed algorithm is competitive to the state-of-the-art trackers on challenging benchmark video sequences with occlusion, rotation, shifting and illumination changes.

Key words visual tracking; kernel sparse representation; multi-feature association; particle filter; overlapped fragment

摘要 大多数现有的基于稀疏表示的跟踪器仅采用单个目标特征来描述感兴趣的目标,因而在处理各种复杂视频时不可避免会出现跟踪不稳定的情况.针对这个问题,提出一种基于多特征联合稀疏表示的粒子滤波跟踪算法.该算法的主要思想是对随时间不断更新的字典模板和抽样粒子的局部块依据其位置进行分类,用字典中所有类别块对抽样粒子的局部块进行稀疏表示,而仅用与字典中具有相同类别的

局部块及表示系数进行重构,根据重构误差构建似然函数以确定最佳粒子(候选目标),实现对目标的精确跟踪.该方法不仅实现了局部块的结构稀疏性,而且充分考虑了粒子之间的依赖关系,提高了跟踪精度.将算法进一步推广到采用基于核的多种特征描述,经混合范数约束并利用 KAPG(kernelizable accelerated proximal gradient)方法求解联合特征的稀疏系数.定性和定量的实验结果均表明该算法在目标发生遮挡、旋转、尺度变化、快速运动、光照变化等各种复杂情况下,依然可以准确地跟踪目标.

关键词 视觉跟踪;核稀疏;多特征联合;粒子滤波;重叠分块

中图法分类号 TP391.4

视觉目标跟踪是计算机视觉研究领域中的重要任务,可广泛应用于自动监控、机器人导航、人机接口等很多领域^[1].对于适用真实场景的稳健视觉跟踪算法,需要有效地处理目标外观及姿态的改变、快速运动、遮挡、光照影响、背景杂波等多方面的挑战.尽管人们在这方面已经做了大量的研究工作,但是仍存在很多困难有待解决.

近年来,稀疏表示(稀疏编码)被成功应用于视觉跟踪中^[2-7].稀疏表示跟踪器的主要思想是将每个候选目标表示成字典模板的一种稀疏线性组合,跟踪过程中需要对字典模板进行动态更新以迎合目标外观的改变.特别地,在目标出现部分遮挡时,基于稀疏表示的跟踪器则能显示出稳健的跟踪性能.然而,稀疏表示跟踪器在处理每帧时都需要执行复杂的 L1 最小化计算,在粒子滤波框架下计算量将随着抽样粒子数呈线性增加的趋势,从而限制了跟踪器的实时应用. Mei 等人^[8]和 Li 等人^[9]对稀疏表示算法进行了改进以提高跟踪速度. Mei 提出最小误差界和遮挡检测的有效 L1 跟踪器,从一个线性最小二乘公式中快速计算得到最小误差界,用以指导粒子滤波框架中的粒子再抽样,再抽样过程不损失精度,并且将大多数无用粒子在求解计算复杂的 L1 最小化问题之前去除掉,这样可大大减少计算量. Li 采用降维和正交匹配追踪(OMP)算法用以加速 L1 跟踪器.另外, Bao 等人^[10]则利用加速近邻梯度法(APG)求解以改进 L1 跟踪器.

目标描述及目标模型是实现精确跟踪的关键要素. Guo 等人^[11]提出了一种特征流形模型,将目标的局部特征建模成特征流形的形式,用以描述目标外观的平滑变化.并利用特征与流形之间的匹配过程确定目标的位置. Li 等人^[12]提出一种简单却富含信息量的目标特征描述模型——判别非正交二进制子空间(non-orthogonal binary subspace, NBS)模型,采用 NBS 描述目标,而利用判别信息有效区分目标和背景,从而实现对目标的跟踪.在基于稀疏表

示的目标跟踪中, Mei 等人^[2-3,8]采用了全局描述的目标外观模型,通过解 L1 最小化问题跟踪目标. Liu 等人^[5]提出了基于局部稀疏模型的跟踪算法,利用稀疏系数的直方图和均值漂移算法实现对目标的跟踪.然而,该方法是基于静态的局部稀疏字典,当场景中出现类似目标时会出现跟踪失败.针对以上缺点, Jia 等人^[13]提出了一种基于自适应局部结构稀疏模型的有效跟踪方法,该方法在目标区域进行图像重叠分块,充分利用目标的部分信息以及空间信息,并采用 alignment pooling 的方法处理目标局部块,最终实现目标的精确定位和遮挡处理.鉴于目标图像的重叠分块能更好地处理遮挡,因此本文对目标区域的处理也采用重叠分块的方式.

Zhong 等人^[14]针对目标外观发生显著变化所引起的挑战,提出同时利用全局模板和局部描述的稳健外观模型,发展了基于稀疏的判别分类器(sparsity-based discriminative classifier, SDC)和基于稀疏的生成模型(sparsity-based generative model, SGM): SDC 模型用于有效区分目标和背景, SGM 模型考虑块的空间信息以便处理遮挡问题; Wang 等人将经典的 PCA 算法和稀疏表示相结合提出了一种新的在线目标跟踪算法^[15],接着于 2013CVPR 上又提出了一种最小软阈值平方(least soft-threshold squares, LSS)算法^[16],用以建模跟踪中的高斯-拉普拉斯误差项分布,从而推出一种新的 LSS 距离度量准则用以度量观测样本和字典之间的差异,这种度量准则与通常的最小平方度量准则相比,可以更有效地处理跟踪过程中的异常值.

稀疏表示跟踪器的跟踪阶段一般基于粒子滤波框架,现有算法大都单独求取各个粒子的稀疏表示系数,并没有考虑各粒子之间的相互关系.特别是在目标外观发生明显变化时,忽略粒子之间的关系限制了粒子的表示,会使得跟踪器偏移真实目标.为了充分利用粒子之间相互关系信息, Zhang 等人先后提出了多任务跟踪器(multi-task tracking, MTT)^[17]

和结构化多任务跟踪器(S-MTT)^[18],将每个粒子的稀疏表示视为一个单任务,MTT通过 $l_{p,q}$ 混合范数计算粒子的联合稀疏系数,联合系数只是将所有粒子一起考虑的全局形式,通过观察发现某些任务(粒子表示)之间有更密切的关系,因此又进一步考虑粒子之间的局部空间结构关系,从而将MTT扩展成S-MTT. Mei等人^[8]的L1跟踪器是MTT框架下的一种特例.与此同时,Zhang等人^[19]为了考虑粒子间的关系,将低秩引入稀疏表示中,通过低秩稀疏学习得到稳健的跟踪.

以上方法有些在求解L1最小化问题上进行改进,有些考虑了各粒子之间的相互关系,还有些对粒子之间的关系加以约束并采用有效的稀疏系数求解方法,但这些方法都无一例外地仅使用目标图像的灰度特征,而没有考虑目标的其他有效特征.虽然灰度特征作为一种基本特征描述图像目标是可行的,然而针对不同的视频序列属性(例如光照、尺度变化、遮挡、运动模糊、快速运动、背景杂波、低分辨率等),它却不一定是最适合的特征描述,因此可联合其他一些特征来共同描述目标,不同特征之间的共享和互补可以改善单个特征的性能. Yuan等人^[20]

已将这一思想成功运用于图像分类. Wu等人^[21]为了更好地评估各种跟踪器的跟踪性能,将现有跟踪器所处理的各种视频序列按属性进行分类;而Li等人^[22]在视觉目标跟踪综述中总结了目标的各种视觉特征描述. 本文仅以灰度特征和HOG特征融合为例,表明多特征融合稀疏表示跟踪器的有效性,此种特征融合可以扩展到其他各类特征.

本文算法的主要思想如下:为了充分利用目标图像的结构信息,首先对目标图像区域进行重叠分块处理,计算每个重叠块的灰度特征和HOG特征,联合这2种特征对目标重叠块进行稀疏表示;考虑候选粒子(样本)之间的相互关系,采用核化加速近邻梯度算法(kernelizable accelerated proximal gradient, KAPG)联合求解候选粒子(样本)的稀疏系数;接着依据分类的思想,将不同位置的目标块看成不同类别,利用字典中与候选样本块具有相同类别的块及表示系数对块进行重构,依据重构误差构建似然函数以确定最佳候选样本,实现对目标的精确跟踪. 此外,在跟踪过程中对字典模板进行自适应更新^[13]以迎合目标外观的变化. 本文跟踪算法的框架结构如图1所示:

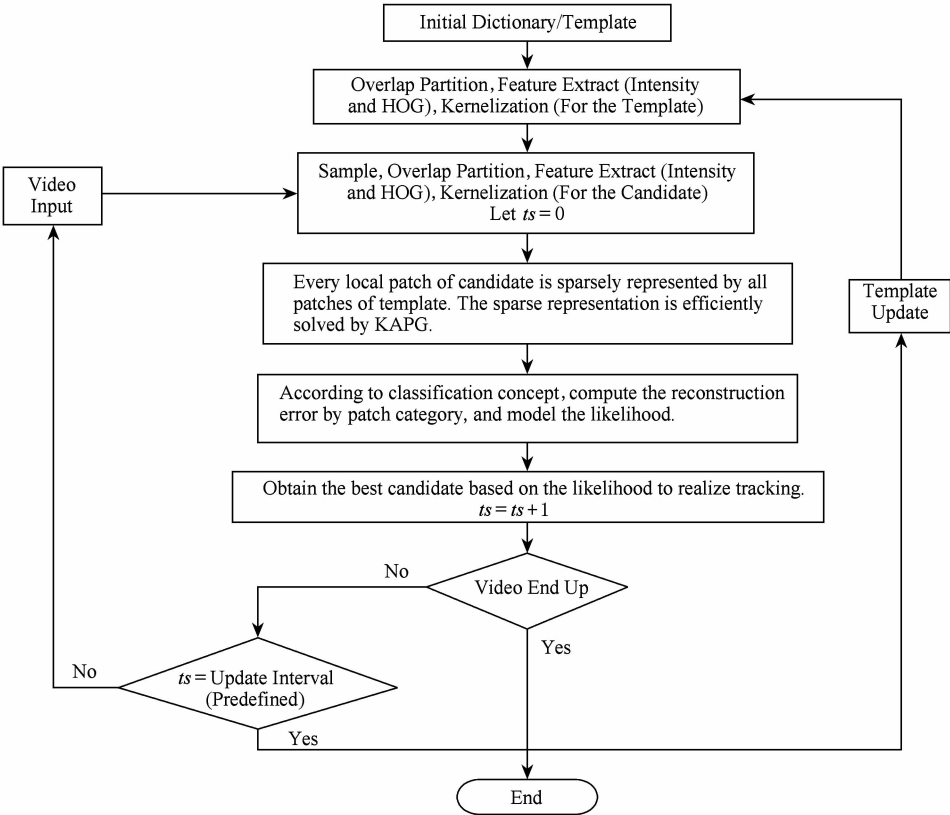


Fig. 1 Flow chart of the algorithm.
图1 本文算法结构框图

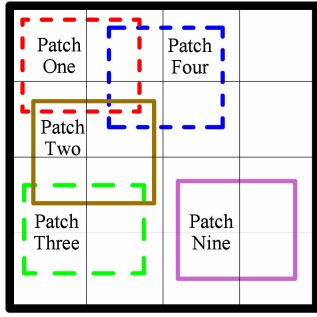
1 多特征联合稀疏表示

本文受多任务稀疏线性回归模型^[23-24]以及多任务联合稀疏表示成功用于视觉识别^[20]的启发,由此将多任务(特征)联合稀疏表示用于视觉跟踪中。

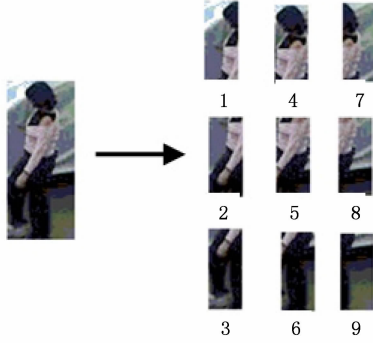
本节主要对字典目标模板和抽取的候选目标样本进行重叠分块处理,分别求取每个局部块的灰度特征和 HOG 特征,将不同位置块看成不同类别,用字典中所有块对候选目标块进行稀疏表示。

1.1 目标图像区域的重叠分块

已知目标模板图像集 $T = \{T_1, T_2, \dots, T_n\}$, 对每个目标模板 $T_i (i=1, \dots, n)$ 进行重叠分块, 事先给定块的大小, 重叠分块的步长, 分块按照从上到下、从左到右的方式进行, 将每个目标模板分成一些



(a) Diagram of overlap partition



(b) The local patches of target

Fig. 2 Overlap partition of target area.

图2 目标图像区域重叠分块

1.2 目标局部块的特征描述

利用分类的思想将不同位置的块看成不同类别, 则每个目标图像的块数就相当于类别数. 假设采用 K 个不同的特征, 对于每个特征索引 $k=1, \dots, K$, 模板字典重新表示为 $\mathbf{X}^k = (\mathbf{X}_1^k, \mathbf{X}_2^k, \dots, \mathbf{X}_J^k) \in \mathbb{R}^{m_k \times p}$ (注: 这里的表示与 1.1 节模板字典的表示略有不同, 以后的公式都基于此表示), 其中 J 为类别数 (局部块的数量), $\mathbf{X}_j^k \in \mathbb{R}^{m_k \times p_j}$ 表示第 j 类局部块的第 k 个特征, m_k 是第 k 个特征的维数, p_j 是第 j 类局部块的数量. 本算法各个类的局部块数量都是相同的, 是因为根据局部块的位置划分类别的缘故, 相同位置的局部块属于同一类, 从而每一类局部块的数量就等于模板图像的个数, 即 $p_j = n$. $\sum_{j=1}^J p_j = p = n \times J$ 是字典模板中所有局部块的数量. 本文选取 2 种不同的特征: 灰度特征和 HOG 特征, 从而 $K=2$, 但考虑算法的通用性, 文后的特征数仍沿用 K 表示。

局部图像块, 并用这些局部图像块构成模板字典 $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{(n \times r)}) \in \mathbb{R}^{m \times (n \times r)}$, 其中 m 是局部图像块的维数, n 是目标模板的数量, r 是单个目标模板所分成的局部块的数量. 由于每个局部块对应目标的一个固定区域, 因此所有的局部块一起可以描述目标的完整结构. 重叠分块示意图以及目标图像的分块结果如图 2 所示. 跟踪过程中抽取的目标候选样本也按照这种相同的方式进行重叠分块, 用 $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_r) \in \mathbb{R}^{m \times r}$ 表示. 用模板字典中所有块分别对候选目标的每个局部块进行线性表示, 于是每个候选局部块便得到一个相应的表示系数. 由于候选局部块仅与模板字典中与它位置相同的局部块具有较高的相似度, 相应的表示系数就大, 而其他位置的表示系数几乎为 0, 因而候选局部块的表示系数具有明显的稀疏特性。

1.3 抽取候选目标样本

基于粒子滤波框架, 跟踪过程中, 假设对于新到来的每帧图像, 需要抽取 N 个粒子 (样本). 抽取粒子的方式: 基于相邻帧之间目标移动很小的特点, 给定一帧新的图像, 在对应于上一帧跟踪目标的位置周围进行高斯采样 (服从 0 均值, 一个较小方差的分布). 选取目标的运动仿射变换参数作为状态变量 $\mathbf{S}_t = (x_t, y_t, \theta_t, s_t, \alpha_t, \phi_t)$, 其中 x_t, y_t 分别表示目标在第 t 帧 x, y 方向上的坐标, θ_t 表示目标在第 t 帧的旋转角度, s_t 表示目标在第 t 帧变化的尺度, α_t 表示目标在第 t 帧的高宽比, ϕ_t 表示目标在第 t 帧的倾斜角. 抽取得到的候选样本同样按照 1.1 节的方法进行重叠分块处理。

1.4 联合稀疏表示

候选样本 (抽样粒子) 采用与字典模板相同的特征, 且其每块用字典中所有块进行稀疏表示. 用 $\mathbf{y}_j^k = ((\mathbf{y}_j^k)^{(1)}, (\mathbf{y}_j^k)^{(2)}, \dots, (\mathbf{y}_j^k)^{(N)}) \in \mathbb{R}^{m \times N}$ 表示第 j 类

N 个候选样本的第 k 个特征,则它可被字典模板块线性表示如下:

$$\mathbf{y}_j^k = \mathbf{X}^k \mathbf{w}_j^k + \mathbf{e}_j^k, k = 1, \dots, K, \quad (1)$$

其中, $\mathbf{w}_j^k \in \mathbb{R}^{p \times N}$ (其中 $p = \sum_{j=1}^J p_j$) 是第 j 类 N 个候选样本在特征 k 下的表示系数, $\mathbf{e}_j^k \in \mathbb{R}^{m_k \times N}$ 是残差项. 用 $\mathbf{w}^k = (\mathbf{w}_1^k, \mathbf{w}_2^k, \dots, \mathbf{w}_J^k) \in \mathbb{R}^{p \times (N \times J)}$ 表示第 k 个特征的表示系数, $\mathbf{w}_j = ((\mathbf{w}_j^1})^\top, (\mathbf{w}_j^2})^\top, \dots, (\mathbf{w}_j^K})^\top)^\top \in \mathbb{R}^{(K \times p) \times N}$ 表示联合所有 k 个特征对应于第 j 类模板的表示系数. 因此,多特征联合稀疏表示可以公式化为以下 $l_{2,1}$ 混合范数正则化问题的解:

$$\min_{\mathbf{w}_j} \frac{1}{2} \sum_{k=1}^K \|\mathbf{y}_j^k - \mathbf{X}^k \mathbf{w}_j^k\|_2^2 + \lambda \|\mathbf{w}_j^\top\|_{2,1}, \quad (2)$$

其中, $\|\mathbf{w}_j^\top\|_{2,1} = \sum_{i=1}^N \|(\mathbf{w}_j^\top)_i\|_2$, $(\mathbf{w}_j^\top)_i$ 是矩阵 $\mathbf{w}_j$ 转置后的第 i 行.

式(2)联合粒子进行稀疏表示时,不仅体现了单个粒子局部块由字典模板表示的结构稀疏性,而且也考虑了各个粒子之间的相互依赖关系,这一点可由式(2)的约束项 $l_{2,1}$ 混合范数 $\|\mathbf{w}_j\|_{2,1}$ 得以解释. 由于粒子采样是在前一帧跟踪结果附近按照零均值、固定方差进行高斯采样,因而采样得到的一些粒子离真实值很近,与模板相似程度高,而另外一些粒子则属于异常值. 在联合所有粒子进行稀疏表示时, $l_{2,1}$ 混合范数 $\|\mathbf{w}_j\|_{2,1}$ 的 1 范数求和项 $\sum_{i=1}^N$ 约束了粒子之间的相互关系,使得不为 0 的系数集中到与模板相似的少数粒子上,从而体现了粒子与粒子之间的相似和差异关系. 对于上述模型的优化,可选择应用广泛的加速近邻梯度法 (APG)^[25-27] 对式(2)进行有效求解.

2 联合稀疏表示的核扩展

第 1 节内容主要是基于原始特征向量的稀疏表示. 然而,在许多视觉问题中,由于提取的各种特征维数不一致,为了统一维数方便后续处理,特征描述(向量)常被编码成相似矩阵或核矩阵的形式,为了联合多个核特征,我们将原始特征空间扩展到 RKHS (reproducing kernel Hilbert space) 来重新讨论稀疏表示问题^[20].

2.1 目标特征的核化

由于局部块的不同特征描述具有不同的维数,为了解决特征之间维数不一致的问题,需要对字典模板和候选目标进行核化处理(kernelization). 1. 2

节中提到,当考虑第 k 个特征时,对应于第 k 个特征的字典模板为 $\mathbf{X}^k = (\mathbf{X}_1^k, \mathbf{X}_2^k, \dots, \mathbf{X}_J^k) \in \mathbb{R}^{m_k \times p}$ (核化前),其中 J 为块的类别数, m_k 是第 k 个特征的维数, p 是字典模板中总的块数;对应于第 k 个特征的第 j 类候选样本表示为 $\mathbf{y}_j^k \in \mathbb{R}^{m_k}$ (核化前),则按照以下方式对字典 $\mathbf{X}^k$ 和候选样本 $\mathbf{y}_j^k$ 进行核化处理:

$$\mathbf{X}^k \leftarrow (\mathbf{X}^k)^\top \mathbf{X}^k, \mathbf{y}_j^k \leftarrow (\mathbf{X}^k)^\top \mathbf{y}_j^k,$$

显然核化后,对于任意特征 k ,都有 $\mathbf{X}^k \in \mathbb{R}^{p \times p}$, $\mathbf{y}_j^k \in \mathbb{R}^p$,从而达到了维数的一致.

2.2 再生核 Hilbert 空间(RKHS)的稀疏表示

我们认为在再生核 Hilbert 空间中线性表示模型是稀疏的. 直观上,核方法是对于每个特征 k 使用一个非线性函数 ϕ^k 将字典模板和候选样本从原始特征空间映射到另一个高维的再生核 Hilbert 空间(RKHS),在此高维空间中,对于一些给定的核函数 g^k ,有 $\phi^k(\mathbf{x}_i)^\top \phi^k(\mathbf{x}_j) = g^k(\mathbf{x}_i, \mathbf{x}_j)$. 在新的空间,我们将式(2)重新写成:

$$\min_{\mathbf{w}_j} \frac{1}{2} \sum_{k=1}^K \|\phi^k(\mathbf{y}_j^k) - \phi^k(\mathbf{X}^k) \mathbf{w}_j^k\|_2^2 + \lambda \|\mathbf{w}_j^\top\|_{2,1}, \quad (3)$$

其中, $\phi^k(\mathbf{X}^k) = [\phi^k(\mathbf{X}_1^k), \phi^k(\mathbf{X}_2^k), \dots, \phi^k(\mathbf{X}_J^k)]$.

设 $\mathbf{G}^k = \phi^k(\mathbf{X}^k)^\top \phi^k(\mathbf{X}^k)$ 是对应第 k 个特征的字典模板核矩阵, $\mathbf{h}_j^k = \phi^k(\mathbf{X}^k)^\top \phi^k(\mathbf{y}_j^k)$ 是对应第 k 个特征的第 j 个候选样本的核向量. 利用核矩阵的一种简单方式就是直接将 $\mathbf{h}_j^k$ 和 $\mathbf{G}^k$ 作为提取的新特征,则式(2)又可重新写为

$$\min_{\mathbf{w}_j} \frac{1}{2} \sum_{k=1}^K \|\mathbf{h}_j^k - \mathbf{G}^k \mathbf{w}_j^k\|_2^2 + \lambda \|\mathbf{w}_j^\top\|_{2,1}. \quad (4)$$

采用核化加速近邻梯度法(kernelizable accelerated proximal gradient, KAPG)求解上述稀疏问题式(4)得到相应的稀疏系数.

2.3 KAPG 方法求解式(4)

KAPG 算法是 APG 算法的核扩展. 一般地, APG 迭代对当前系数矩阵序列 $\{\mathbf{W}^t = [\mathbf{w}_j^{k,t}]\}_{t \geq 1}$ 和 aggregation 矩阵序列 $\{\mathbf{V}^t = [\mathbf{v}_j^{k,t}]\}_{t \geq 1}$ 交替进行更新. APG 算法每次迭代由 2 步构成: 1) 广义梯度映射阶段, 固定 $\mathbf{V}^t$, 更新 $\mathbf{W}^{t+1}$; 2) aggregation 阶段, 通过线性组合 $\mathbf{W}^{t+1}$ 和 $\mathbf{W}^t$, 更新 $\mathbf{V}^{t+1}$.

2.3.1 广义梯度映射阶段

已知当前矩阵 $\mathbf{V}^t$, 根据文献[23]中的结论更新 $\mathbf{W}^{t+1}$ 如下:

$$\begin{aligned} \mathbf{w}^{k,t+1} &= \mathbf{v}^{k,t} - \eta \nabla^k, k = 1, \dots, K, \\ \mathbf{w}_j^{t+1} &= \max\left(0, 1 - \frac{\lambda \eta}{\|\mathbf{w}_j^{t+1}\|_2}\right) \mathbf{w}_j^{t+1}, j = 1, \dots, J, \end{aligned} \quad (5)$$

其中, $\nabla^k = -\mathbf{h}^k + \mathbf{G}^k \mathbf{v}^{k,t}$, η 是步长参数.

2.3.2 Aggregation 阶段

在这一阶段,按照以下方式构建 $\mathbf{w}^{t+1}$ 和 $\mathbf{w}^t$ 的线性组合,更新 $\mathbf{v}^{t+1}$:

$$\mathbf{v}^{k,t+1} = \mathbf{w}^{k,t+1} + \frac{\alpha_{t+1}(1-\alpha_t)}{\alpha_t}(\mathbf{w}^{k,t+1} - \mathbf{w}^{k,t}), \quad (6)$$

按照惯例,一般将其中 α_t 设置为 $\frac{2}{t+3}$. 完整的 KAPG 算法参见算法 1.

算法 1. 利用 KAPG 算法求解多特征联合稀疏系数.

输入:核矩阵 $\{\mathbf{G}^k, \mathbf{h}^k\}_{k=1}^K$ 、正则化参数 λ 和步长 η ;

输出: $\mathbf{w}$.

① 初始化: $t \leftarrow 0, \mathbf{v}^{k,0} = 0, \mathbf{w}^{k,0} = 0, \alpha_0 = 1$;

② while 不收敛 do

③ $\mathbf{w}^{k,t+1} = \mathbf{v}^{k,t} - \eta(-\mathbf{h}^k + \mathbf{G}^k \mathbf{v}^{k,t}), k = 1, \dots, K$;

④ $\mathbf{w}_j^{t+1} = \max\left(0, 1 - \frac{\lambda\eta}{\|\mathbf{w}_j^{t+1}\|_2}\right) \mathbf{w}_j^{t+1}, j = 1, \dots, J$;

⑤ $\alpha_{t+1} = \frac{2}{t+3}$;

⑥ $\mathbf{v}^{k,t+1} = \mathbf{w}^{k,t+1} + \frac{\alpha_{t+1}(1-\alpha_t)}{\alpha_t}(\mathbf{w}^{k,t+1} - \mathbf{w}^{k,t}), k = 1, \dots, K$;

⑦ $t \leftarrow t+1$;

⑧ end while

2.4 局部块的重构误差

当用字典模板所有块对候选样本的某一类局部块(即某一位置块)进行线性表示时,显然,与候选样本局部块类别相同(即局部块位置相同)的字典模板局部块所对应的表示系数值较大.因而接下来求候选样本每块的重构误差时,我们仅利用字典中与候选样本块具有相同类别的局部块及其表示系数进行重构,而舍弃掉其他类别不同的局部块及其表示系数,这样处理的目的是使得重构误差尽可能小.因此,对于第 i 个候选样本(抽样粒子)的第 j 类块,重构误差(残差) $residual_j^{(i)}$ 计算如下:

$$\begin{aligned} residual_j^{(i)} &= \sum_{k=1}^K \theta^k \|\boldsymbol{\phi}^k((\mathbf{y}_j^k)^{(i)}) - \\ &\quad \boldsymbol{\phi}^k(\mathbf{X}_j^k)(\hat{\mathbf{w}}_j^k)^{(i)}\|_2^2 = \\ &\quad \sum_{k=1}^K \theta^k [(\mathbf{H}_j^k)^{(i)} - 2(\mathbf{h}_j^k)^{(i)}(\hat{\mathbf{w}}_j^k)^{(i)} + \\ &\quad ((\hat{\mathbf{w}}_j^k)^{(i)})^T \mathbf{G}_j^k (\hat{\mathbf{w}}_j^k)^{(i)}], i = 1, \dots, N, \end{aligned} \quad (7)$$

其中, $\mathbf{G}_j^k = \boldsymbol{\phi}^k(\mathbf{X}_j^k)^T \boldsymbol{\phi}^k(\mathbf{X}_j^k)$, $\mathbf{H}_j^k = \boldsymbol{\phi}^k(\mathbf{y}_j^k)^T \boldsymbol{\phi}^k(\mathbf{y}_j^k)$. $\hat{\mathbf{w}}_j^k$ 是第 j 类块在第 k 个特征下的最优表示系数中仅保留与字典中第 j 类块相应的系数(即用字典中所有

类块进行稀疏表示,但仅用字典中相同类的块进行重构). $\{\theta^k\}_{k=1}^K$ ($\sum_k \theta^k = 1$) 是核权值,用以度量不同特征在最后决策上的置信度大小,这里我们取等权值.

2.5 构建似然函数确定最佳粒子

似然函数的构建主要利用 2.4 节得到的重构误差.将候选样本每个局部块的重构误差 $residual_j^{(i)}$ 进行累加,即 $err^{(i)} = \sum_{j=1}^J residual_j^{(i)}$ 为第 i 个候选样本总的重构误差.接下来对所有候选样本的 $err^{(i)}$ 求最小值,最小 $err^{(i)}$ 对应的候选样本为最佳样本(最佳粒子) i^* , 即:

$$i^* = \arg \min_i \sum_{j=1}^J residual_j^{(i)}, \quad (8)$$

其中, $residual_j^{(i)}$ 的计算参见式(7).

3 模板更新

跟踪过程中如果采用固定不变的字典模板,则当目标外观或姿态发生变化、光照不均以及出现遮挡时往往会导致跟踪失败;然而如果字典模板更新太频繁,误差将会累积最终会导致跟踪器发生漂移,因而需要设计一种合理的模板更新方法.模板更新的主要思想是用当前好的跟踪结果去替代字典中已经过时的目标模板.本文采用 Jia 等人^[13]的模板更新方法,将子空间学习的方法引入稀疏表示以适应目标外观的变化,同时也减少了遮挡目标模板的影响.

3.1 更新原则

如果整个跟踪过程模板保持固定不变,则容易导致跟踪失败;倘若模板更新太过频繁,由于更新产生的误差累积又将会使得跟踪器发生漂移,因此本文采取每隔 5 帧更新 1 次模板的方式.在大多数跟踪方法中,早期的跟踪结果相对精确一些,因此它们应该在字典模板中存储更长的时间,平衡新旧目标模板的一种有效方式就是给它们分配不同的更新概率,一般来说,希望新模板更新快,旧模板更新慢.按照这样的原则首先产生一个累积概率序列如下:

$$L_p = \left\{0, \frac{1}{2^{n-1}-1}, \frac{3}{2^{n-1}-1}, \dots, \frac{2^{i-1}-1}{2^{n-1}-1}, \dots, 1\right\}, \quad i = 1, 2, \dots, n, \quad (9)$$

其中, n 表示字典中模板的个数,可以看出该累积序列的特点是前面密集后面稀疏.接着在 0 到 1 之间产生一个随机数,根据随机数的大小在上述累积序列中所处的位置来确定需要更新的模板.通过这样

的更新方式将会使得旧的模板更新较慢,而新的模板更新较快,从而可以减少模板漂移问题.

3.2 更新方法

通过联合稀疏表示和子空间学习来实现对字典模板的更新.我们先搜集当前的字典模板和最近几帧目标的跟踪结果,然后执行增量学习方法.目标的估计值可建模成 PCA 基向量和遮挡模板的线性组合如下:

$$\mathbf{p} = \mathbf{U}\mathbf{q} + \mathbf{e} = [\mathbf{U} \ \mathbf{I}] \begin{bmatrix} \mathbf{q} \\ \mathbf{e} \end{bmatrix}, \quad (10)$$

其中, $\mathbf{p}$ 表示当前时刻目标的最佳估计值, $\mathbf{U}$ 是由字典模板的特征基向量构成的矩阵,它由当前字典模板和最近几帧跟踪结果计算得到. $\mathbf{q}$ 是特征基向量的系数, $\mathbf{e}$ 表示 $\mathbf{p}$ 中被遮挡或干扰的像素.由于因遮挡或噪声引起的误差是随机的并且稀疏的,因此我们可以将此问题看成以下 l_1 正则化最小二乘问题:

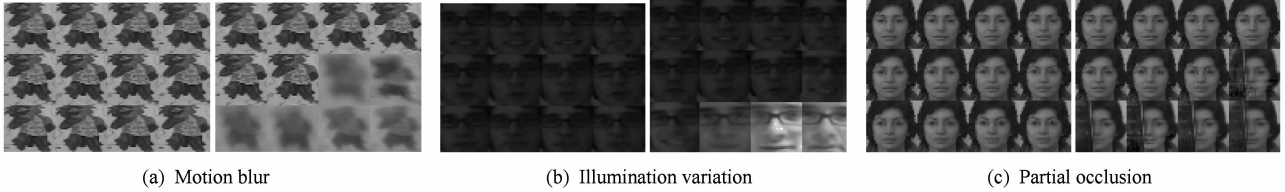


Fig. 3 Initial template and updated template for several sequences have different attributes.

图3 不同属性视频序列的初始模板和更新模板对比

4 跟踪算法

本文在粒子滤波框架下实现对感兴趣目标的跟踪.给定直到时刻 t 的观测集 $\mathbf{z}_{1:t} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_t\}$, 目标在时刻 t 的最佳状态可由最大后验概率确定:

$$\hat{\mathbf{S}}_t = \arg \max_{\mathbf{S}_t^{(i)}} p(\mathbf{S}_t^{(i)} | \mathbf{z}_{1:t}), \quad (12)$$

其中, $\mathbf{S}_t^{(i)}$ 表示时刻 t 第 i 个抽样粒子的状态.后验概率 $p(\mathbf{S}_t | \mathbf{z}_{1:t})$ 可由贝叶斯理论按照下面的方式递归得到:

$$p(\mathbf{S}_t | \mathbf{z}_{1:t}) \propto p(\mathbf{z}_t | \mathbf{S}_t) \int p(\mathbf{S}_t | \mathbf{S}_{t-1}) \times p(\mathbf{S}_{t-1} | \mathbf{z}_{1:t-1}) d\mathbf{S}_{t-1}, \quad (13)$$

其中, $p(\mathbf{S}_t | \mathbf{S}_{t-1})$ 是状态转移概率用来描述动态模型, $p(\mathbf{z}_t | \mathbf{S}_t)$ 是观测似然函数用来描述观测模型.

在粒子滤波框架中,后验概率 $p(\mathbf{S}_t | \mathbf{z}_{1:t})$ 可由抽样粒子集 $\{\mathbf{S}_t^{(i)}\}_{i=1}^N$ 的加权分布近似表示,每个粒子的重要性权值为 $l_t^{(i)}$.一般地,抽样粒子 $\mathbf{S}_t^{(i)}$ 从转移概率 $p(\mathbf{S}_t | \mathbf{S}_{t-1})$ 中进行抽取,从而重要性权值按如下方式更新:

$$\min_{\mathbf{c}} \|\mathbf{p} - \mathbf{H}\mathbf{c}\|_2^2 + \lambda \|\mathbf{c}\|_1, \quad (11)$$

其中, $\mathbf{H} = [\mathbf{U} \ \mathbf{I}]$, $\mathbf{c} = [\mathbf{q} \ \mathbf{e}]^T$, λ 是一个正则化参数.

通过求解式(11)得到 $\mathbf{q}$ 和 $\mathbf{e}$,将重构图像 $\hat{\mathbf{p}} = \mathbf{U}\mathbf{q}$ 用于替代被更新的模板.可以看出,遮挡模板的稀疏 $\mathbf{e}$ 用来解释存在的噪声和遮挡,可避免噪声被更新进入字典模板,而仅由 PCA 基向量构建的重构图像对遮挡的影响并不敏感.因此采用这种字典模板更新方法,可以适应目标外观改变并能很好地处理部分遮挡.

图3列出了不同属性视频序列的初始模板和更新模板,分别对应运动模糊、光照不均和部分遮挡这3种情况,其中每种情况的第1幅图片是初始字典,第2幅图片是更新后的字典(字典包含12个目标模板).可以看出,字典前面的旧模板几乎不更新,更新过程基本集中在字典后面的最近几个模板中.由此可见,该模板更新策略能很好地适应目标外观的变化.

$$l_t^{(i)} = l_{t-1}^{(i)} p(\mathbf{z}_t | \mathbf{S}_t^{(i)}). \quad (14)$$

因此,抽样粒子重要性权值的更新主要依赖于观测似然函数 $p(\mathbf{z}_t | \mathbf{S}_t)$,似然函数在跟踪中起到至关重要的作用.本文似然函数的构建参见2.5节:

$$p(\mathbf{z}_t | \mathbf{S}_t) \propto \min \left\{ \sum_{j=1}^J \text{residual}_j^{(i)} \right\}_{i=1}^N, \quad (15)$$

最终时刻 t 的最佳粒子 $\hat{\mathbf{S}}_t = \arg \max_{\mathbf{S}_t^{(i)}} p(\mathbf{S}_t^{(i)} | \mathbf{z}_{1:t}) = \mathbf{S}_t^{(i^*)}$,其中 i^* 由式(8)获得.多特征稀疏表示的跟踪算法参见算法2.

算法2. 基于多特征联合稀疏表示的跟踪算法.

输入:时刻 t 当前帧、字典模板 $\mathbf{T}_{t-1}$, N 个抽样粒子 $\{\mathbf{S}_{t-1}^{(i)}\}_{i=1, \dots, N}$;

输出:目标最佳状态 $\mathbf{S}_t^{(i^*)}$ 、更新的字典模板 $\mathbf{T}_t$.

① 对字典模板以及 N 个抽样粒子产生的候选样本进行重叠分块,提取每块的像素灰度值和 HOG 特征并进行核化处理;

② 用字典模板中所有块对候选样本中每一块进行多特征联合稀疏表示,利用 KAPG 算法求解式(4),得到粒子的稀疏表示系数 $\mathbf{w}$;

- ③ 将抽样粒子不同位置的块看成不同的类别,由式(7)求解各抽样粒子每块的重构误差 $residual_j^{(i)}$;
- ④ 基于步骤③求得的重构误差 $residual_j^{(i)}$,由式(8)构建似然函数从而确定最佳粒子序号 i^* ,最终得到目标最佳状态 S_i^* ;
- ⑤ 根据本文第3节的模板更新策略更新得到新模板 T_i .

5 实验结果

为了评估我们跟踪器的性能,我们分别在15个挑战性视频上进行跟踪实验. 这些视频序列覆盖了绝大多数挑战性场景,例如严重遮挡、运动模糊、光照不均、目标尺寸及外观变化、目标发生旋转以及复杂背景等. 所采用的挑战性视频序列分别为: Occlusion1, Occlusion2, Caviar1, Caviar2, Caviar3, DavidOutdoor, DavidIndoor, Singer1, Car4, Car11, Deer, Football, Jumping, Owl, Face. 其中视频序列 Caviar1, Caviar2, Caviar3 可以从 CAVIAR 数据集^[28]下载; Car4, Car11, DavidIndoor 可以从在线资源^[29]下载; Football, Singer1, Deer 可以从在线资源^[30]下载; 其他视频序列可以从 Yang Ming-Hsuan 的主页上下载^①.

整个实验中设正则化常数 $\lambda=0.001$. 对于每个视频序列,目标在第1帧的初始位置是手动进行提取的,首先采用一个简单跟踪器跟踪视频序列前12帧,将跟踪得到的目标作为初始字典的目标模板. 将目标模板的尺寸大小限定为 32×32 ,局部块的大小是 16×16 ,分块步长为8,按照从上到下,从左到右的顺序对目标图像进行分块,则每个目标模板最终可重叠分成9块. 字典模板每5帧更新一次.

我们将本文算法与现有的12种跟踪方法进行对比: incremental visual tracker (IVT)^[31], fragment-based tracker (FragT)^[32], Tracking-learning-detection (TLD)^[33], multiple instance learning tracker (MIL)^[34], visual tracking decomposition (VTD)^[35], APGL^[10], LSAT^[5], SCM^[14], OSPT^[15], adaptive structural local sparse appearance tracker (ASLSA)^[13], Least soft-threshold squares tracker (LSST)^[16], structured multi-task sparse tracker (S-MTT)^[18].

5.1 定量评估

我们采用2种度量准则来定量评估跟踪器的性能. 第1种是真实值与跟踪估计值中心位置(像素点)的相对误差,称为中心位置误差. 表1给出了各算法

Table 1 Average Center Error (in Pixels)
表 1 平均中心位置误差(像素点)

Image Sequence	FragT	IVT	MIL	VTD	TLD	APGL1	MTT	LSAT	SCM	ASLAS	OSPT	LSST	Ours
Occlusion1	5.6	9.2	32.3	11.1	17.6	6.8	14.1	<u>5.3</u>	3.2	10.8	<u>4.7</u>	<u>5.3</u>	6.6
Occlusion2	15.5	10.2	14.1	10.4	18.6	6.3	9.2	58.6	4.8	<u>3.7</u>	<u>4.0</u>	3.1	<u>3.7</u>
Caviar1	5.7	45.2	48.5	3.9	5.6	50.1	20.9	1.8	0.9	<u>1.4</u>	1.7	<u>1.4</u>	<u>1.1</u>
Caviar2	5.6	8.6	70.3	4.7	8.5	63.1	65.4	45.6	<u>2.5</u>	62.3	2.2	<u>2.3</u>	2.7
Caviar3	116.1	66.0	100.2	58.2	<u>44.4</u>	68.6	67.5	55.3	2.2	2.2	45.7	<u>3.1</u>	2.2
DavidOutdoor	90.5	53.0	38.4	61.9	173.0	233.4	65.5	101.7	64.1	87.5	5.8	<u>6.4</u>	<u>7.8</u>
DavidIndoor	148.7	<u>3.1</u>	34.3	49.4	13.4	10.8	13.4	6.3	3.4	3.5	<u>3.2</u>	4.3	2.6
Singer1	22.0	8.5	15.2	4.1	32.7	3.1	41.2	14.5	<u>3.7</u>	5.3	4.7	<u>3.5</u>	5.1
Car4	179.8	2.9	60.1	12.3	18.8	16.4	37.2	<u>3.3</u>	3.5	4.3	<u>3.0</u>	2.9	4.4
Car11	63.9	2.1	43.5	27.1	25.1	<u>1.7</u>	1.8	4.1	1.8	2.0	2.2	<u>1.6</u>	1.5
Deer	92.1	127.5	66.5	11.9	25.7	38.4	<u>9.2</u>	69.8	36.8	8.0	<u>8.5</u>	10.0	10.8
Football	16.7	18.2	16.0	4.1	11.8	12.4	<u>6.5</u>	14.1	10.4	18.0	33.7	7.6	<u>5.1</u>
Jumping	58.4	36.8	9.9	63.0	3.6	8.8	19.2	55.2	<u>3.9</u>	39.1	5.0	4.8	<u>4.6</u>
Owl	148.0	141.4	148.9	86.8	8.2	104.2	184.3	110.7	<u>7.3</u>	<u>7.6</u>	47.4	6.2	10.1
Face	48.8	69.7	134.7	141.4	22.3	148.9	127.2	<u>16.5</u>	125.1	95.1	24.1	12.3	<u>12.4</u>
Average	67.8	40.2	55.5	36.7	28.6	51.5	45.5	37.5	18.2	23.4	<u>13.1</u>	5.0	<u>5.3</u>

Note: The best three results are shown in bold, double underline and underline.

① <http://faculty.ucmerced.edu/mhyang/pubs.html>

跟踪不同视频序列的平均中心位置误差. 中心位置误差越小,跟踪越精确;另一种称为跟踪重叠(成功)率,依据 PASCAL VOC^[36]中的检测原则,可按照如下方式定义:给定跟踪估计值 R_T 和真实值 R_G ,跟

踪重叠(成功)率为 $score = \frac{area(R_T \cap R_G)}{area(R_T \cup R_G)}$. 表 2 给出了各算法针对不同视频序列的平均跟踪重叠(成功)率. $score$ 值越大跟踪越精确.

Table 2 Success Rate of Tracking Methods
表 2 平均重叠率

Image Sequence	FragT	IVT	MIL	VTD	TLD	APGL1	MTT	LSAT	SCM	ASLAS	OSPT	LSST	Ours
Occlusion1	0.90	0.85	0.59	0.77	0.65	0.87	0.79	<u>0.90</u>	0.93	0.83	<u>0.91</u>	0.89	0.88
Occlusion2	0.60	0.59	0.61	0.59	0.49	0.70	0.72	0.33	<u>0.82</u>	0.81	<u>0.84</u>	0.86	0.78
Caviar1	0.68	0.28	0.25	0.83	0.70	0.28	0.45	0.85	0.91	<u>0.90</u>	<u>0.89</u>	<u>0.89</u>	0.91
Caviar2	0.56	0.45	0.26	0.67	0.66	0.32	0.33	0.28	<u>0.81</u>	0.35	0.71	<u>0.80</u>	0.89
Caviar3	0.13	0.14	0.13	0.15	0.16	0.13	0.14	0.58	<u>0.87</u>	0.82	0.25	<u>0.85</u>	0.91
DavidOutdoor	0.39	0.52	0.41	0.42	0.16	0.05	0.42	0.36	0.46	0.45	<u>0.77</u>	<u>0.76</u>	0.78
DavidIndoor	0.09	0.69	0.23	0.23	0.50	0.63	0.53	0.72	0.75	<u>0.77</u>	<u>0.76</u>	0.75	0.88
Singer1	0.34	0.66	0.34	0.79	0.41	<u>0.83</u>	0.32	0.52	0.85	0.78	<u>0.82</u>	0.80	<u>0.82</u>
Car4	0.22	0.92	0.34	0.73	0.64	0.70	0.53	<u>0.91</u>	<u>0.89</u>	<u>0.89</u>	0.92	0.92	<u>0.91</u>
Car11	0.09	0.81	0.17	0.43	0.38	<u>0.83</u>	0.58	0.49	0.79	0.81	0.81	<u>0.84</u>	0.87
Deer	0.08	0.22	0.21	0.58	0.41	0.45	0.60	0.35	0.46	<u>0.62</u>	<u>0.61</u>	0.58	0.66
Football	0.57	0.55	0.55	0.81	0.56	0.68	<u>0.71</u>	0.63	0.69	0.57	0.62	0.69	<u>0.79</u>
Jumping	0.14	0.28	0.53	0.08	<u>0.69</u>	0.59	0.30	0.09	0.73	0.24	0.69	0.65	<u>0.72</u>
Owl	0.09	0.22	0.09	0.12	0.60	0.17	0.09	0.13	<u>0.79</u>	0.78	0.48	<u>0.81</u>	0.82
Face	0.39	0.44	0.15	0.24	0.62	0.14	0.26	<u>0.69</u>	0.36	0.21	0.68	<u>0.76</u>	0.77
Average	0.35	0.51	0.32	0.50	0.51	0.49	0.45	0.52	<u>0.74</u>	0.66	0.72	<u>0.79</u>	0.83

Note: The best three results are shown in bold, double underline and underline.

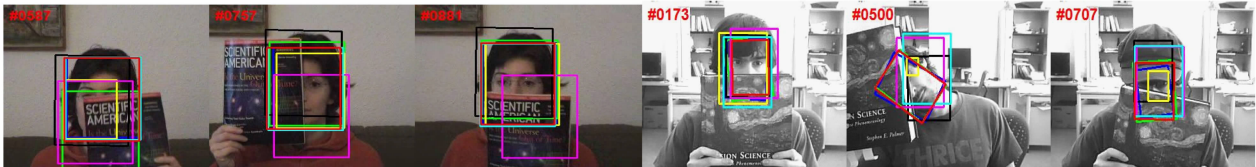
5.2 定性评估

1) 严重遮挡. 我们在严重遮挡 (Occlusion1, Occlusion2, Caviar3) 以及长时间部分遮挡 (Caviar1, Caviar2, DavidOutdoor, Woman) 的几个视频序列上,对算法进行测试. 图 4(a)(b)(c)(h)中 Woman 显示了在目标出现严重遮挡或长时间部分遮挡时,我们的算法依旧能很好地跟踪目标. 我们的跟踪器能够精确地跟踪目标:一方面是因为我们联合采用了目标的灰度特征和 HOG 特征,实现了特征间的优势互补;另一方面在计算粒子的稀疏系数时考虑了粒子之间的相互关系. 图 5 显示了采用单特征(单独采用图像灰度特征或 HOG 特征)和采用多特征联合(联合图像灰度特征和 HOG 特征)描述目标的跟踪结果对比. 目标框分别对应仅采用灰度特征、仅采用 HOG 特征以及联合灰度和 HOG 特征的跟踪结果. 不难看出,在没有出现遮挡时,三者都可以较好地跟踪目标,一旦出现长时间部分遮挡等挑战性情况,单特征跟踪器的跟踪精度将显著下降. 图 4 以

Occlusion1 视频为例,灰度特征跟踪器性能最差,最终丢失目标;HOG 特征跟踪器性能善可,但跟踪精度不高;联合灰度和 HOG 特征的跟踪器在整个跟踪过程中则能一直很好地锁定目标. 从另一个角度也可以反映出,对于 Occlusion1 这样的视频属性类,采用 HOG 特征比采用灰度特征能更好地描述目标.

在 Woman 序列中,与其他跟踪器出现漂移相比,我们的算法和 ASLAS 算法都能很好地处理长时间的部分遮挡,是因为采用了目标重叠分块的思想,这种处理对于部分遮挡稳健. 在 DavidOutdoor 序列中,其他跟踪器很早就丢失目标,困难在于目标发生了严重遮挡和姿态的显著变化,我们的跟踪器由于采取了多特征融合以及较好的字典模板更新方法因而可以有效地锁定目标.

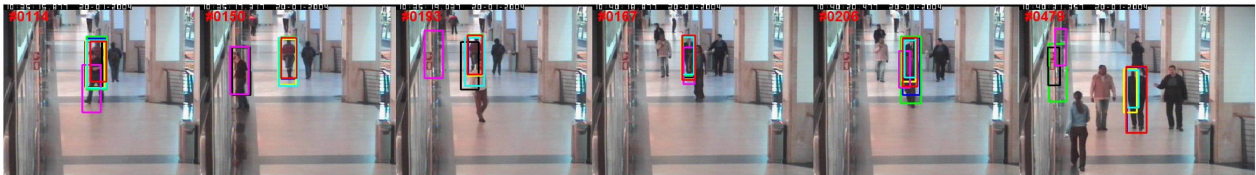
2) 光照变化. 图 4(d)(e)显示了光照明显变化 (DavidIndoor, Singer1, Car4) 以及姿态改变 (DavidIndoor),尺度大小变化 (Singer1, Car4) 的视频序列



(a) Occlusion1 and Occlusion2



(b) Caviar3 and DavidOutdoor



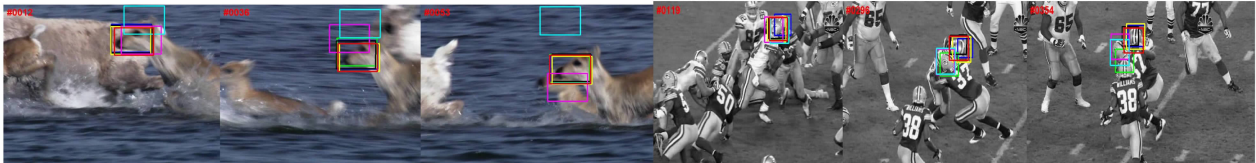
(c) Caviar1 and Caviar2



(d) DavidIndoor and Singer1



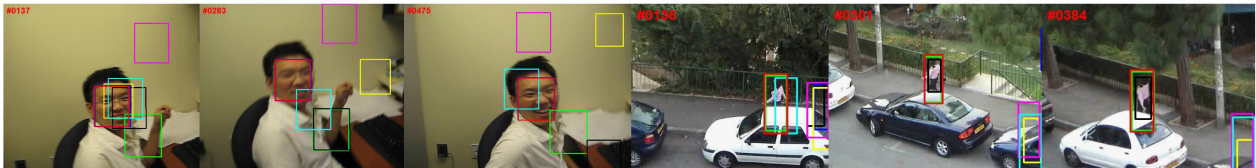
(e) Car4 and Car11



(f) Deer and Football



(g) Jumping and Owl



(h) Face and Woman

— LSST — ASLAS — MTT — VTD — MIL — Frag — Ours

Fig. 4 Tracking results of our algorithm on 16 challenging video sequences.

图 4 对 16 个挑战性视频序列的跟踪结果抽样

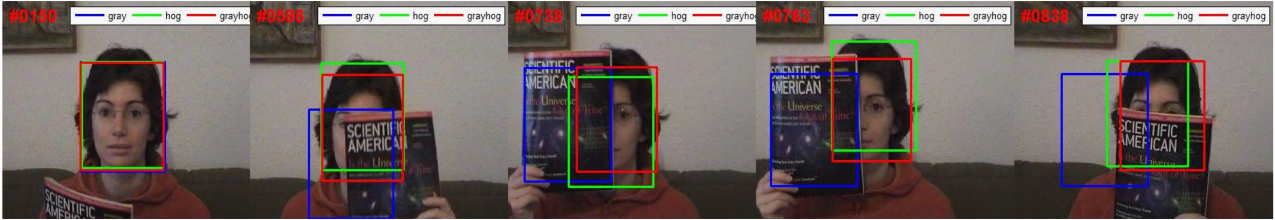


Fig. 5 Comparison of tracking results on single feature and multi-feature.

图 5 单个特征与多特征联合描述的跟踪结果对比

跟踪结果. 我们看到, MTT, MIL, Frag 跟踪器在这些情况跟踪效果不好, 我们的跟踪器和 LSST 跟踪器则能很好地跟踪.

3) 背景杂波: 图 4(e)(f) 显示了具有明显背景杂波(Car11, Deer, Football)的视频序列跟踪结果. 同时这些视频序列还具有其他一些挑战性因素, 例如 Car11 包含光照变化, Deer 中的快速运动, Football 还存在部分遮挡. 我们的跟踪器在视频序列 Car11 和 Football 上的中心位置误差和跟踪重叠率都达到很好的指标.

4) 运动模糊: 图 4(g)(h) 显示了快速运动以及运动模糊(Jumping, Owl, Face)的视频序列跟踪结果. 一般来说, 在出现运动模糊时很难预测目标的准确位置, ASLAS 跟踪器在这种情况下趋向于失败, 在 Jumping 序列中, MIL 尚能保持住目标, 但在 Owl 序列中就完全丢失目标. 我们的跟踪器和 LSST 跟踪器在这 4 个运动模糊视频序列中则显示了较好的跟踪性能.

6 结 论

本文提出了一种基于目标重叠分块多特征核稀疏表示的有效视觉跟踪方法. 利用目标的多种特征, 并采用基于核的特征描述, 通过 KAPG(kernelizable accelerated proximal gradient)方法求解联合特征的稀疏系数, 既弥补了单独采用图像灰度特征描述目标模型的不足, 又充分利用粒子之间的相互关系联合求解稀疏系数. 另外, 对目标区域的重叠分块方式能很好地处理视频序列中的部分遮挡, 以及联合利用基于增量空间的模板更新方法对于视频序列中的光照变化, 目标姿态改变则有稳健的跟踪效果.

近年来, 各种目标特征描述方法和目标外观模型相继被提出, 本文在目标特征描述上仅采用像素灰度值特征和 HOG 特征. 由于不同属性(严重遮

挡、光照不均、快速运动、姿态、尺度大小变化等)视频序列中的目标, 有其各自合适的特征描述及目标模型, 因而将来的工作计划包括: 研究目前常见的各种属性视频序列中目标的最佳特征描述, 针对视频序列的不同属性或同一个视频序列在不同时间段表现出的属性变化, 自适应地选取与之匹配的目标特征建立目标外观模型, 从而发展一种多特征(多目标外观模型)自适应的视觉跟踪方法.

参 考 文 献

[1] Zhang Fengjun, Zhao Ling, An Guocheng, et al. Mean shift tracking algorithm with scale adaptation [J]. Journal of Computer Research and Development, 2014, 51(1): 215-224 (in Chinese)
(张风军, 赵岭, 安国成, 等. 一种尺度自适应的 Mean Shift 跟踪算法[J]. 计算机研究与发展, 2014, 51(1): 215-224)

[2] Mei Xue, Ling Haibin. Robust visual tracking using L1 minimization [C] //Proc of Int Conf on Computer Vision. Los Alamitos, CA: IEEE Computer Society, 2009: 1436-1433

[3] Mei Xue, Ling Haibin. Robust visual tracking and vehicle classification via sparse representation [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2011, 33(11): 2259-2272

[4] Liu Baiyang, Yang Lin, Huang Junzhou, et al. Robust and fast collaborative tracking with two stage sparse optimization [C] //Proc of European Conf on Computer Vision. Berlin: Springer, 2010: 624-637

[5] Liu Baiyang, Huang Junzhou, Yang Lin, et al. Robust visual tracking with local sparse appearance model and k -selection [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2011: 1-8

[6] Tang Zhengyuan, Zhao Jiajia, Yang Jie, et al. Infrared target tracking algorithm based on sparse representation model [J]. Infrared and Laser Engineering, 2012, 41(5): 1389-1395 (in Chinese)

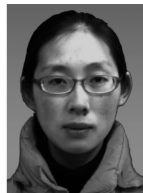
- (唐峥远, 赵佳佳, 杨杰, 等. 基于稀疏表示模型的红外目标跟踪算法[J]. 红外与激光工程, 2012, 41(5): 1389-1395)
- [7] Xu Ruyi, Chen Jingying. Lucas-Kanade tracking based on sparse representation [J]. Journal of Image and Graphics, 2013, 18(3): 283-289 (in Chinese)
(徐如意, 陈靓影. 稀疏表示的 Lucas-Kanade 目标跟踪[J]. 中国图象图形学报, 2013, 18(3): 283-289)
- [8] Mei Xue, Ling Haibin, Wu Yi, et al. Minimum error bounded efficient l_1 tracker with occlusion detection [C] // Proc of IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2011: 1257-1264
- [9] Li Hanxi, Shen Chunhua, Shi Qinfeng. Real-time visual tracking with compressing sensing [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2011: 1305-1312
- [10] Bao Chenglong, Wu Yi, Ling Haibin, et al. Real time robust l_1 tracker using accelerated proximal gradient approach [C] // Proc of IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2012: 1830-1837
- [11] Guo Yanwen, Chen Ye, Tang Feng, et al. Object tracking using learned feature manifolds [J]. Computer Vision and Image Understanding, 2014, 118: 128-139
- [12] Li Ang, Tang Feng, Guo Yanwen, et al. Discriminative nonorthogonal binary subspace tracking [G] //LNCS 6313; Proc of the 11th European Conf on Computer Vision. Berlin: Springer, 2010: 258-271
- [13] Jia Xu, Lu Huchuan, Yang Minghsuan. Visual tracking via adaptive structural local sparse appearance model [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2012: 1822-1829
- [14] Zhong Wei, Lu Huchuan, Yang Minghsuan. Robust object tracking via sparsity-based collaborative model [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2012: 1838-1845
- [15] Wang Dong, Lu Huchuan, Yang Minghsuan. Online object tracking with sparse prototypes [J]. IEEE Trans on Image Processing, 2013, 22(1): 314-325
- [16] Wang Dong, Lu Huchuan, Yang Minghsuan. Least soft-threshold squares tracking [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2013: 2371-2378
- [17] Zhang Tianzhu, Ghanem B, Liu Si, et al. Robust visual tracking via multi-task sparse learning [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2012: 2042-2049
- [18] Zhang Tianzhu, Ghanem B, Liu Si, et al. Robust visual tracking via structured multi-task sparse learning [J]. International Journal of Computer Vision, 2013, 101(2): 367-383
- [19] Zhang Tianzhu, Ghanem B, Liu Si, et al. Low-rank sparse learning for robust visual tracking [C] //Proc of European Conf on Computer Vision. Berlin: Springer, 2012: 470-484
- [20] Yuan Xiaotong, Yan Shuicheng. Visual classification with multi-task joint sparse representation [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2012: 3493-3500
- [21] Wu Yi, Lim J, Yang Minghsuan. Online object tracking: A benchmark [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2013: 1-8
- [22] Li Xi, Hu Weiming, Shen Chunhua, et al. A survey of appearance models in visual object tracking [OL]. [2013-10-01]. <http://arxiv.org/abs/1303.4803>
- [23] Obozinski G, Taskar B, Jordan M. Joint covariate selection and joint subspace selection for multiple classification problems [J]. Journal of Statistics and Computing, 2010, 20(2): 231-252
- [24] Zhang J. A probabilistic framework for multi-task learning [R]. Pittsburgh: Carnegie Mellon University, 2006
- [25] Nesterov Y. Gradient methods for minimizing composite objective function [R]. Belgium: CORE, Catholic University of Louvain, 2007
- [26] Tseng P. On accelerated proximal gradient methods for convex-concave optimization [OL]. [2013-10-01]. <http://www.csie.ntu.edu.tw/~b97058/tseng/papers/apgm.pdf>
- [27] Chen Xi, Pan Weike, Kwok J, et al. Accelerated gradient method for multi-task sparse learning problem [C] //Proc of Int IEEE International Conf on Data Mining. Los Alamitos, CA: IEEE Computer Society, 2009: 746-751
- [28] The University of Edinburgh. CAVIAR test case scenarios [OL]. [2013-10-01]. <http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1>
- [29] Ross D, Lim J, Lin R, et al. Incremental learning for robust visual tracking [OL]. [2013-10-01]. <http://www.cs.toronto.edu/~dross/ivt>
- [30] Kwon J, Lee K M. Visual tracking decomposition [OL]. [2013-10-01]. <http://cv.snu.ac.kr/research/~vtd>
- [31] Ross D, Lim J, Lin R, et al. Incremental learning for robust visual tracking [J]. International Journal of Computer Vision, 2008, 77(1/2/3):125-141
- [32] Adam A, Rivlin E, Shimshoni I. Robust fragments-based tracking using the integral histogram [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2006: 798-805

[33] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection [J]. IEEE Trans on Pattern Analysis Machine Intelligence, 2012, 34(7): 1409-1422

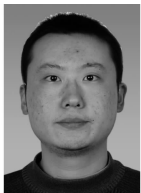
[34] Babenko B, Yang Minghsuan, Belongie S. Visual tracking with online multiple instance learning [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2009: 983-990

[35] Kwon J, Lee K M. Visual tracking decomposition [C] //Proc of IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2010: 1269-1276

[36] Everingham M, Van Gool L, Williams C, et al. The pascal visual object classes (voc) challenge [J]. International Journal of Computer Vision, 2010, 88(2): 303-338



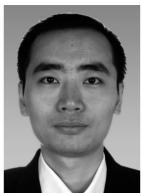
Hu Zhao Hua, born in 1981. PhD and associate professor at Nanjing University of Information Science & Technology. Her main research interests include computer vision, object tracking and pattern recognition.



Yuan Xiaotong, born in 1980. PhD and professor at Nanjing University of Information Science & Technology. His main research interests include machine learning, data mining, and computer vision(xtyuan1980@gmail.com).



Li Jun, born in 1983. PhD candidate at the School of Computer Science and Technology, Nanjing University of Science and Technology, China. His current research interests include deep learning, sparse representation and stability analysis of recurrent neural networks(junl.njust@gmail.com).



He Jun, born in 1978. PhD and lecturer at Nanjing University of Information Science and Technology. His main research interests include algorithms, analysis, and applications of optimization tools to image/signal processing and computer vision (hejun.zz@gmail.com).

《计算机研究与发展》征订启事

《计算机研究与发展》(Journal of Computer Research and Development)是中国科学院计算技术研究所和中国计算机学会联合主办、科学出版社出版的学术性刊物,中国计算机学会会刊. 主要刊登计算机科学技术领域高水平的学术论文、最新科研成果和重大应用成果. 读者对象为从事计算机研究与开发的研究人员、工程技术人员、各大专院校计算机相关专业的师生以及高新企业研发人员等.

《计算机研究与发展》于 1958 年创刊,是我国第一个计算机刊物,现已成为我国计算机领域权威性的学术期刊之一. 并历次被评为我国计算机类核心期刊,多次被评为“中国百种杰出学术期刊”. 此外,还被《中国学术期刊文摘》、《中国科学引文索引》、“中国科学引文数据库”、“中国科技论文统计源数据库”、美国工程索引(Ei)检索系统、日本《科学技术文献速报》、俄罗斯《文摘杂志》、英国《科学文摘》(SA)等国内外重要检索机构收录.

国内邮发代号:2-654;国外发行代号:M603
国内统一连续出版物号:CN11-1777/TP
国际标准连续出版物号:ISSN1000-1239

联系方式:
100190 北京中关村科学院南路 6 号《计算机研究与发展》编辑部
电话: +86(10)62620696(兼传真);+86(10)62600350
Email:crad@ict.ac.cn
http://crad.ict.ac.cn