

基于可变形模型的目标跟踪算法

马俊凯^{1,2,3}, 罗海波^{1,2}, 常 铮^{1,2}, 惠 斌^{1,2}, 周晓丹⁴, 侯德飞⁵

(1. 中国科学院沈阳自动化研究所, 辽宁 沈阳 110016; 2. 中国科学院光电信息处理实验室, 辽宁 沈阳 110016; 3. 中国科学院大学, 北京 100049; 4. 空军驻湖北地区军事代表室, 湖北 武汉 430000; 5. 空军装备部装备采购局, 北京 100843)

摘要:近年来目标跟踪技术的研究已经有了很大的进展,但目标的遮挡和形变仍然是目标跟踪算法面临的重大挑战。针对这些问题提出了一种基于可变形模型的目标跟踪算法。首先,利用可变形模型对跟踪目标进行表达,该模型将目标分为若干子块,目标的特征由局部子块特征和全局特征共同构成。将目标的特征和子块之间的空间关系结合起来,给出了对目标的一个统一的相似度量函数。然后,在线训练一个结构化输出支持向量机作为分类器,该分类器的输出是可变形模型中目标的结构化描述。利用该分类器可以在视频及图像序列中准确地检测到目标,完成跟踪。通过实验比较,该算法的跟踪性能优于其他主流跟踪算法,尤其在目标发生遮挡和形变的时候仍能准确跟踪。

关键词: 可变形模型; 结构化分类器; 在线学习; 目标跟踪

中图分类号: TP391.4 **文献标志码:** A **DOI:** 10.3788/IRLA201746.0928001

Visual tracking algorithm based on deformable parts model

Ma Junkai^{1,2,3}, Luo Haibo^{1,2}, Chang Zheng^{1,2}, Hui Bin^{1,2}, Zhou Xiaodan⁴, Hou Defei⁵

- (1. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China;
2. Key Laboratory of Opt-Electronic Information Processing, CAS, Shenyang 110016, China;
3. University of Chinese Academy of Sciences, Beijing 100049, China;
4. Military Representative Office of Air Force in Hubei Area, Wuhan 430000, China;
5. Air Force Reserve Equipment Procurement Bureau, Beijing 100843, China)

Abstract: In recent years, the technology of target tracking has been greatly developed, but occlusion and deformation of the target were still the major challenges in tracking algorithms. To address these problems, a tracking algorithm based on deformable parts model (DPM) was proposed. Firstly, DPM was used to represent the target object. DPM divided the target into several small parts. The feature of the target was composed of the local feature of each part and the global feature of the entire object, then DPM defined a uniform similar function based on the object feature and spatial relationship of each pair of parts. Secondly, a structured output support vector machine (structured SVM) was trained online as the classifier, the output of the structured SVM was the structured description of the object. The target in videos or image sequences could be tracked by the detection result of the classifier. Experimental results

收稿日期: 2017-01-10; 修订日期: 2017-02-20

作者简介: 马俊凯(1989-), 男, 博士生, 主要从事机器学习、图像处理、目标跟踪等方面的研究。Email: majunkai@sia.cn

导师简介: 罗海波(1967-), 男, 研究员, 博士生导师, 博士, 主要从事光电成像、图像处理、目标识别、目标跟踪等方面的研究。

Email: luohb@sia.cn

demonstrate that the proposed methods outperform other popular trackers, especially with the challenge of object's occlusion and deformation.

Key words: deformable parts model; structured output classifier; online learning; object tracking

0 引言

近年来计算机视觉飞速发展,其中智能视频监控、人机交互、行为识别、机器人视觉等技术已日趋成熟。目标跟踪算法是这些技术的重要组成部分,目标跟踪算法的精确性、鲁棒性对这些技术的发展、应用起着至关重要的作用。虽然目标跟踪算法层出不穷,但光照变化、部分遮挡、复杂背景以及目标变形等仍然是目标跟踪算法面临的重大挑战。

目前主流的目标跟踪方法主要分为两种:生成式算法和判别式算法。生成式算法通常先要生成目标的表观模型,然后在每一帧图片中寻找一个与目标表观模型相似度最高的区域作为目标跟踪的结果^[1-3]。Zhong 等人^[1]提出了一种 L1 跟踪算法,利用稀疏字典作为目标的表观模型,使用粒子滤波来估计目标的状态。Ross 等人^[2]使用一个低维的子空间来描述物体,通过增量式的学习方法更新目标的表观模型,以适应跟踪过程中目标的变化。崔雄文^[3]等人采用 SIFT 特征构造目标的 PCA 子空间,使用粒子滤波理论框架来对目标进行跟踪。判别式算法通常将目标跟踪问题看作一个机器学习中的分类问题^[4-6],通过训练出一个分类器来对目标和背景进行区分,从而在每一帧中跟踪到目标。Grabner 等人^[4]使用在线学习的方法在跟踪过程中训练一个 AdaBoost 分类器,利用这个分类器在每一帧中检测目标。Kalal 等人^[5]将跟踪(Tracking)、检测(Detection)、学习(Learning)三个模块融合起来提出了 TLD 跟踪算法,该算法将目标检测模块加入到跟踪算法中,在每帧输入图像中检测目标,利用检测到的目标位置与跟踪器的输出结果做融合得到目标的位置。Henriques^[7]等人将跟踪问题转化为一个回归问题,利用训练样本的循环特性以及核相关滤波器得到一个高速的跟踪算法 KCF。邢运龙^[8]等人在 KCF 的基础上加入了对光照变化的处理等,提高了 KCF 的鲁棒性。

在大多数的判别式跟踪算法中用到的分类器是

二分类器,即分类器的输出为 1、0(或 1、-1),分别表示是或不是目标。在二分类器的训练过程中采用正负两类样本进行训练。这类算法首先在第一帧中人为标出目标,然后使用分类器在后续的每帧图像中跟踪目标。具体处理过程如下:(1)使用分类器将目标从背景中分割出来;(2)标记正负样本,一般选取跟踪到的目标为正样本,在距离样本较远的位置或与样本重合度较低的位置取负样本;(3)使用正负样本更新分类器。虽然这类算法在跟踪中取得了较为理想的效果,但仍然存在一些缺陷。在分类器更新过程中通常是选取跟踪器输出的位置上的图片块为正样本,在正样本周围取负样本。例如在正样本周围与正样本重合 30%的图像块和与正样本重合 10%的图像块都被标记为负样本,这两个样本在分类器更新的过程中起到的作用是一样的,而笔者等期望的是与正样本重合度不同的负样本在分类器的训练过程中占有不同的权重。基于这个目的,Hare 等人^[6]提出了名为 Struck 的跟踪算法,该方法将目标看作一个整体,将两帧之间目标的位移向量看作结构化的输出,使用结构化的支持向量机(structured SVM)作为分类器。该文献中分类器将两帧之间目标的位置变化看作结构化的输出变量,大大提高了分类器的性能。利用此分类器,Struck 的跟踪性能获得了很大的提高。由于其使用整个目标图像块中的特征来表示目标,故在对目标细节描述上稍显不足。

以上方法取得了较好的效果,但对于解决目标跟踪中存在的目标变形以及目标遮挡等问题,仍然效果不佳。文中提出利用可变形模型对目标进行表达,可有效解决这两类问题。可变形模型采取先将目标分块再"分而治之"的思想。在对目标分块的基础上对各子块之间的空间关系进行度量,用来衡量其形变的大小,可变形模型给出了目标的一个统一的相似度度量。同时,采用结构化的分类器在跟踪过程中区分出目标和背景,较传统的二分类器而言,也大大提高了目标跟踪精度。

1 可变形模型的目标表示方法

近些年基于分块的可变形模型(Deformable Parts Model)在目标检测领域取得了令人瞩目的成果。Felzenszwalb 等人于 2010 年使用可变形模型进行目标检测^[9],获得了著名的目标检测比赛“PASCAL VOC”的终身成就奖。可变形模型在目标检测中达到的效果充分证明了可变形模型在目标表达上的鲁棒性,故文中提出将可变形模型用于目标跟踪中。

对于任意一个目标,首先将目标分为若干个子块,由这些图像子块以及子块之间的连接关系构成该目标的可变形模型。可将一个目标的可变形模型表示为一个图 $G=(V,E)$,其中 V 是图的节点,表示目标的每个子块; E 是图的边,表示各个子块之间的空间位置关系。这个模型是一个通用的可变形模型。为了适用于目标跟踪算法,同时提高问题求解的效率,笔者在文中为以上模型加上特定的约束条件,即将该模型中的图 G 设定为一个深度为 2 的树,每颗树包含一个根节点与若干个子节点。根节点表示跟踪目标的整体,子节点对应每个子块。对一个具体的目标,其表示方法如图 1 所示。该目标包含五个节点及四条边,五个节点分别记作 v_0, v_1, \dots, v_4 ,其中 v_0 表示根节点,包含了目标的整体, v_1, \dots, v_4 分别表示目标的四个子块。四条边 e_{01}, \dots, e_{04} 分别表示根节点与分块 1 至 4 号子块的连接。使用一个包含目标整体的矩形框 $l_0=(c_0, r_0, h_0, w_0)$ 来表示根节点 v_0 ,其中 c_0, r_0 分别表示矩形框左上角在图片中的行列位置; h_0, w_0 分别表示矩形框的高度和宽度。 $l_i=(c_i, r_i, h_i, w_i), i=1, 2, \dots, n$ 分别表示 n 个子节点。提取每个节点中矩形框内图像的特征 $f(l_i)$,将这些特征与每个矩形框左上角的行列坐标结合起来,共同构成节点的特征 $v_i=(\phi(l_i), \Delta c_i, \Delta r_i)$,其中 $\Delta c_i=c_i-c_0; \Delta r_i=r_i-r_0$ 。

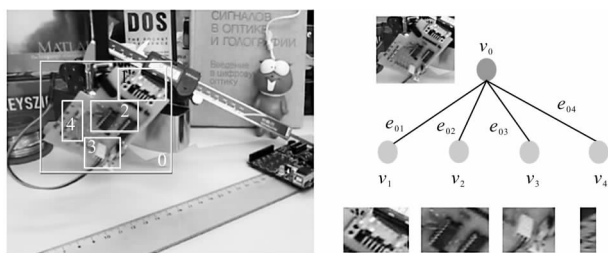


图 1 目标的可变形模型示意图

Fig.1 Deformable model of object

这样这个模型既包含了目标的局部特征又包含了各个局部的空间位置关系。将每个候选目标表示为各个子块的集合 $L=(l_0, l_1, \dots, l_n)$ 。

该候选目标与目标模型的相似度可以定义为:

$$S = \sum_{i=1}^n m_i(l_i) + \sum_{e_{ij} \in E} e_{ij}(l_i, l_j) \quad (1)$$

式中: $m_i(l_i)$ 为节点 v_i 在位置 l_i 时与目标模型之间的相似度; $e_{ij}(l_i, l_j)$ 为节点 v_i 与节点 v_j 分别在位置 l_i 以及 l_j 时两者的空间位置与模型中这两者位置关系之间发生形变误差的度量。以上这两个方面共同衡量了候选的目标与目标模型的相似程度。

定义可变形模型的表达之后就可以给出模型的参数表示。整个模型包括根节点以及每个子节点的参数 $w=(w_0, w_1, \dots, w_n)$, 其中 w_0 表示根节点参数, $w_i, i=1, 2, \dots, n$ 为子节点的参数。每个节点的参数 w_i 又由相似度参数 w_{ai} 以及空间形变参数 w_{di} 共同组成, 即 $w_i=(w_{ai}, w_{di})$ 。在模型参数表达的基础上公式(1)的准确定义为:

$$S(L) = \sum_{i=0}^n \langle w_i, v_i \rangle = \sum_{i=0}^n (\langle w_{ai}, \phi(l_i) \rangle + \langle w_{di}, (\Delta c_i, \Delta r_i) \rangle) = \sum_{i=0}^n \langle w_{ai}, \phi(l_i) \rangle + \sum_{i=1}^n \langle w_{di}, (\Delta c_i, \Delta r_i) \rangle \quad (2)$$

公式(2)中等号后边的两项分别等价于公式(1)等号右边的两项。在一幅新图像 x 上检测出目标的过程为: 在新图像上选取在候选样本位置的集合 $L=\{L_1, L_2, \dots, L_m\}$, 在集合 L 中选取使公式(2)达到最大值的一个候选样本作为目标模型最佳的匹配:

$$L^* = \operatorname{argmax}_{L_i \in L} (S(L_i)) \quad (3)$$

2 结构化分类器

文中采用的可变形模型记录了各子块的空间位置,也就是说包含了目标中丰富的结构信息。如果在跟踪中使用传统的二分类器,除了有引言中所述的不足之外,也没有充分利用到可变形模型的优势,所以在跟踪中采用结构化支持向量机作为分类器来对目标位置进行预测。结构化支持向量机^[10]学习出一个从特征空间 x 到输出空间 y 的预测函数 $f: x \rightarrow y$ 。 y 中的元素是结构化的变量,例如:图,有序序列、字符

串序列等。训练样本是输入-输出对 $(x, y) \in x \times y$ 的集合。在训练样本上训练出判别函数 $f: x \times y \rightarrow R$ 预测函数可以定义为:

$$f(x) = \operatorname{argmax}_{y \in Y} F(x, y) \quad (4)$$

即对一个任意输入的 x 选取能使判别函数 $F(x, y)$ 达到最大值的 y 作为预测结果。

在跟踪算法中,每一帧图像作为输入变量,选取上一节中目标及各子块的位置为结构化分类器的输出变量 y , 即 $y = (l_0, l_1, \dots, l_n)$ 。这样结构化支持向量机中的预测函数等价于公式(3), 而判别函数则等价于公式(2)所示的相似度函数。结合目标的参数模型及公式(2), 给出文中所使用的结构化分类器中判别函数的表达式:

$$F(x, y; w) = \langle w_i, \phi(x, y) \rangle = \sum_{i=0}^n \langle w_i, v_i \rangle = \sum_{i=0}^n (\langle w_{di}, \phi_d(l_i) \rangle + \langle w_{di}, \phi_d(\Delta c_i, \Delta r_i) \rangle) \quad (5)$$

与支持向量机相同, 结构化支持向量机也是一种最大间隔分类器。其参数 w 可以通过求解以下二次规划问题得到:

$$\begin{aligned} \min_w \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t. } \forall i: \quad & \xi_i \geq 0 \\ & \forall i, \forall y \neq y_i: \langle w, \delta \phi_i(y) \rangle \geq \Delta(y_i, y) - \xi_i \end{aligned} \quad (6)$$

式中: $\langle w, \delta \phi_i(y) \rangle = \langle w, \phi(x_i, y_i) - \phi(x_i, y) \rangle$; C 为正则化参数; ξ_i 为松弛变量; $\Delta(y_i, y)$ 为两个输出变量之间的误差函数, 当 $y = \bar{y}$ 时有 $\Delta(\bar{y}, y) = 0$, $\Delta(\bar{y}, y)$ 随着 y 与 \bar{y} 之间差别的增大而增大。公式(6)中约束条件的意义是: 对于每一帧输入图像 x_i , 通过公式(4)求出这幅图像中目标的最佳匹配后, 将其记为 y_i , 则在之后的训练过程中, 训练样本 (x_i, y_i) 被看作是“最正确”的样本, $\forall y \in Y, \forall y \neq y_i$ 均有以下公式成立:

$$\begin{aligned} \langle w, \delta \phi_i(y) \rangle &= \langle w, \phi(x_i, y_i) - \phi(x_i, y) \rangle = \\ & F(x_i, y_i) - F(x_i, y) > 0 \end{aligned} \quad (7)$$

这样, 对于判别函数来说, 总是在“最正确”的样本 (x_i, y_i) 上取得最大值。

为了方便求解公式(6), 利用拉格朗日对偶原理, 引入拉格朗日乘子 β , 并且经过适当的变量代换将其转化为对偶形式^[9]:

$$\begin{aligned} \max_{\beta} \quad & \sum_{i, y} -\Delta(y, y_i) \beta_i^y - \sum_{i, y, j, \bar{y}} \frac{1}{2} \beta_i^y \beta_j^{\bar{y}} k(x_i, y, x_j, \bar{y}) \\ \text{s.t. } \forall i, \forall y: \quad & \beta_i^y \leq \delta(y, y_i) C \\ & \forall i: \sum_y \beta_i^y = 0 \end{aligned} \quad (8)$$

式中: 当 $y = \bar{y}$ 时 $\delta(y, \bar{y})$ 的值为 1, 否则为 0; $k(x_i, y, x_j, \bar{y})$ 为核函数, 用来对原始问题中的线性模型进行非线性扩展:

$$k(x_i, y, x_j, \bar{y}) = \langle \phi(x_i, y), \phi(x_j, \bar{y}) \rangle \quad (9)$$

转换为对偶问题之后, 原问题中的判别函数可以表示为:

$$F(x, y) = \sum_{i, \bar{y}} \beta_i^{\bar{y}} k(x_i, \bar{y}, x, y) \quad (10)$$

这样, 求解公式(8)得到对偶变量 β_i^y 的值即为模型的参数值。对对偶变量 β_i^y 做如下解释: 对于大部分样本 (x_i, y) , 其对应的对偶变量的值为 0, 只有支持向量 (x_i, \bar{y}) 对应的对偶变量 $\beta_i^{\bar{y}} \neq 0$ 。且对于每一个 x_i , 只存在一个样本 (x_i, y_i) 对应的 $\beta_i^y < 0$, 该样本为正支持向量; 另外存在样本 (x_i, y_i) , 对应的 $\beta_i^y < 0$, 这些样本为负支持向量。由公式(10)可以看出, 在跟踪过程中, 只需储存一个支持向量的集合 S 及其对应的对偶变量 β_i^y 的值, 就能得到预测函数 $f(x)$ 。文中使用基于序贯最小优化算法(SMO)改进的在线优化算法 LaRank 来优化参数 β 的值^[11], 达到实时更新模型的目的。

结合以上算法, 给出了图 2 所示的文中的跟踪算法流程。首先, 在视频的第一帧中标出所需跟踪的目标, 初始化目标的可变形模型以及结构化支持向

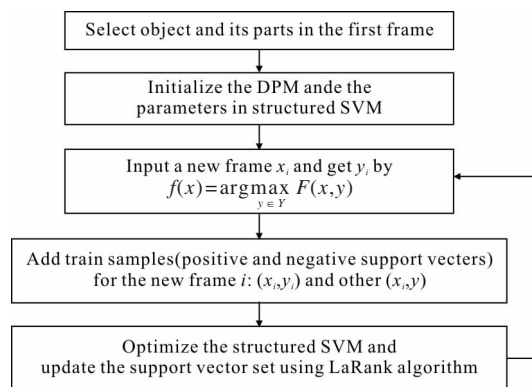


图 2 算法流程图

Fig.2 Flow chart of proposed method

量机的参数。然后对每一帧新图像 x_t 来说,使用结构化支持向量机的预测函数 $f(x)$ 可得到目标的位置 y_t 。确定 (x_t, y_t) 为正支持向量,并在这帧图像中找到一个负支持向量,使用 LaRank 算法,更新结构化支持向量机参数 β 以及支持向量集合 S 。

3 预测函数求解算法

文中使用可变形模型来表示目标,一个目标由多个子块组合构成。假设目标由一个根节点以及 $N-1$ 个子节点构成,候选目标可表示为 $L=(l_0, l_1, \dots, l_{N-1})$,同时每个子块的候选个数为 M ,即每个 l_i 有 M 个取值,则 L 共有 M^N 个候选取值。在这种情况下,结构化的输出空间的规模将呈指数形式增长。若对于一个目标取 $M=200, N=4$ 则每次使用公式(4)对一个输入 x 做预测需要计算 16×10^8 个样本的判别函数 $F(x, y)$ 的值。这就严重限制了可变形模型在跟踪中的应用。

通过分析公式(10)的形式,对其做等价变换并采用动态规划的算法来求解公式(4)中的最大值问题。

利用核函数具有的线性组合性质^[12],可以对公式(10)进行等价变换:

$$F(x, y; w) = \sum_{i, \bar{y}} \beta_i^{\bar{y}} k(x_i, \bar{y}, x, y) = \sum_{i, \bar{y}} \beta_i^{\bar{y}} (k_f^0 + k_d^0 + k_f^1 + k_d^1 + \dots + k_f^N + k_d^N) \quad (11)$$

式中: k_f^j 、 k_d^j 分别为用于计算第 j 个子块相似度的核函数以及变形量的核函数。相似度核函数 k_f^j 度量的是在位置 l_j 上的图像块与第 j 个子块的相似度。每个形变量核函数 k_d^j 度量的是第 j 个子块在位置 l_j 处与根节点在 l_0 处时目标的形变量。由此可将公式(4)表示为:

$$f(x; w) = \operatorname{argmax}_{y \in Y} F(x, y; w) = \operatorname{argmax}_{y \in Y} \sum_{i, \bar{y}} \beta_i^{\bar{y}} (k_f^0 + k_d^0 + k_f^1 + k_d^1 + \dots + k_f^N + k_d^N) = \operatorname{argmax}_{y \in Y} \sum_{i, \bar{y}} \left[\beta_i^{\bar{y}} (k_f^0(l_0) + \sum_{j=1}^n (k_f^j(l_j) + k_d^j(d(l_0, l_j)))) \right] = \operatorname{argmax}_{l_0 \in L_0} \sum_{i, \bar{y}} \beta_i^{\bar{y}} k_f^0(l_0) + \operatorname{argmax}_{l_j \in L_j} \sum_{i, \bar{y}} \beta_i^{\bar{y}} (k_f^j(l_j) + k_d^j(d(l_0, l_j))) \quad (12)$$

式中: $d(l_0, l_j)$ 为 l_0 与 l_j 之间的距离函数。通过这个公式把求解一个可变形模型整体 $y=(l_0, l_1, l_2, \dots, l_n)$ 的相似度响应最大值的过程分解成对根块求最大相似度响应位置 $y_0=l_0$, 以及对每个子块求其各自的最大相似度响应位置 $y_j=l_j$ 的问题。即公式(12)完成了从对一个变量组合求最值到对各个变量单独求最值的转换。从公式(12)最下一行的求和式还能看出每个子块的最大值位置还依赖于根块位置 l_0 。为了表述清晰,将公式(12)中最后一个等号右边第一项定义为:

$$S_{\text{root}}(l_0) = \sum_{i, \bar{y}} \beta_i^{\bar{y}} k_f^0(l_0) \quad (13)$$

第二项定义为:

$$S_{\text{part}}(l_j; l_0) = \sum_{i, \bar{y}} \sum_{j=1}^n \beta_i^{\bar{y}} (k_f^j(l_j) + k_d^j(d(l_0, l_j))) \quad (14)$$

公式(12)求解步骤如下:固定一个 l_0 , 求出在此 l_0 下 $S_{\text{part}}(l_j; l_0)$ 的最大值并记录对应的 l_j , 求出所有 l_0 取值下各子块的 $S_{\text{part}}(l_j; l_0)$ 之后, 将其与 $S_{\text{root}}(l_0)$ 相加, 此时结果即为公式(12)的整体响应函数。最后找出整体响应函数的最大值, 该最大值对应的位置即为最优的 l_0 , 并且根据 $S_{\text{part}}(l_j; l_0)$ 中记录的 l_j 信息确定第 j 子块的最优位置。其中 $S_{\text{part}}(l_j; l_0)$ 可看作一个权值 $\beta_i^{\bar{y}}$ 与距离变换函数^[13] (Distance Transforms Function) 的乘积, 利用相关参考文献[13]中的距离变换函数求解算法可高效求解上式, 将预测函数算法的时间复杂度由 $O(M^N)$ 降低到 $O(M \cdot N)$ 。

图 3 展示了针对 Sylvester 序列中的一帧图像

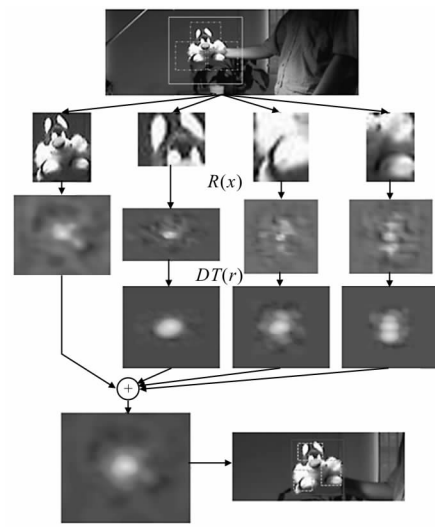


图 3 预测函数流程图

Fig.3 Flow chart of the predication function

求解预测函数(即预测目标所在位置)的流程。首先,使用滑动窗口在目标候选区域上对一个根块及三个子块进行稠密采样,再求出每个子块在这些区域上的相似度响应图,利用每个子块的相似度响应图求出距离变换响应图,最后将三个子块的距离响应图与根块的相似度响应图加起来。最大值响应值的位置即为预测的目标位置,子块的位置确定方法如上所述。

4 实验结果

文中的实验细节如下。

从图像块中提取 Haar 特征来对每个块的相似度进行度量。对每一个图像块,提取其中不同位置,不同范围的 32 个区域,对于每个区域,提取如图 4 所示的六类 Haar 特征。这样每个图像块对应一个特征向量,该特征向量的维数是 $32 \times 6 = 192$ 。



图 4 六种 Haar 特征

Fig.4 Six types Haar features

相似度核函数使用高斯核函数:

$$k(x_i, y_i, x, y) = \exp(-\sigma \|\phi(x_i, y_i) - \phi(x, y)\|) \quad (15)$$

式中: $\sigma = 0.2$ 。该实验中使用距离函数形式为 $d(l_0, l_i) = w_c \Delta c_i + w_r \Delta r_i$ 来度量图像子块之间的形变量。结构化的支持向量中,正则化参数 C 取 100。

可变形模型使用分块的策略可根据应用场景的不同而灵活选择。为了保证跟踪算法的稳定性,要尽量保证选择的分块区域与周围的区域区分性较大,这样才能保证各个子块跟踪结果的可靠性,在此基础上预测出来的目标准确较高。分块数目不限定。如图 5 中 Sylvester 序列上,选择目标(玩具)的头部,两脚共三个子块,这三个子块对周围的区域来说具有较强的区分性。

为了验证文中算法的有效性,在公开的视频跟踪数据集^[4]上做了测试。图 5 显示了该算法在 Sylvester 序列上第 1、230、550、670 帧上的跟踪结果(列出第 1 帧的结果作为其他帧中目标的形变参考,以下几个序列中采用同样的做法)第 220 帧上目标缩小,该算法指示出各子块的位置相互靠近,第

550 帧上目标向右发生了偏转,第 670 帧中,目标发生了向左偏转,在这些形变下本算法均能很好的跟踪到子块,并根据子块的位置关系预测出目标的位置。从图中也可以看出当目标发生形变时各子块相对位置的变化。

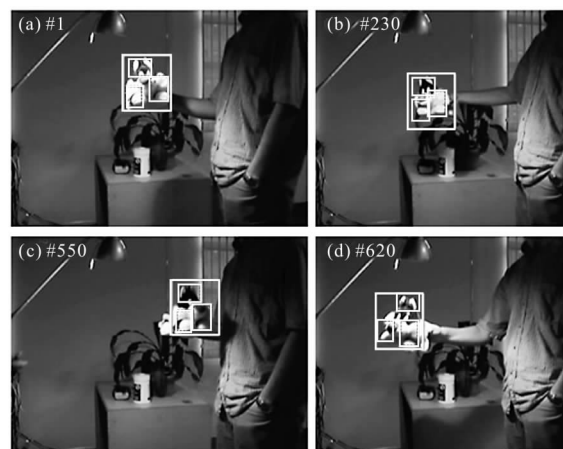


图 5 Sylvester 序列结果示例

Fig.5 Results of the Sylvester sequence

图 6 展示了该算法在 Dancer2 序列中的结果。可以看出当目标发生形变时,各个子块的跟踪结果之间的位置关系也发生了变化。在第 125 帧中,目标由站姿改成了半蹲状态,形状发生了较大的变化,文中算法能准确检测出这种形变关系并且预测出目标较为准确的位置。

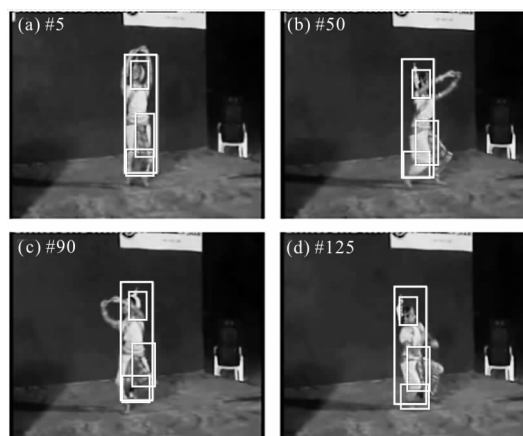


图 6 Dancer2 序列结果示例

Fig.6 Results of the Dancer2 sequence

图 7 显示了 Board 序列的跟踪结果,从第 260 帧以及第 344 帧中可以看到,当目标变大或缩小时,跟踪到的子块的位置关系也变得紧凑或发散。在第 485 帧中,目标发生了非平面旋转,从结果可以看

出,该算法仍能很好地跟踪到目标。

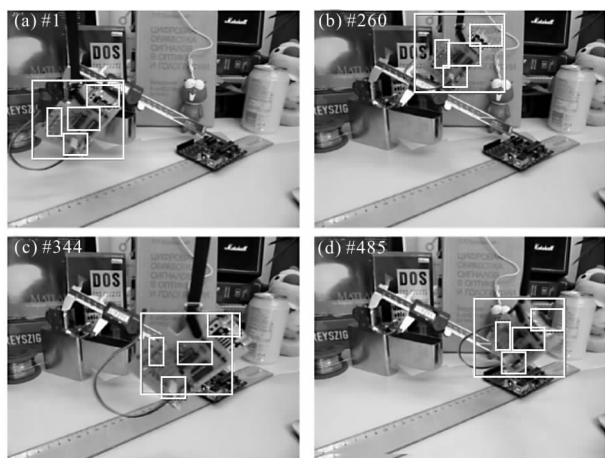


图 7 Board 序列结果示例

Fig.7 Results of the Board sequence

图 8 显示了该算法在Gym 序列上的跟踪结果,在此序列中,选取上身和腿为两个分块,目标发生形变,如在第 70 帧上身前探,第 416 帧中转身,第 650 帧中俯身等,文中算法的模型可以检测到子块相对位置的变化,并预测出完整目标的位置。

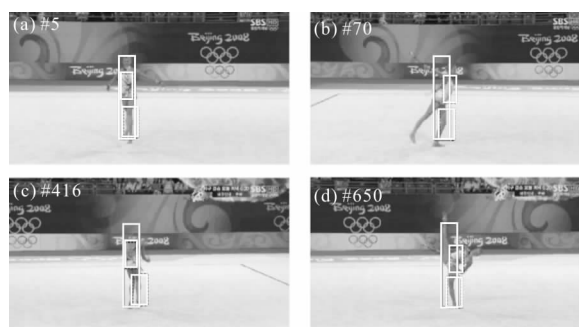


图 8 Gym 序列结果示例

Fig.8 Results of the Gym sequence

为了评估算法的性能,将文中算法与目前最新的跟踪算法 LIAPG^[15]、TLD^[5]、Struck^[6]、KCF^[7]做了对比。图 9 显示在以上四个视频的若干目标发生形变的时刻,文中的算法与比较算法的对比结果。明显可以看出文中算法较其他算法更为准确。为了更准确地测评文中算法,使用了跟踪结果与目标真实位置的中心位置误差(CLE)作为评价指标在 Board、Dog1、FaceOcc1、Gym、Dancer2、LemminSylvester、Trellis 这八个视频序列上对算法做了测试。图 10

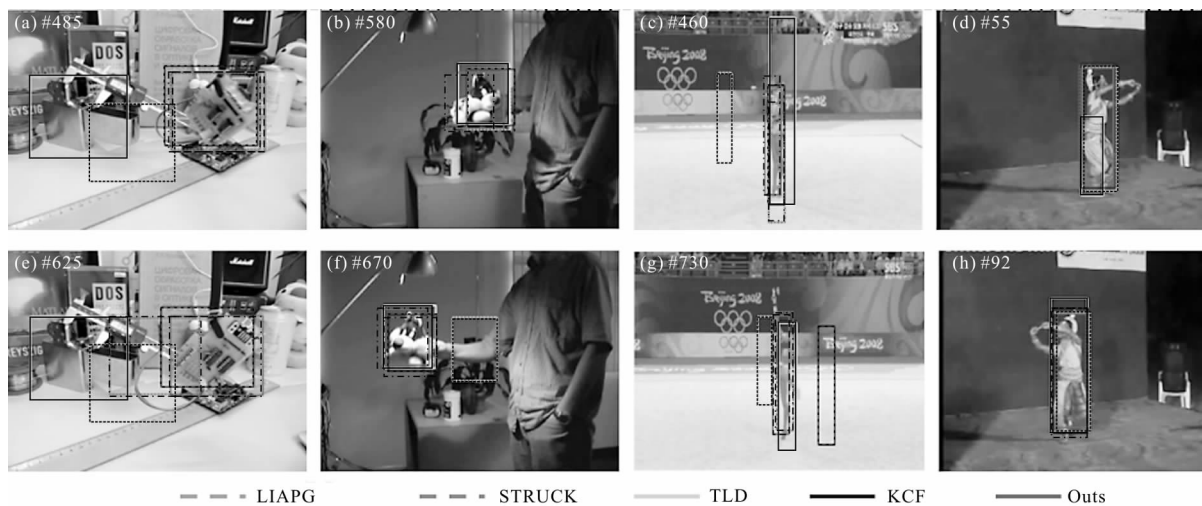


图 9 典型帧的算法对比图

Fig.9 Comparison of algorithms in typical frame

显示了文中算法与其他四种对比算法在以上八个视频序列上的 CLE 结果。图中横坐标表示中心位置误差的阈值(以像素为单位),范围从 0 个像素到 50 个像素,纵坐标表示跟踪结果的 CLE 小于对应横坐标阈值的成功率。以下采用 CLE 阈值为 20 个像素时算法的跟踪成功率来评价算法:在 FaceOcc1 序列上领先第二名算法(KCF)19.8%;在

Board 序列中文中算法的领先第二的算法(Struck) 16.88%;在 Gym 序列上文算法领先第二名(KCF)16.6%;在 Lemming 序列中文中算法领先第二名(TLD)2.6%;在 Dancer2、Dog1、Sylvester 序列上,文中算法与其他算法基本持平,均在较高水平。在 Trellis 序列中,虽然在 20 像素阈值位置上文中算法与 KCF 基本持平,但在小于 20 像素阈值下,

跟踪性能优于较好的 KCF 算法和 Struck 算法。

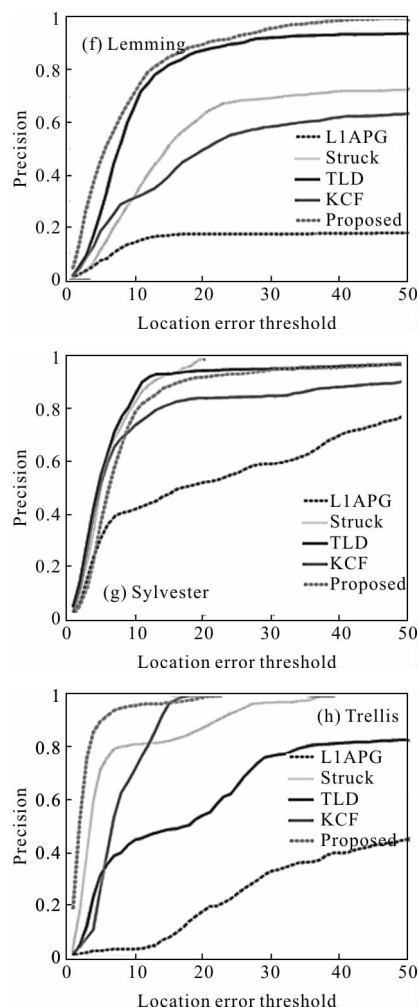
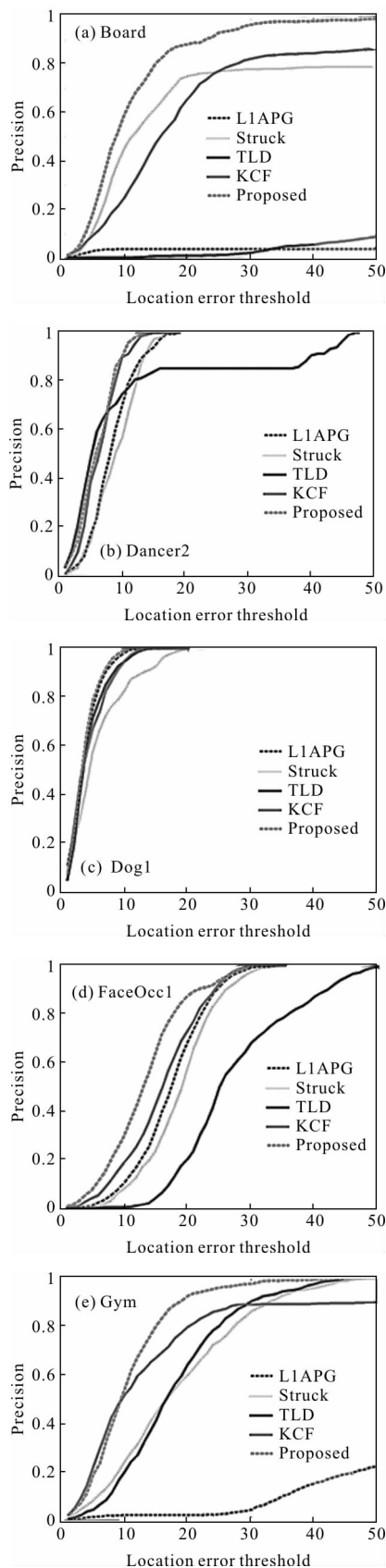


图 10 中心位置误差对比图

Fig.10 Comparison of center location error

5 结 论

文中针对目标跟踪中出现的目标形变及遮挡问题,提出了一种基于可变形模型的目标跟踪算法。该算法使用可变形模型表示目标,既能表征目标的局部和全局特征,又能表示出目标局部之间的空间位置关系。采用结构化输出的分类器能更加精确地将目标从背景中分离出来。从在公开的视频测试集上的实验结果来看,文中算法精度较高,尤其是在目标发生形变的情况下仍有较好的表现。

在文中算法的框架下,还可以进行扩展来进一步提升算法性能。如使用不同的特征来表示目标、尝试不同的核函数等。

参考文献:

[1] Mei X, Ling H. Robust visual tracking and vehicle

- classification via sparse representation [J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2011, 33(11): 2259–2272.
- [2] Ross D A, Lim J, Lin R S, et al. Incremental learning for robust visual tracking [J]. *International Journal of Computer Vision*, 2008, 77(1–3): 125–141.
- [3] Cui Xiongwen, Wu Qinzhong, Jiang Ping, et al. Affine-invariant target tracking based on subspace representation [J]. *Infrared and Laser Engineering*, 2015, 44(2): 769–774. (in Chinese)
崔雄文, 吴钦章, 蒋平, 等. 子空间模型下的仿射不变目标跟踪[J]. 红外与激光工程, 2015, 44(2): 769–774.
- [4] Grabner H, Bischof H. Online boosting and vision[C]// 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), 2006, 1: 260–267.
- [5] Kalal Z, Mikolajczyk K, Matas J. Tracking–learning–detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(7): 1409–1422.
- [6] Hare S, Golodetz S, Saffari A, et al. Struck: structured output tracking with kernels [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(10): 2096–2109..
- [7] 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.
- [8] Xing Yunlong, Li Aihua, Cui Zhigao, et al. Moving target tracking algorithm based on improved Kernelized correlation filter [J]. *Infrared and Laser Engineering*, 2016, 45(S1): S126004. (in Chinese)
邢运龙, 李艾华, 崔智高, 等. 改进核相关滤波的运动目标跟踪算法[J]. 红外与激光工程, 2016, 45(S1): S126004.
- [9] Felzenszwalb P, Grishick R, Mcallester D. Object detection with discriminatively trained part based models[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(9): 1627–1645.
- [10] Tschantzaris I, Joachims T, Hofmann T. Large margin methods for structured and interdependent output variables [J]. *Journal of Machine Learning Research*, 2005, 6: 1453–1484.
- [11] Bordes A, Bottou L, Gallinari P, et al. Solving multiclass support vector machines with LaRank [C]// International Conference on Machine Learning, ACM 2007: 89–96.
- [12] Bishop C M. Pattern Recognition and Machine Learning [M]. New York: Springer, 2006: 294–323.
- [13] Felzenszwalb P F, Huttenlocher D P. Distance transforms of sampled functions [J]. *Theory of Computing*, 2004, 19(8): 415–428.
- [14] Wu Y, Lim J, Yang M, et al. Object tracking benchmark [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1834–1848.
- [15] Bao C, Wu Y, Ling H, et al. Real time robust L1 tracker using accelerated proximal gradient approach[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2012: 1830 – 1837.