

基于最大池图匹配的 形变目标跟踪方法

王治丹, 蒋建国, 齐美彬

(合肥工业大学计算机与信息学院, 安徽合肥 230009)

摘要: 该文提出了一种基于最大池图匹配的形变目标跟踪算法, 适用于跟踪目标产生较大形变或者严重遮挡等场合. 此方法首先将目标搜索区域过分割为候选目标部件并建立动态图表示, 即目标部件的表象特征和它们之间的几何位置关系. 然后采用最大池图匹配算法, 得到目标图和候选图中部件的匹配关系, 从而确定出目标位置的置信图. 联合考虑目标整体和目标部件对目标位置的支持, 投票决定出精确的目标位置. 在各种形变目标的跟踪序列测试下, 该算法与其他跟踪器的对比验证了其有效性和鲁棒性.

关键词: 视觉目标跟踪; 动态图表示; 最大池图匹配

中图分类号: TP751 **文献标识码:** A **文章编号:** 0372-2112 (2017)03-0704-08

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2017.03.030

Deformable Object Tracking Based on Max-pooling Graph Matching

WANG Zhi-dan, JIANG Jian-guo, QI Mei-bin

(School of Computer and Information, Hefei University of Technology, Hefei, Anhui 230009, China)

Abstract: This paper develops a novel deformable object tracking algorithm based on max-pooling graph matching, which can be applied in the scenes with large deformations and severe occlusions. The dynamic graph is built based on candidate parts extracted by over-segmentation method from searching area, namely feature representation of candidate parts and geometric structure between them. Based on max-pooling graph matching method, the matching relations between target parts and candidate parts are found to calculate the confidence map of target location. Considering both the support of holistic target and local parts, the optimal target location can be determined. Compared to state-of-the-art methods, experimental results on several deformable sequences demonstrate the effectiveness and robustness of the proposed method.

Key words: visual tracking; dynamic graph representation; max-pooling graph matching

1 引言

视觉跟踪作为计算机视觉领域的基础问题, 在各方面都取得了重要应用, 比如视频监控、轨迹分析以及人机交互等. 尽管之前研究者设计出各种算法并取得重要进展, 但是由于目标在跟踪过程中出现剧烈形变、严重遮挡、不规则运动以及尺度变化等情况, 对形变目标的跟踪仍然是一项棘手的工作.

之前方法大多基于目标整体模型建立复杂的特征表达来描述目标的表象变化, 比如稀疏表示^[1-3], 子空

间学习^[4]以及 SVM 模型^[5]等. 由于存在目标形变和遮挡, 很多固定尺度的跟踪器在跟踪过程中会引入背景噪声或者丢失前景信息, 导致跟踪失败. 为增强其适应度, 一些基于部件的跟踪器^[6-8]不断被提出. 文献[6]基于 SLIC 过分割算法^[17]将目标分割成超像素目标部件, 建立表象模型来区分前景和背景区域. 但是由于较少考虑目标部件之间的结构信息, 当前景和背景表象相似时容易混淆目标. 在此基础上, 文献[7]综合考虑结构信息, 基于频谱分析算法将跟踪问题转化为图匹配问题, 频谱分析算法的精度决定了跟踪效果. 相似地, 文

献[8]利用目标结构信息,训练了一个在线的结构化隐式 SVM 模型来预测目标中各部件的位置.但是由于假设目标尺度不变,当目标发生形变的时候会出现跟踪困难.

考虑到跟踪目标可能会出现巨大形变和面临严重遮挡,本文提出了基于部件的最大池图匹配的跟踪方法(Max-pooling Graph matching based Tracker, MGT).首先对于目标搜索区域,算法基于图像分割技术提取出候选目标部件建立候选图,并与建立好的目标图模型进行匹配.其次采用更为精确的最大池(max-pooling)图匹配方法,建立起目标图和候选图之间的部件匹配关系.在此基础上得到目标位置的置信图(confidence map),通过采样确定目标的最优位置.最后,为了避免仅考虑局部目标部件的贡献造成的鉴别力不够,引入了整体目标的特征表达参与目标位置投票,以提高跟踪鲁棒性.本文的主要贡献:提出了基于最大池图匹配方法的形变目标跟踪算法,同时考虑整体目标和目标部件对目标位置的支持,并基于目标位置置信图采样策略获得最优的目标位置.

2 本文提出的方法

在目标出现较大形变或者严重遮挡的时候,目标形状会发生剧烈变化.如果使用整体框(bounding box)来描述目标,势必会滤掉部分前景或者引入背景噪声,无法给出精确的目标表达.文献[9,10]认为目标在运动过程中其局部部件的结构不变性得以保持,并取得良好效果.根据此假设,本文基于动态图结构表示目标部件,即目标部件的表象特征(表象信息),以及他们之间的相对位置关系(结构信息).然后采用最大池图匹配方法找到目标图和当前搜索区域内候选图之间各部件的匹配关系,从而帮助定位出目标位置.

2.1 建立动态图结构

动态图结构模型主要是用以描述目标部件的表象特征和他们之间的空间位置关系.我们定义动态图 $G = (V, E)$,其中包括:(1)图节点(graph node) $v \in V$,代表了从目标搜索区域中提取的超像素候选部件;(2)图边(graph edge) $e \in E$,表示部件之间的近邻关系.动态图定义如下:

$$\begin{cases} V = \{v_i\}_{i=1}^n \\ E = \{e | \forall v_i, v_j \in e, d(v_i, v_j) \leq \varepsilon\} \end{cases} \quad (1)$$

其中 v_i 为图节点, n 为图节点数目. e 为图边, $d(v_i, v_j)$ 表示图节点之间的距离.定义具有近邻关系的相邻节点之间的距离小于距离 $\varepsilon = 1.5 \sqrt{W \cdot H/N}$. $\{H, W\}$ 为目标搜索区域的长宽, N 为超像素数目.

2.2 提取动态图节点

如果采用目标搜索区域中所有超像素候选部件作为图节点,不仅会提高计算复杂度,也会引入不必要的背景噪声.所以首先采用图像分割方法得到粗略的前景区域,尽可能地滤掉背景噪声,再收集相对可信的候选目标部件作为图节点.根据图割(Graph Cut)算法^[11],建立基于马尔科夫随机场(Markov Random Field, MRF)的前景分割的能量方程:

$$E(f|M) = \sum_{p \in P} D(f_p|M) + \sum_{p,q \in N} V_{p,q}(f_p, f_q) \quad (2)$$

其中 $D(f_p|M)$ 和 $V_{p,q}(f_p, f_q)$ 分别为分割模型的一元能量项(unary energy)和二元能量项(binary energy). f 为候选目标部件集合 P 在当前目标搜索区域的标签, $f_p \in \{0, 1\}$ 为指示函数,当候选部件 p 属于前景时 $f_p = 1$ 否则 $f_p = 0$. 图像分割的表象模型记为 M , 包含归一化后的 HSV 颜色直方图 M^{hsv} 和线性 SVM 分类器 M^{svm} .

2.2.1 一元能量项

能量方程的一元能量项表示候选目标部件属于前景的似然性,由目标部件的 HSV 颜色直方图模型和 SVM 鉴别式模型生成,计算公式如下所示:

$$D_p(f_p|M) = \lambda_1 D_p^{\text{hsv}}(f_p|M^{\text{hsv}}) + \lambda_2 D_p^{\text{svm}}(f_p|M^{\text{svm}}) \quad (3)$$

其中 λ_1 和 λ_2 为平衡系数. D_p^{hsv} 为部件 p 中像素属于模型 $M^{\text{hsv}} = \{M_b^{\text{hsv}}, M_f^{\text{hsv}}\}$ 产生式能量(generative energy),计算公式如下:

$$D_p^{\text{hsv}}(f_p|M^{\text{hsv}}) = \begin{cases} -\frac{1}{n_p} \sum_{i=1}^{n_p} H_f(c_i), & f_p = 1 \\ -\frac{1}{n_p} \sum_{i=1}^{n_p} H_b(c_i), & f_p = 0 \end{cases} \quad (4)$$

其中 M_f^{hsv} 和 M_b^{hsv} 分别为前景和背景 HSV 颜色直方图,其中 c_i 为像素 i 在 HSV 颜色直方图中对应值, n_p 为部件 p 中包含的像素个数. $H_f(c_i)$ 和 $H_b(c_i)$ 分别为 c_i 在前景模型和背景模型中的似然估计值(likelihood). D_p^{svm} 为通过模型 M^{svm} 计算得到的鉴别式能量(discriminative energy),即 SVM 分类器鉴别候选部件 p 的前景分数,如下式所示.平衡参数 λ_3 可以控制引入的候选部件的数量.

$$D_p^{\text{svm}}(f_p|M^{\text{svm}}) = \begin{cases} -M^{\text{svm}}(p), & M^{\text{svm}}(f_p) \geq 0, f_p = 1 \\ M^{\text{svm}}(p), & M^{\text{svm}}(f_p) \geq 0, f_p = 0 \\ -\lambda_3 \cdot M^{\text{svm}}(p), & M^{\text{svm}}(f_p) < 0, f_p = 1 \\ \lambda_3 \cdot M^{\text{svm}}(p), & M^{\text{svm}}(f_p) < 0, f_p = 0 \end{cases} \quad (5)$$

2.2.2 二元能量项

二元能量项表示相邻部件之间的相关性或平滑度,即具有相似表象的相邻部件分配相同标签时具有较小的能量.计算如下所示:

$$V_{p,q}(f_p, f_q) = I(f_p \neq f_q) \cdot \exp(-d_x(z_p, z_q)/\delta_c^2) \quad (6)$$

其中当标签 $f_p \neq f_q$ 时判别式 $I(\cdot)$ 返回 1, 否则返回 0. $d_x(z_p, z_q)$ 计算相邻候选部件 p 和 q 的 HSV 直方图表象特征之间的卡方距离, δ_c 为平滑系数.

根据图割算法^[11], 最小化能量方程(2)中的能量解, 实现图像的前景分割. 搜集满足条件的标签 f_p , 将前景候选部件(超像素)加入到图节点集合中, 记为 $V = \{v_i\}_{i=1}^n = \{p | f_p = 1\}$.

2.3 图匹配策略

2.3.1 图匹配最优化建模

本文通过计算两图之间的匹配关系实现目标定位. 首先我们在给定的第一帧建立目标图 $G = (V, E)$, 然后在当前帧目标搜索区域建立候选图 $G' = (V', E')$. 图匹配问题的解为两图可能匹配对的子集 $V_c \subset V \times V'$, 记为二值化赋值矩阵 $\mathbf{X} \in \{0, 1\}^{n \times n'}$, 其中 n 和 n' 分别表示图 G 和图 G' 中的节点数目. 如果节点 $v_i \in V$ 匹配上节点 $v'_a \in V'$, 则 $X_{ia} = 1$, 否则 $X_{ia} = 0$. 将矩阵 \mathbf{X} 各列向量串联, 得到二值化赋值向量 $\mathbf{x} \in \{0, 1\}^{nm}$. 我们引入目标函数 $f(\mathbf{x})$ 度量了目标图和候选图的相似性, 定义如下:

$$f(\mathbf{x}) = \sum_{x_{ij}=1} s_E(e_{ij}, e'_{ab}) + \sum_{x_{ia}=1} s_V(v_i, v'_a) = \mathbf{x}^T \mathbf{A} \mathbf{x} \quad (7)$$

其中一元相似性函数 $s_V(v_i, v'_a)$ 表示两图中节点的相似性, 二元相似性函数 $s_E(e_{ij}, e'_{ab})$ 表示两图中边的相似性. 这个相似性函数通常用一个匹配类属矩阵 \mathbf{A} 来表示, 其非对角元素为 $A_{ia;ib} = s_E(e_{ij}, e'_{ab})$, 对角元素为 $A_{ia;ia} = s_V(v_i, v'_a)$. 通过最大化目标函数 $f(\mathbf{x})$, 求得最优匹配解 \mathbf{x}^* . 建立匹配模型^[12] 如下所示:

$$\mathbf{x}^* = \arg \max_{\mathbf{x}} \mathbf{x}^T \mathbf{A} \mathbf{x} \quad (8a)$$

$$\text{s. t. } \sum_{i=1}^n x_{ia} \leq 1, \quad \sum_{a=1}^{n'} x_{ia} \leq 1, \quad (8b)$$

其中匹配类属矩阵 \mathbf{A} 的非对角元素定义为 $A_{ia;ib} = \exp(-\|(l_i - l_j) - (l_a - l_b)\|_2 / \delta_g^2)$, 表示匹配对的一致性, 即相邻匹配对之间的结构保持程度. l 表示节点的中心位置, δ_g 为平滑系数. 而对角元素 $A_{ia;ia} =$

$\exp(-d_x(z_i, z_a) / \delta_c^2)$, 表示匹配对的可靠性, 即该匹配对中节点表象的相似性. $d_x(z_i, z_a)$ 计算节点 i 和节点 a 的 HSV 直方图表象特征的卡方距离.

2.3.2 图匹配策略

文献[8]同样是基于图匹配方法的跟踪器, 相比本文算法而言最重要的区别是图匹配策略. 图 1 给出了两种跟踪器的图匹配策略, 即如何找到当前匹配对 (match pair) $i \sim a$ 的支持项 $\{j \rightarrow i, b \rightarrow a\}$ 来描述该匹配的可靠性. 其中图左边表示目标图节点, 图右边表示候选图节点.

图 1(a) 展示的是文献[7]使用的图匹配策略, 即考虑目标图节点 j 在候选图中所有 n 个近邻节点形成的支持项 $\{b_k \rightarrow a\}_{k=1}^n$ 的对 $i \sim a$ 的贡献, 这种方式被称作求和池 (sum pooling) 或者平均池 (average pooling). 但是其精确度常常受到大量无关候选匹配对的影响, 使得匹配最优化问题陷入局部极小值. 例如很多与目标图中支持项 $j \rightarrow i$ 方向并不一致的候选图支持项会参与计算.

图 1(b) 表示的是本文采用的图匹配策略, 仅仅选择了与 $j \rightarrow i$ 最相似的 $b_{\max} \rightarrow a$ 作为支持项描述匹配对 $i \sim a$ 的可靠性, 称之为最大池支持项. 图 1(c) 表示最大池匹配方式, 即目标图匹配对中的每一个节点支持项都只使用候选图中的最大池支持项, 并将其相关结构一致性分数作为匹配似然度 (见 2.4 节).

相比于文献[8], 本文采用的最大池匹配策略的好处是: (1) 通常候选图中大部分支持项与相应的目标图支持项不相符, 最大池匹配策略可以忽略掉大量无关项, 提高匹配可靠性分数的精确度; (2) 最大池匹配策略通常只给予正确的匹配对较高的权重. 由于与目标支持项相似度高的候选图支持项很难为噪声, 只要在较大近邻区域内选择相似度最高的支持项并不会过滤掉正确的支持项, 反而可以允许更大的结构变化, 对发生剧烈形变的目标跟踪有利.

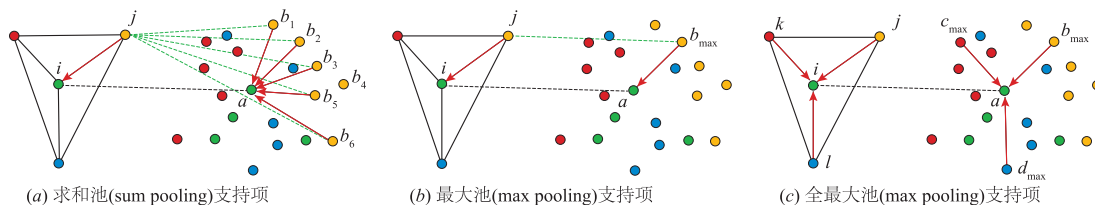


图1 图匹配支持方式

2.3.3 最大池图匹配

由于图匹配模型式(8)为二次整数规划 NP-hard 问题, 我们将可行解空间放松到连续空间, 即 $\mathbf{x} \in [0, 1]^{nm}$, 对目标函数 $f(\mathbf{x})$ 进行求解. $f(\mathbf{x})$ 在当前迭代次数

k 的解 \mathbf{x}_k 的一阶泰勒展开式为:

$$f(\mathbf{x}) \approx f(\mathbf{x}_k) + (\mathbf{x} - \mathbf{x}_k)^T \mathbf{A} \mathbf{x}_k \quad (9)$$

基于式(7)的定义, 在 2.3.2 节分析基础上, 定义最大池乘积 $(\mathbf{A} \otimes \mathbf{x})_{ia}$ 如下所示^[12], 用以取代式(8)中的 $\mathbf{A} \mathbf{x}$ 项:

$$(\mathbf{A} \otimes \mathbf{x})_{ia} = \mathbf{x}_{ia} \mathbf{A}_{ia;ia} + \sum_{j \in N_i} \max_{b \in N_a} \mathbf{x}_{jb} \mathbf{A}_{ia;jb} \quad (10)$$

其中第一项表示匹配对 $i \sim a$ 的可靠性,第二项表示匹配对 $i \sim a$ 近邻中最大池支持项对该匹配对的贡献,如图 1(b) 所示. 根据式(9)的梯度下降算法,目标函数 $f(\mathbf{x})$ 可通过以下迭代步骤求解:首先通过 $\mathbf{x}_{ia} \leftarrow \mathbf{x}_{ia} \mathbf{A}_{ia;ia} + \sum_{j \in N_i} \max_{b \in N_a} \mathbf{x}_{jb} \mathbf{A}_{ia;jb}$ 对每一候选匹配对 $i \sim a$ 更新;然后归一化解 $\mathbf{x} \leftarrow \mathbf{x} / \|\mathbf{x}\|_2$,直到 \mathbf{x} 收敛为止. 根据文献[12],迭代会在 10 到 50 次之间收敛. 算法的复杂度为 $O(nmk^2)$,其中 n 和 m 分别表示目标图和候选图中节点数量, k 表示用以计算匹配度的支持项个数. 通过离散化最优解 \mathbf{x} 得到匹配对向量 \mathbf{y} ,从而得到目标图和候选图节点之间的匹配关系 $C = \{i \sim a \mid y_{ia} = 1\}$.

2.4 计算目标位置

得到式(2)生成的前景候选部件和匹配关系 C 生成的匹配候选部件,我们可以得到目标位置的置信图. 在此基础上,最优的目标位置应该包含尽量多的前景像素. 其中目标位置为中心位置 l_c^* 和尺度 $s^* = (w, h)$, w 和 h 分别表示目标的宽和高. 借鉴文献[8]中目标位置确定方法,分两步进行. 第一步计算目标初始中心位置 $l_c = \sum \varpi_{ia} \cdot l_c^{ia} / \sum \varpi_{ia}$. 其中 l_c^{ia} 为匹配对 $i \sim a$ 的相对位置,而 ϖ_{ia} 表示其匹配可靠性分数,计算公式如下所示:

$$\varpi_{ia} = \mathbf{A}_{ia;ia} \left(\frac{1}{|N_i|} \sum_{j \in C \mid j \in N_i} \mathbf{A}_{ia;jb} \pi_{jb} + \frac{1}{|N_b|} \sum_{j \in C \mid b \in N_a} \mathbf{A}_{ia;jb} \pi_{jb} \right) \quad (11)$$

其中 $j \in N_i$ 和 $b \in N_a$ 分别表示节点 j 为节点 i 的近邻,节点 b 为节点 a 的近邻. 上式中 π_{jb} 的计算式为:

$$\pi_{jb} = \mathbf{A}_{jb;jb} \left(\frac{1}{|N_j| - 1} \sum_{kk' \in C \mid k \in N_j \setminus i} \mathbf{A}_{jb;kk'} + \frac{1}{|N_b| - 1} \sum_{kk' \in C \mid k' \in N_j \setminus a} \mathbf{A}_{jb;kk'} \right) \quad (12)$$

其中 $N_j \setminus i$ 为排除节点 i 之后的节点 j 的近邻节点集合. 这样初始中心位置 l_c 可以通过加权投票各近邻支持项对中心位置的贡献算出.

基于上一帧目标尺度 s ,以及得到初始目标中心位置 l_c 之后,第二步需要使用震荡项 $\{\delta_l, \delta_s\}$ 将目标中心和目标尺度调整到视觉中更好的位置,使得预测的跟踪目标框可以包含最多的前景像素. 最优目标位置 $\{l_c^*, s^*\}$ 的分数计算如下:

$$\{l_c^*, s^*\} = \xi(\delta_l, \delta_s) \cdot \arg \max_{\delta_l, \delta_s} \{ \omega \cdot \theta'(\delta_l, \delta_s) + \theta^c(\delta_l, \delta_s) - \theta^n(\delta_l, \delta_s) \} \quad (13)$$

其中 $\theta'(\delta_l, \delta_s)$ 和 $\theta^c(\delta_l, \delta_s)$ 分别代表目标中心 $l_c + \delta_l$ 和尺度为 $s + \delta_s$ 确定的目标框里面,匹配候选部件和非匹配候选部件的像素个数; $\theta^n(\delta_l, \delta_s)$ 表示背景部件中的

像素个数. 震荡项 $\{\delta_l, \delta_s\}$ 设置为 2.1 节定义的近邻距离 $\varepsilon = 1.5 \sqrt{W \cdot H/N}$. 系数 ω 表明了匹配候选部件的可信度. 为了增强目标位置定位的鲁棒性,我们不仅考虑匹配候选部件对目标位置的贡献,同时引入相邻两帧目标特征之间的相似性函数 $\xi(\delta_l, \delta_s) = \exp(-m \cdot d_\chi(z_t, z_{t-1}))$, 分配给当前帧目标框权重值,使得最小的目标框包含最多的目标部件. 其中 $m = 10$ 为量化系数, $d_\chi(z_t, z_{t-1})$ 表示前后两帧目标区域的表象特征的卡方距离,用于控制表象特征,由 HSV 直方图特征和 LBPHF 纹理特征^[18] 串联得到. 最后通过粒子采样随机产生一些目标位置,并选择取得最大分数的位置作为预测的跟踪目标位置.

2.5 模型更新

由于目标表象在跟踪过程中可能发生剧烈变化,为了生成较精确的候选目标部件,需要每隔 η 帧更新 MRF 分割模型中的表象模型. 具体做法是将目标跟踪框内的部件作为正样本,而框外部件作为负样本训练分类器. 为了减少噪声样本对表象模型的影响,更新机制同时从初始化第一帧和当前帧的目标搜索区域中提取样本. HSV 直方图更新如下所示:

$$M^{\text{hsv}} = M_0^{\text{hsv}} + M_t^{\text{hsv}} \quad (14)$$

其中 M_0^{hsv} 和 M_t^{hsv} 分别表示根据第一帧和当前帧的样本建立的 HSV 直方图, M^{hsv} 为更新后再归一化的 HSV 直方图. 此外,目标图 $G = (V, E)$ 更新可以捕捉跟踪目标发生的形变,通过对图中的节点状态(包括相对位置和表象特征)进行增加和删除实现. 当目标图节点 i 匹配上候选图节点 a ,同时满足匹配可靠性分数 $\varpi_{ia} > \theta_c$ 、表象相似度条件 $d_\chi(p_i, p_a) < \theta_a$ 和近邻条件 $\|l_i - l_a\| < \varepsilon$, 认为成功匹配. 其中 $\theta_c = 0.4$ 、 $\theta_a = 0.5$ 和 ε 分别为匹配可靠性分数阈值、表象阈值和近邻距离. 当候选图节点 a 未被目标图节点匹配上,并满足节点 a 和目标图中所有节点的距离大于近邻距离时,即 $\|l_i - l_a\| > \varepsilon$ 加入节点 a 状态. 当目标图节点 i 在连续 $N_f = 5$ 帧内未能成功匹配,则删除该节点 i .

3 实验与分析

本文选择了 11 段极具挑战性的视频序列(即 avatar, waterski, dancer, cliff-dive, transformer, bluecar, lipinski, torus, neymar, yunakim 和 diving),所要处理的跟踪挑战包括形变问题、遮挡问题、不规则运动和尺度变化等方面. 因为需要跟踪的所有目标都是非刚性的,非常适合用来评价我们的方法. 我们比较了 10 种现有方法,包括 IVT^[4], Struck^[5], STC^[13], CT^[19], Frag^[14], SPT^[6], SCM^[3], LSL^[7], LGT^[15] 和 DGT^[8].

3.1 实验参数设置

实验环境为主频 2.9GHz,内存 16GB 的 PC 机,算

法采用 Matlab 程序实现,在未经优化情况下跟踪器运行速度为 0.2fps. 实验中固定算法的所有参数,同时对比的跟踪方法也按照作者推荐的默认参数设置. 本文中采用的实验参数使用如下经验值:组成目标候选部件的超像素平均含有 100 个像素,其数目范围设定为 [100,400]. 在 2.5 节模型更新中 HSV 直方图每帧都更新, SVM 分类器每 $\eta = 3$ 帧更新一次. 对于分割公式 (3)、(5) 中预定的平衡参数设置为 $\lambda_1 = 0.9, \lambda_2 = 1.0, \lambda_3 = 15.0$. 表象特征使用各通道 6 位 HSV 直方图表示. 匹配类属矩阵 \mathbf{A} 中平滑系数 $\delta_c = \delta_g = 0.5$. 式 (13) 中系数 $\omega = 3.5$, 位置和尺度扰动项 $\delta_l = \delta_s = 10$.

3.2 实验结果分析

相比于以前方法采用平均位置误差 (Mean Location Error) 和跟踪成功率 (Success Rate) 来评估跟踪方法的性能,本文基于文献 [16] 使用了近年来更流行的曲线评价方式 (One-Pass Evaluation), 即给出跟踪成功率分别与位置误差和目标覆盖率大小的关系. 图 2(a) 表示成功率评价曲线 (success plots), 图 2(b) 表示精确度评价曲线 (precision plots). 从图 2 中两组曲线分数可以看出,本文提出的算法相比其他 10 种跟踪器具有更好的效果.

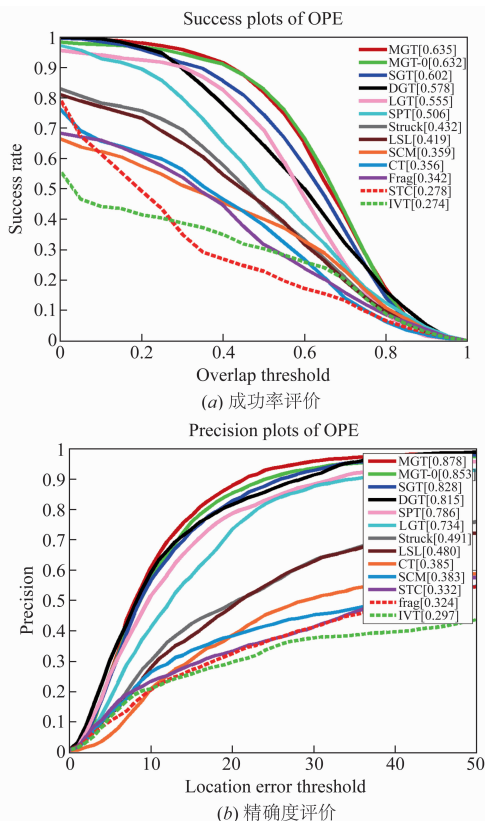


图2 成功率和精确度曲线

进一步地,我们在图 3 给出了序列下最优 4 种跟踪器的跟踪效果,选择代表序列为 avatar, waterski, trans-

former, bluecar 和 diving. 为了讨论方便,本文分别通过 4 种跟踪挑战进行分析.

3.2.1 剧烈形变

序列 avatar, waterski, transformer 和 diving 中目标发生了剧烈形变,使得固定尺寸的跟踪算法很难精确表示目标表象,容易造成跟踪失败(如 CT^[19]跟踪 avatar). 结合图 2,基于部件的跟踪器^[6,7,15]相比基于整体目标的跟踪器^[4,13,19]对形变目标跟踪精确度更高. 这是由于基于部件的跟踪器专注于局部部件表象而非全局目标表象,对目标结构变化不敏感.

3.2.2 目标遮挡

序列 avatar, waterski, transformer 和 bluecar 中的目标会出现一定程度的遮挡. 尽管一些跟踪算法声称对遮挡鲁棒,但是由于不能应对形变或者不规则运动等跟踪挑战,它们在实验测试中性能并不理想. 本文提出算法可以通过计算能成功匹配的目标部件比例检测遮挡发生的时刻,从而处理目标遮挡.

3.2.3 不规则运动

在序列 diving 中目标出现了快速运动、翻转和跳跃等复杂动作,基于固定目标尺寸的跟踪器^[8,14]很难精确表示. 而本文提出的跟踪器能在这类序列中表现出比较好的性能,得益于基于 MRF 的分割模型在目标搜索区域内提供了较为可靠的候选目标部件. 另外更高精度的部件匹配算法也抵御了不规则运动的不利影响.

3.2.4 尺度变化

跟踪目标会发生比较大的尺度变化,而基于部件的跟踪器相比于基于整体的跟踪器在处理尺度变化时具有更好的鲁棒性和自由度. 如图 3 所示,序列 avatar 和 waterski 中目标变化很大,以至于其他跟踪算法很难找到合适的目标框大小. 而本文提出的算法采用最大池图匹配方法得到精度更高的目标部件关联关系(见 2.3.2 节). 在较大的尺度变化挑战下,可以得到更精确的目标位置.

3.3 模型分析

本文首先分析了模型中的几个重要参数在不同取值下对跟踪效果的影响,包括组成目标候选部件的超像素数量 N (取值范围为 50, 100, 150, 200, 250, 记为 MGTsp), 分割式 (3) 中的平衡参数 $\{\lambda_1, \lambda_2\}$ (固定 $\lambda_2 = 1.0$, 参数 λ_1 取值范围为 0.5, 0.9, 1.0, 1.5, 2.0, 记为 MGTwt) 和式 (13) 中系数 ω (取值范围为 2.0, 2.5, 3.0, 3.5, 4.0, 记作 MGTcf). 超像素数量控制了目标部件的尺寸大小. 如图 4(a) 所示,本文提出的方法在各种取值下的超像素数量具有相似的效果,表明了模型的鲁棒性. 分割模型体现了产生式模型和鉴别式模型的共同作用,而两者合适的比例参数可以筛选出更精确的候选目标部件集合. 如图 4(b) 所示,当 $\lambda_1 = 0.9$ 时效果最

好. 系数 ω 表明了匹配结果对最后跟踪结果的可信度影响, 如果取得太大或者太小都会高估或低估其权重,

影响和背景像素的平衡性. 如图 4(c) 所示, 当 $\omega = 3.5$ 时效果最好.

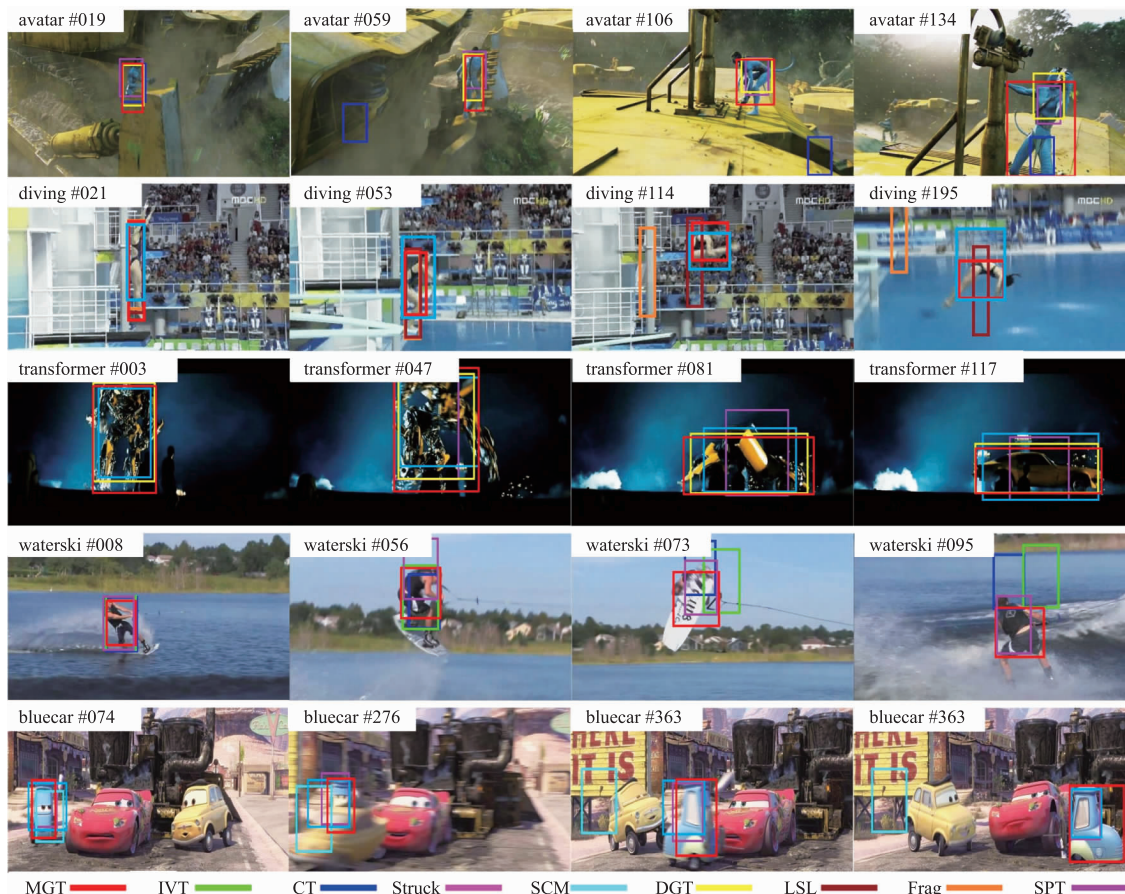


图3 跟踪器在各序列下跟踪效果

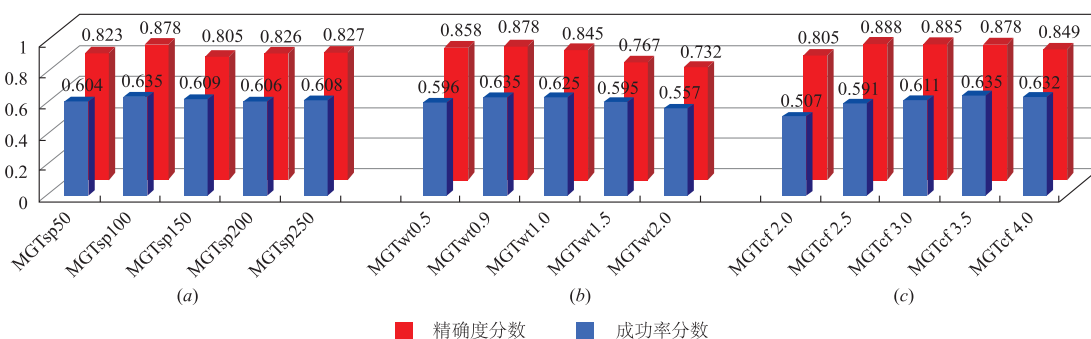


图4 模型参数分析

另外, 如图 2 所示, 相比于基于求和池或平均池图匹配的跟踪方法 (SGT), 基于最大池图匹配的跟踪方法 (MGT-0) 具有更高精确度. 进一步地, 同时考虑整体目标和局部目标部件的贡献, 本文方法 (MGT) 相比于只考虑局部部件表象模型 (MGT-0), 在效果上稍微有所提高. 图 5 给出了序列 avatar 下 SGT 和 MGT 的匹配跟踪情况, 其中在第 40 帧中 SGT 未匹配成功的目标上部部件在 MGT 方法下被成功匹配, 而第 110 帧中 SGT 匹

配到目标右边的噪声部件. 定义序列每一帧的目标部件平均匹配度 $\rho = 1/N \cdot \sum_i N_i^{\text{matched}} / N_i^{\text{fore}} \times 100\%$, 其中 N 为序列帧数, N_i^{matched} 和 N_i^{fore} 分别表示第 i 帧中成功匹配的节点数和目标部件数. 在序列 avatar 中, MGT 算法的匹配率为 61.01%, 高于 SGT 的 55.41%. 综上所述, 本文提出方法取得了较好的效果提升, 实验证明了方法的有效性.

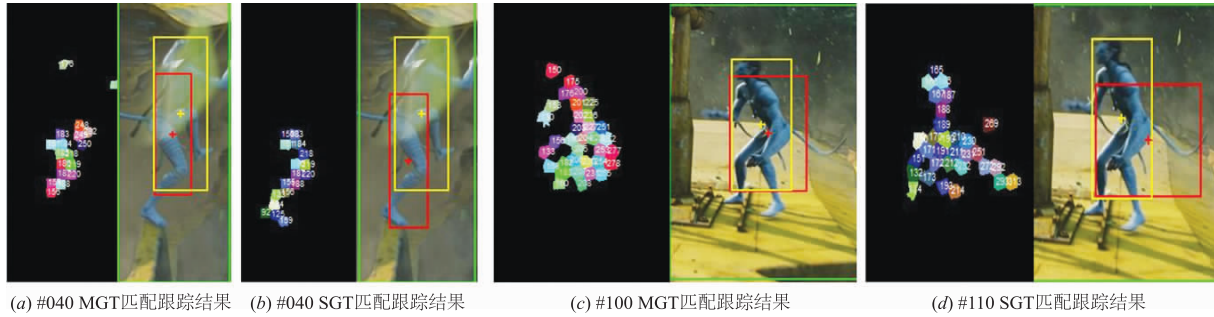


图5 最大池和求和池匹配结果对比

4 结束语

本文提出了一种基于最大池图匹配的形变目标跟踪方法,将跟踪问题转化为图匹配问题.其中目标图和候选图中部件的匹配精确度决定了跟踪性能.同时考虑整体目标和局部部件对目标位置的支持,进一步提升跟踪鲁棒性.实验在 11 个序列上比较了最近的 10 种跟踪算法,同时参数分析证明了本文提出算法的有效性.未来可考虑引入局部多目标部件之间的结构关系,进一步增强局部结构的保持性,提升跟踪效果.

参考文献

- [1] X Mei, H Ling. Robust visual tracking using L1 minimization [A]. IEEE International Conference on Computer Vision [C]. Kyoto: IEEE, 2009. 1436 – 1443.
- [2] T Zhang, B Ghanem, S Liu, et al. Robust visual tracking via multi-task sparse learning [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Providence: IEEE, 2012. 2042 – 2049.
- [3] D Du, H Qi, W Li, L Wen, Q Huang, S Lyu. Online deformable object tracking based on structure-aware hypergraph [J]. Transactions on Image Processing, 2016, 25 (8): 3572 – 3584.
- [4] J Lim, D A Ross, R-S Lin, et al. Incremental learning for visual tracking [A]. IEEE Advances in Neural Information Processing Systems [C]. Vancouver: IEEE, 2004. 793 – 800.
- [5] S Hare, A Saffari, P H Torr. Struck: Structured output tracking with kernels [A]. IEEE International Conference on Computer Vision [C]. Springs: IEEE, 2011. 263 – 270.
- [6] S Wang, H Lu, F Yang, M-H Yang. Superpixel tracking [A]. IEEE International Conference on Computer Vision [C]. Barcelona: IEEE, 2011. 1323 – 1330.
- [7] R Yao, Q Shi, C Shen, et al. Part-based visual tracking with online latent structural learning [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Portland: IEEE, 2013. 2363 – 2370.
- [8] Z Cai, L Wen, Z Lei, N Vasconcelos, et al. Robust deformable and occluded object tracking with dynamic graph [J]. IEEE Transactions on Image Processing, 2014 (12): 5497 – 5509.
- [9] L Wen, D Du, Z Lei, et al. JOTS: Joint online tracking and segmentation [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Boston: IEEE, 2015. 2226 – 2234.
- [10] J Chang, J W F III. Topology-constrained layered tracking with latent flow [A]. IEEE International Conference on Computer Vision [C]. Sydney: IEEE, 2013. 161 – 168.
- [11] Y Boykov, V Kolmogorov. An experimental comparison of mincut/max-flow algorithms for energy minimization in vision [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, (9): 1124 – 1137.
- [12] Cho M, Sun J, Duchenne O, et al. Finding matches in a haystack: A max-pooling strategy for graph matching in the presence of outliers [A]. IEEE International Conference on Computer Vision and Pattern Recognition [C]. Columbus: IEEE, 2014. 2091 – 2098.
- [13] K Zhang, L Zhang, Q Liu, et al. Fast visual tracking via dense spatio-temporal context learning [A]. European Conference on Computer Vision [C]. Zurich: Springer International Publishing, 2014. 127 – 141.
- [14] A Adam, E Rivlin, I Shimshoni. Robust fragments-based tracking using the integral histogram [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. New York: IEEE, 2006 (1). 798 – 805.
- [15] L Cehovin, M Kristan, A Leonardis. Robust visual tracking using an adaptive coupled-layer visual model [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013 (4): 941 – 953.
- [16] Y Wu, J Lim, M-H Yang. Online object tracking: A benchmark [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Portland: IEEE, 2013. 2411 – 2418.
- [17] R Achanta, A Shaji, K Smith, A Lucchi, P Fua, S Süsstrunk. SLIC superpixels compared to state-of-the-art superpixel methods [J]. IEEE Transactions on Pattern A-

- analysis and Machine Intelligence, 2012, 34 (11): 2274 - 2282.
- [18] Zhao G, Ahonen T, Matas J, et al. Rotation-invariant image and video description with local binary pattern features [J]. IEEE Transactions on Image Processing, 2012 (4): 1465 - 1477.
- [19] K Zhang, L Zhang, M-H Yang. Real-time compressive tracking [A]. European Conference on Computer Vision [C]. Florence: Springer International Publishing, 2012. 864 - 877.

作者简介



王治丹 女,1989 年生,江苏东台人,硕士,研究方向:数字图像处理、目标检测以及智能视频监控系統.

E-mail:wangzhidan_dsp@mail.hfut.edu.cn



蒋建国 男,1955 年生,安徽广德人,教授,博士生导师,研究方向:分布式智能系統、视频信息处理与分析以及 DSP 技术应用.

E-mail:jgjiang@hfut.edu.cn.cn



齐美彬 男,1969 年生,安徽东至人,教授,硕士生导师,研究方向:视频编码、视觉监控以及 DSP 技术应用.

E-mail:qimeibinhf@gmail.com