基于多分类器融合的多视角目标检测算法

尹维冲 路 通
(南京大学计算机软件新技术国家重点实验室 南京 210093)

摘 要 提出了从任意视角图像中检测视觉目标的新框架,并通过多视角分类器球面(Multi-View Detector Sphere, MVDS)模型对不同视角分类器之间的关系进行建模,以描述多个视角在识别视觉目标过程中的视角关联。首先,对不同视角的特定目标进行建模,对每个视角训练一个分类器;其次,通过将视角球面三角化,将视角球面均匀划分成若干三角面片,面片顶点所代表的视角之间的关系用以刻画不同视角分类器之间的关系。最后,对于来自未训练视角的目标,可通过融合球面上相邻视角分类器的输出给出其正确的检测结果。在多个公共数据集上的实验结果表明了该算法的有效性和准确性。

关键词 多视角,目标检测,MVDS
中图法分类号 TP391.4 **文献标识码** A

Novel Framework for Multi-view Object Detection through Combining Multiple Classifiers

YIN Wei-chong LU Tong
(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China)

Abstract We proposed a novel framework for detecting generic objects from arbitrary viewpoints described by varied object appearances. Our key insight is to exploit the multi-view detection patterns established by a number of detectors from different viewpoints and their relationships through the Multi-View Detector Sphere (MVDS), reflecting the underlying intrinsic structure for detecting multi-view objects. We first modeled the annotated objects from different viewpoints, and then triangulated the sphere into a number of uniformly distributed meshes to represent the explicit correspondences across view detectors. As a result, multi-view objects from untrained viewpoints can be detected by combining the outputs of the adjacent view detectors on the sphere. Our experiments on several public datasets give promising results for the experimental object classes.

Keywords Multi-view, Object detection, MVDS

1 引言

近年来,多视角目标检测在计算机视觉和多媒体领域受到了越来越多的关注[1-6],而检测的对象也由多视角人脸[2,3]和行人[4]扩展到一般的多视角目标[1,5]。该研究中最具挑战的问题之一是如何检测任意视角的目标类别。同类目标不同实例之间颜色、纹理等外观信息差别很大,而且由于多视角的影响,不同视角之间的同类目标差别也很大,因此,解决该类问题必须同时处理类内外观和观察视角变化的双重影响。目前,对同类目标外观的建模已有很多高效的算法,但对不同视角的同类目标间的视觉建模目前还缺乏研究,其主要原因是同类目标的不同视角之间底层特征较难找到有效关联,导致了相关算法的不精确性。同时考虑目标外观和多视角因素来建立多视角目标检测的算法模型更具挑战。

多视角目标类检测的算法大致可以分为两类:基于3D模型的方法和基于2D模型的方法。基于3D模型的方法通过学习训练数据来建立对该目标类的3D几何表示[5,6]。3D模型由于其对训练数据有一定要求(如训练数据中同一目标实例具有众多视角下的图像),且在训练和测试过程中需要较大的计算量,因此模型较为复杂,难以应用。基于2D模型[1,2,7,8]的方法为目标类的每个视角训练一个检测分类器,之后在测试阶段通过融合各个分类器的输出给出最终目标检测结果。不同的方法在融合的策略上有很大的不同,一些方法简单地选择分类器输出中置信最高的作为输出结果,但没有用到相邻视角的分类器所提供的检测信息,忽略了相邻视角之间的目标具有在外观和几何上的相似性这一内在联系。最近的一些方法意识到这个问题并对其进行了改进,如Liu等[7]提出了通过分类器插值生成新分类器的方法,解决了不同视角分类器的融合问题,但该算法是建立在"单个视角分类器可以表示成视角参数函数"这一假设条件上的,对实际应用而言这个假设很难满足。此外,Thomas等[1]通过建立不同视角的ISM(Implicit Shape Model)模型来实现对多视角目标类的检测;Razavi等[9]将单视角的ISM模型扩展到多视角检测,将多视角目标类简化为不同目标类来检测。

到稿日期:2012-09-30 返修日期:2012-12-22 本文受国家自然科学基金(61272218),教育部新世纪优秀人才支持计划(NCET-11-0232)资助。
尹维冲(1988-),男,硕士生,主要研究方向为计算机视觉,E-mail:weichong2153@163.com;路 通(1976-),男,博士,副教授,CCF高级会员,主要研究方向为多媒体技术与计算机图形学。

针对 3D 模型建立模型计算量较大，以及 2D 模型无法充分利用相邻视角分类器间约束的问题，本文提出了一个融合多视角分类器的目标检测新框架，其将不同视角的分类器组织在一个视角球面上，通过不同视角在视角球面上的位置关系来刻画分类器之间的关联，从而构建了多视角分类器球面模型(Multi-View Detector Sphere，MVDS)。在不同数据集上的多组实验证实了该模型的有效性。

2 多视角分类器球面模型

目标的任意视角由角度、高度和距离 3 个参数确定。当固定距离参数时，所有的视角组成了一个视角球面，视角球面上的每一个点对应于目标的某个特定视角。不同视角之间的关系可以通过在视角球面上的点的位置关系来确定；通过视角之间的位置约束，可以间接刻画出视角对应的分类器之间的关系，从而有效融合来自不同视角分类器的信息。

本文利用视角球面与每个视角对应的目标检测分类器来构建多视角分类器球面模型 MVDS(其中球面上的每个点与该点对应视角的分类器相关)。具体做法是：首先选取球面上的部分视角点作为关键视角点，这些关键视角点均匀地分布在整个视角球面上。接下来将视角球面三角化成离散的三角面片，这些三角面片不重复地均匀覆盖整个视角球面；每个三角形由 3 个关键视角点组成，覆盖了视角球面上的一定视角区域。接下来采用 Gentle Adaboost[10] 算法训练关键视角的分类器。最后，在进行多视角目标检测时，通过该模型融合各个关键视角分类器给出的输出，给出最终目标检测的结果。对于训练时未出现过的未知目标视角，模型可通过融合相邻视角分类器给出检测结果。此外，本文模型还避免了建立 3D 模型所需的复杂计算量，同时有效利用了不同视角分类器提供的信息。

本文模型及目标检测的流程如图 1 所示。首先，利用训练图像为每个关键视角训练一个 boosting 分类器。利用这些分类器构建出 MVDS 模型，由于训练时图像含有视角标注，因此可以手动地将关键视角分类器对应到视角球面上，之后将整个球面三角化，建立不同视角分类器之间的关联。之后对于输入的测试图像，根据分类器的特征选择，计算相应的特征，每个分类器对该图像可以计算出与原图像大小相同的置信图 CI(Confidence Image)，并利用所建立的球面模型对所有置信图进行三角融合，产生视角球面上每个三角形的置信图 TCI(Triangle Confidence Image)。最后根据这些三角形的置信图给出最终目标检测结果。

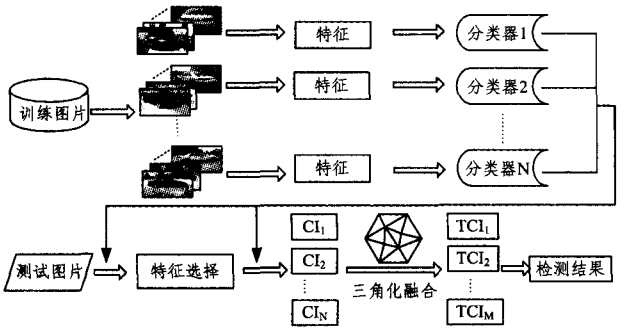

图 1 本文算法流程

3 训练单视角分类器

本节描述了对于给定的输入图像利用特定视角的分类器如何计算出对应的置信图(CI)。我们利用 Gentle AdaBoost 算法训练特定视角的分类器；对于输入图像中的每个像素位置，分类器给出该位置是否是目标中心的置信值。然后，通过该分类器给出对应的置信图。

3.1 特征提取

训练分类器所使用的目标外观采用变换角点特征[11] 来刻画，特征的生成依赖于通过如下步骤提取的模板库(template patches)。

1. 从训练集中随机选择 d 个训练图像，缩放图像，使得其中的目标大小基本相同；

2. 对每个选择的图像，使用 Canny 边缘检测算法获取位于目标包围盒内的 n 个角点；

3. 对于每个角点，提取以该角点为中心的图像区域作为一个模板 tp_f，区域的大小从预定义的范围中随机选择，同时记录该模板相对于目标中心位置的位置偏移 w_f。

由此得到的模板库的大小是 $D=d*n$。对于训练图像 I，通过以下步骤提取该图像中的特征向量：

1. 对于模板库中的每一模板 tp_f，计算 tp_f 和 I 之间的归一化相关度，由此得到一个和 I 尺寸相同的过程图像；

2. 将步骤 1 产生的过程图像和模板的位置偏移 w_f 作卷积。

至此，对于每一张图像 I 中的每一个像素点，计算出了一个 D 维的特征向量，其每一维的值反映了该像素点作为目标中心点的可能性。以上获取特征向量的步骤可以由以下公式给出。

$$v^f(x,y)=(I\otimes tp_f)*w_f \tag{1}$$

式中，$v^f(x,y)$ 代表位置 (x,y) 处的特征向量的第 f 维，$\otimes$ 代表归一化相关度计算，$*$ 代表卷积操作。

我们将训练图像目标中心位置的特征向量作为正样本，背景中的特征向量作为负样本，为每个视角训练一个分类器。训练阶段所有图像都缩放到目标大小基本相同，并计算这些目标包围盒的平均大小，将其作为该视角分类器包围盒的大小。

3.2 基于 boosting 的分类器

boosting 算法是通过结合多个弱分类器产生高准确率强分类器的有效方式。boosting 算法需不断拟合一个如下形式的增量模型[10]：

$$H(v)=\sum_{m=1}^{M}h_m(v,\gamma_m) \tag{2}$$

式中，v 是特征向量，M 是 boosting 次数，h_m 表示一个弱分类器，γ_m 代表其参数。拟合主要是优化以下形式的损失函数：

$$J=E[e^{-zH(v)}] \tag{3}$$

式中，z 表示样本 v 的真实值标签，$z\in\{+1,-1\}$，$H(v)$ 表示分类的结果。本文算法使用 Gentle AdaBoost[10]，该方法通过自适应牛顿迭代法将 J 的优化转化为最小化加权平方误差。每个弱分类器选择 $h_m(v,\gamma_m)=a\delta(v^f>\theta)+b$，$v^f$ 是特征向量

第 f 个维度的值，θ 是阈值，δ 是指示函数，a 和 b 是回归参数。每个弱分类器只选择特征向量的一个维度，这样最终的强分类器只用到了特征向量的 M 个维度，起到了降维的作用，大大减少了测试时的计算量。

给定了正负样本之后，本文利用该 boosting 算法训练出分类器，步骤如下。

1. 初始化样本的权值：

$w_i=1/N, i=1,2,\cdots,N, H(v)=0$

2. 对 $m=1,2,\cdots,M$，执行

a)通过最小化加权平方误差 J_{use} 选取一个弱分类器 $h_m(v,\gamma_m)$：

$$J_{use}=\sum_{i=1}^{N}\omega_i(z_i-h_m(v,\gamma_m))^2 \tag{4}$$

b)更新分类器：

$$H(v)\leftarrow H(v)+h_m(v,\gamma_m) \tag{5}$$

c)更新样本的权值：

$$\omega_i\leftarrow\omega_i e^{-z_i h_m(v_i,\gamma_m)}, i=1,2,\cdots,N \tag{6}$$

3. 输出最终的分类器。

与分类任务不同的是，对每一个测试样本，本文算法利用分类器输出的连续实数值，该值可以作为该样本是否是目标中心点的置信度量。

对于测试图像中的每一个像素，根据已训练分类器的特征选择，只需要计算特征向量的一部分维度，之后利用分类器给出置信值。由此对于测试图像，每一个特定视角的分类器可计算出其对应的置信图 CI。

4 多视角融合

给定各个视角分类器计算出的多张置信图，本节描述如何利用这些置信图计算出最终的目标检测结果，检测结果包括图像中目标出现的位置和大小，以及对该目标视角的估计。本文的主要思路是：对于未知视角的图像，它可能不与球面上的任何一个关键视角精确关联，但总是可将其对应到球面上某个三角形内，并通过融合三角形 3 个顶点分类器，来推测出未知视角下的检测结果。

给定输入图像 I 和多张置信图$\{CI_i\}$，通过如下步骤计算出特定球面三角形 t 的置信图 TCI_t。

1. 计算三角形 t 3 个顶点分类器的权值，(x,y) 表示位置：

$$\omega_i^t(x,y)=\frac{CI_i^t(x,y)}{\sum_{k=1}^{3}CI_k^t(x,y)} \tag{7}$$

2. 计算融合之后的球面三角形置信图：

$$TCI_t(x,y)=\sum_{i=1}^{3}\omega_i^t(x,y)CI_i^t(x,y) \tag{8}$$

考虑到未知视角距离三角形顶点越近，顶点分类器产生的置信就越高，在计算顶点分类器权值时，加大产生高置信的分类器的权值。在计算出视角球面上所有三角形的置信图之后，将所有三角形置信度中最大的作为检测结果：

$$FCI(x,y)=\max_t TCI_t(x,y) \tag{9}$$

至此，对输入图像中的每个像素点计算出了该点是否为目标中心点的置信度。将最终的置信图通过 parzen 窗口作卷积来平滑该置信图，并找出所有局部最大值点作为目标的中心点。利用预定义阈值 th 过滤这些局部最大值，即可得到最终的检测结果：

$$R(x,y)=FCI(x,y)>th \tag{10}$$

式中，th 为阈值，$R(x,y)\in\{0,1\}$ 表示最终的检测结果，即像素位置是否是目标中心点。对于目标中心点，目标的大小通过与分类器关联的包围盒的大小加权求得，同时可以估计目标的视角范围：

$$T(x,y)=\arg\max_t TCI(x,y) \tag{11}$$

5 实验结果与分析

本文在公共数据集 Leuven sports shoes 和 Leuven motorbikes[1] 数据集上进行了实验，并与其它方法进行了比较。在此基础上，进一步在以公共数据集 3D Objects[5] 为基础的扩展数据集上进行了实验，并将实验结果与最大值方法进行了比较。

5.1 实验 1：公共数据集

实验 1 在公共数据集 Leuven sports shoes 和 Leuven motorbikes[1] 上进行，使用与文献[1]相同的训练集和测试集。由于该数据集中目标视角的变化以及外观有较大差异，这两个类的数据集都非常有挑战性。

对于 sports shoes 类，训练数据包含 16 个视角，分布在两个高度(elevation)上，每个高度上 8 个视角。对于 motorbikes 类，训练库包含了 30 个不同的 motorbikes 实例。测试数据集与文献[1]相同。

训练过程中，对于每个视角，用于产生特征的模板库的大小 $D=400$，训练每个分类器的 boosting 次数 $M=60$。所有的训练图像的标签包括视角标签和目标包围盒。本文将目标缩放到 80×80 像素窗口中，并在多个尺度上进行检测，搜索步长为 1.05。评价结果时，将检测到的目标位置和大小与手工标定的正确值进行比较，使用和文献[1]一致的评价标准：多个检测结果包围盒重复 40%时仅取最大置信的结果，只有检测结果和真实结果包围盒重复 50%以上时才将其视作正确结果。

图 2 展示了本文算法和文献[1]算法在 sports shoes 类上的效果比较，使用的评价标准是目标检测时常用的 P/R(Precision/Recall)曲线。从图 2 可见，本文算法比文献[1]有明显改进。不过，正如文献[1]所述，该测试集中图像构成较为复杂和多变，因此这两种方法都尚未取得很好的效果。

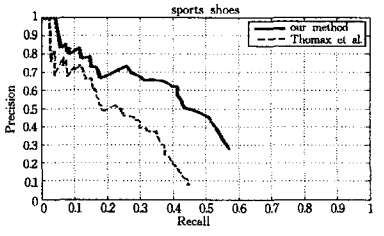

图 2 结果对比(sports shoes 类)

对于 motorbike 类，图 3 给出了本文方法与文献[1]方法

的实验效果对比，从中可见本文算法也取得了更高的平均检测率(Average Precision，AP)。更进一步，本文算法与在同样数据集和测试集上的其它几种方法做了对比，如表 1 所列。

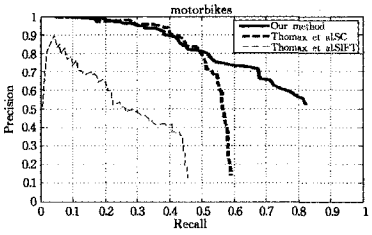

图 3　结果对比(motorbikes 类)

表 1　不同方法 AP 对比

方法	本文算法	Thomas 等[1]	Razavi 等[9]	Savarese 等[5]
AP	69.3%	53.1%	68.4%	75.1%

本文算法相比 Thomas 等[1]和 Razavi 等[9]的方法获得了更高的平均准确率，而 Savarese 等[5]的方法平均准确率高于其它方法，主要是因为后者在模型中建立了不同视角显式的几何关联，但是这种关联需要较高的计算量，且需要数据集中包含所有实例的所有视角的图像，这显然不适用于大的数据集，而本文的方法则没有这方面的限制。

图 4 展示了本文算法的一些检测结果，图中包围盒内为检测的结果，而包围盒的厚度表示该算法对结果的置信度。

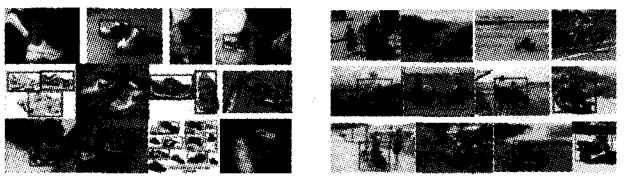

图 4　算法的部分检测结果

5.2　实验 2：扩展数据集

为进一步证实算法性能以及视角数量对算法的影响，本文在另外两个 bike 和 car 类上进行了进一步实验。使用的数据集基于 3D Objects[5]数据集并进行了扩展，对于该数据集的使用和文献[5]中一致，取 7 个实例的图像进行训练，另外取 3 个实例的图像进行测试。对于 car 类，选择了在 2 个高度上的 16 个视角数据，对于 bike 类，选择了 3 个高度上的 24 个视角的数据，这样，在 3D Objects 数据库中所有可用的图像对于每个视角只有 14 个，这对于训练一个 boosting 分类器来说是不足的。本文进一步采集了 259 张 car 类和 387 张 bike 类的图像，使得每一类的每一个视角平均约有 30 张图像。

实验表明，对于 car 类，本文算法获得了 71.4%的平均准确率，高于最大值方法的 59.0%。对于 bike 类，我们进一步探索了在建立模型时，使用的分类器数量对实验结果的影响情况，分别使用了 4、8、12、16、20 和 24 个视角分类器建立模型，并在相同的测试数据集上进行结果评价。图 5 给出了这一实验的结果，从中可见，随着建立模型所使用视角分类器数量的增加，结果的平均准确率在不断提高，当视角接近于致密覆盖整个视角球面之后，平均准确率基本保持不变。测试集中包含很多正侧面视角的图像，故只使用 8 个视角分类器的模型就可以获得很好的效果。

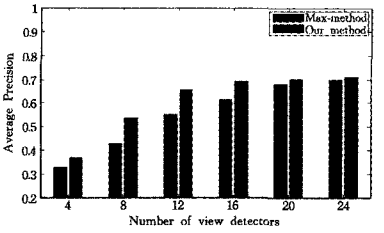

图 5　AP 随视角数量变化图

结束语　本文提出了一种新的多视角目标检测的方法，即通过融合多个视角分类器的输出给出最终的检测结果。与现有方法相比，本文方法易于实现、性能较好。进一步的工作包括自动建立 MVDS 模型。

参 考 文 献

[1] Thomas A, Ferrari V, Leibe B, et al. Toward multi-view object class detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Computer Society, 2006(2): 1589-1596

[2] Viola P, Jones M. Robust Real-Time Face Detection[J]. International Journal of Computer Vision, 2004, 57(2): 137-154

[3] 武勃，黄畅，艾海舟，等. 基于连续 AdaBoost 算法的多视角人脸检测[J]. 计算机研究与发展，2005，42(9)：1612-1621

[4] 徐剑，丁晓青，王生进，等. 多视角多行人目标检测、定位与对应算法[J]. 清华大学学报：自然科学版，2009，49(8)：1139-1143

[5] Savarese S, Li F. 3D generic object categorization, localization and pose estimation[C]// Proceedings of the IEEE International Conference on Computer Vision. Riode Janeiro: IEEE, 2007: 1-8

[6] Su H, Sun M, Li F, et al. Learning a dense multi-view representation for detection, viewpoint classification and synthesis of object categories[C]// Proceedings of the IEEE International Conference on Computer Vision. Kyoto: IEEE, 2009: 213-220

[7] Liu X, Gong H, Yan S, et al. Multi-view object detection by classifier interpolation[C]// Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. Dallas: IEEE, 2010: 826-829

[8] Wu B, Nevatia R. Cluster Boosted Tree Classifier for Multi-View, Multi-Pose Object Detection[C] // Proceedings of the IEEE International Conference on Computer Vision. Riode Janeiro: IEEE, 2007: 1-8

[9] Razavi N, Gall J, Van Gool L. Backprojection revisited: scalable multi-view object detection and similarity metrics for detections [C]// Proceedings of 11th European Conference on Computer Vision. Heraklion: Springer, 2010: 620-633

[10] Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting[J]. The Annals of Statistics, 2000, 28(2): 337-374

[11] Torralba A, Murphy K, Freeman W. Sharing visual features for multiclass and multiview object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(5): 854-869