

文章编号 1004-924X(2018)11-2776-09

多尺度显著性区域提取的模板匹配

逯睿琦, 马惠敏*

(清华大学 电子工程系, 北京 100084)

摘要:针对模板匹配过程中强遮挡、剧烈背景变化及物体非刚性形变等难题,本文提出了一种基于多尺度显著性区域提取的模板匹配算法。算法采用多尺度-显著性特征并行提取的方式:一方面利用空间金字塔模型将参考图像中的模板和待匹配图像中的目标区域分割成不同尺度的网格,采用可形变多相似性度量方法(Deformable Diversity Similarity, DDIS)计算不同尺度下的匹配得分;同时,算法提取模板区域的显著性区域图,形成模板区域的显著性得分;随后,利用显著性得分对不同尺度的匹配得分进行加权融合,在融合得到的匹配得分图上寻找最佳匹配区域。算法与取得目前最好结果的 DDIS 方法相比,AUC(Area Under Curve)指标提升 2.9%。实验结果表明,显著性区域提取使匹配方法更加关注目标物体,削弱背景及遮挡物体对其影响,从而增强模板匹配方法对于背景变化及遮挡的抵抗能力。另外,空间金字塔模型能够增强模板匹配方法对于物体不同尺度下的特征提取,如物体的局部轮廓及结构特征等。二者结合有效地提高了匹配精度。

关键词:模板匹配;可形变多相似性度量;显著性区域提取;空间金字塔

中图分类号:TP391.4 **文献标识码:**A **doi:**10.3788/OPE.20182611.2776

Template matching with multi-scale saliency

LU Rui-qi, MA Hui-min*

(Department of electronics, Tsinghua University, Beijing 100084, China)

* Corresponding author, E-mail: mhmpub@tsinghua.edu.cn

Abstract: Traditional template matching methods suffer from heavy occlusion, intense background change and non-rigid deformation. A multi-scale saliency template matching method is proposed in this article in order to deal with such conditions. The method extracted saliency and multi-scale features in parallel. On the one hand, the template and the target images were first divided into grids of different scales using spatial pyramid model. Deformable Diversity Similarity (DDIS) was calculated under such different grids. On the other hand, saliency map of the template image was calculated using saliency segmentation method. Such saliency maps are then used to weight the scores calculated by DDIS under different grids. Finally, the final score map is calculated by fusing the score maps under different grids. The method proposed achieves 2.9% AUC (Area Under Curve) improvement compared with original DDIS method. Experiments show that salient object segmentation helps the method to focus more on object than background, therefore improve the robustness to background changes and occlusion. Besides, spatial pyramid model makes the method to

收稿日期:2018-06-07;修订日期:2018-06-27.

基金项目:国家重点基础研究计划资助项目(No. 2016YFB0100900);国家自然科学基金资助项目(No. 61171113, No. 61773231)

consider information from different scale, for example, local contours and structural features of an object. Combining these two factors raises the matching accuracy significantly.

Key words: template matching; deformable diversity similarity; salient object segmentation; spatial pyramid model

1 引 言

模板匹配的准确性在物体检测、物体追踪和三维物体重建中起到非常关键的作用。一般来说,模板匹配任务给出参考(源)图像中的模板框,要求在待匹配(目标)图像中找到与模板框内的物体最接近的区域。在自然场景中,模板框内的物体通常会经历非常复杂的变化,例如背景变化、物体形变、遮挡变化等,给模板匹配带来较大的挑战。

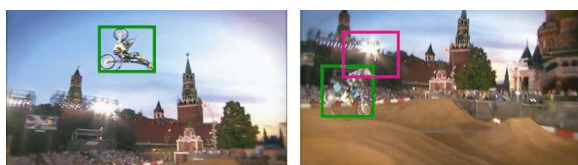
模板匹配方法的效果与其采用的相似性度量方法密切相关,传统的相似性度量一般为像素级别的,例如平方距离、绝对值距离或正规化的互相关等^[1],但当物体的背景出现变化或出现遮挡时会出现较大的误差。在这些度量方法的基础上,提出一些改进^[2-3]以对抗光照变化及噪声,但对于物体的形变则不能较好地处理。为了处理物体的形变问题,传统的方法^[4-5]通过构建物体的形变模型,希望利用给定的参数来描述物体的形变。但由于固定的参数通常只能描述一个种类的形变方式,例如仿射变换等,当物体经历任意的复杂形变时,上述方法则不能有效地对形变进行建模,从而不能正确地进行匹配。另外,传统的方法在考虑匹配时通常会利用模板图像内的所有像素点计算距离并进行匹配,这就使得当物体背后的背景发生改变时,背景区域的误差会发生任意的改变,从而可能严重影响匹配的准确性。为了改进这些缺点,一种建立在最近邻(Nearest-Neighbor)匹配之上的相似性度量方法(Best-Buddies-Similarity, BBS)^[6-7]被提出。其特点为只有当模板中的特征点和目标框中的特征点互为最近邻时,它们才被作为一个有效的匹配对,进而对相似度有贡献。在此方法的基础上,Talmi 等人^[8]进一步提出了可形变多相似性度量(DDIS),该方法与 BBS 相比,主要的特点有:1. 只考虑从目标到模板的单向匹配;2. 显式地考虑了物体的形变。由于只考虑了单向的匹配,因此 DDIS 的计算速

度更快。由于显式地考虑了物体的形变,因此 DDIS 能够对抗一些中等程度的形变,较 BBS 有提升。

本文着眼于自然场景下的模板匹配任务,在自然场景下,由于缺乏先验信息,模板物体通常会经历复杂的变化,例如物体的背景会改变,如图 1(a),物体的受遮挡面积改变,如图 1(b)等。尽管 DDIS 有一定的对抗背景变化以及物体形变的能力,但是当模板中的背景所占比重较大时,对于 DDIS 的结果还是会有较大的影响,如图 1(a)和图 1(b)所示。(a)中由于模板中天空背景占据了较大的面积,因此当物体的背景已经完全改变后,匹配到的目标区域偏向包含了天空背景的区域,而(b)中由于模板中物体基本被树叶遮挡,因此匹配到了树叶而不是杯子。本文观察到,尽管整张图片中存在多种背景及物体,但是在模板附近的区域中,前景物体与背景的对比如较高,因此本文提出使用显著性物体分割对模板附近的区域进行处理,形成显著性区域图,使用显著性区域图对 DDIS 的检测结果进行加权,从而可以抑制背景区域对于相似度的贡献。邵枫^[9]等人也在其文中也有相似的做法,他们提出了一种基于显著性分析的立体图像视觉舒适度检测方法,利用人眼的视觉注意力机制和协方差矩阵得到图像显著图以建立客观图像与主观评价之间的联系。

另外,DDIS 虽然在计算相似度时只依赖存在有效匹配的一个子集点,但其仍然是像素级别的度量方法,即其缺乏对于物体局部结构及轮廓的多尺度的考虑,从而导致匹配的结果与模板中的物体可能有较大程度的偏差,如图 1(c)所示。图(c)中尽管目标物体在两张图像之间几乎没有任何变化,但由于目标内部纹理不明显,因此必须更多的依靠轮廓信息进行匹配,而 DDIS 则未能考虑物体的轮廓信息,导致匹配失败。受到空间金子塔模型的启发,本文将模板图像划分为不同尺度的网格,在网格的格点级别上进行相似度计算,并与像素级别的相似度进行融合,使得匹配时能够多尺度地考虑目标与模板的相似度。

综上,本文综合利用了显著性区域提取以及空间金字塔模型对于目前取得最好结果的 DDIS 方法进行提升,提出了基于多尺度显著性区域提取的模板匹配方法,使得方法能够有效地对抗背景变化、复杂遮挡及物体形变等,进一步提升了 DDIS 的匹配精度。



(a) 物体背景完全改变

(a) Significant background changes of object



(b) 物体被大面积遮挡

(b) Large occlusion of object



(c) 物体内部纹理不明显

(c) Unobvious interior texture

图 1 DDIS 的错误匹配结果

Fig. 1 Error matching results of DDIS

2 算法框架

多尺度显著性区域提取模型的算法框架如图 2 所示。该框架综合利用了显著性分割以及空间金字塔模型,能够有效地对抗背景变化、物体遮挡以及物体形变等复杂情况。首先,图(a)模板经过显著性分割得到模板的显著性得分图,显著性得分图的大小与模板相同,为灰度图。像素的灰度值越大则代表该像素点更有可能属于物体,则在匹配时该像素对于相似度的贡献越大;图(b)模板被分割为不同尺度的网格,分别在不同尺度下在目标图像上进行 DDIS 匹配。同时图(a)中得到的显著性得分图也被分割为相应的网格,并对于匹配得分进行加权,得到不同尺度下的匹配得分图;图(c)将不同尺度下得到的匹配得分图进行

融合,得到最终的匹配结果(绿色为真值框,红色为匹配结果)。下面第 2 章将阐述如何使用显著性分割结果对于 DDIS 结果进行加权,第 3 章将阐述如何在不同尺度下进行 DDIS 匹配以及如何融合不同尺度的匹配得分。

3 基于显著性提取的 DDIS 方法

本章基于原始的 DDIS 方法进行提升,提出了基于显著性提取的 DDIS 方法,表示为 sDDIS。

原始的 DDIS 相似性度量包括外观相似度和形变相似度两个部分,并且待匹配图像中的目标框的大小和参考图像中的模板框大小相同。定义 $p_i, q_j \in R^d$ 分别代表模板框和目标框内的像素点的特征,DDIS 的目标是度量模板框点集 $P = \{p_i\}_{i=1}^N$ 和目标框点集 $Q = \{q_j\}_{j=1}^M$ 的相似度。用符号 a 表示外观(appearance),用符号 l 表示位置,则令 p^a 代表点 p 的外观属性, p^l 代表点 p 的位置属性(对目标框内的点 q 定义相同)。

对于目标框中的任意一点 $q_j \in Q$,定义 $p_i = NN^a(q_j, P) = \underset{p \in P}{\operatorname{argmin}} d(p^a, q_j^a)$,其中 $d(p^a, q^a)$ 为某种距离函数,如欧氏距离。该式表示模板框内的点 p_i 为目标框内的点 q_j 在外观属性上的最近邻,即 NN^a 。因为模板框内的点 p_i 可能是多个目标框内的点的最近邻,因此定义 $\kappa(p_i)$ 为目标框内最近邻是 p_i 的点的总数,即:

$$\kappa(p_i) = |\{q \in Q \mid NN^a(q, P) = p_i\}|. \quad (1)$$

另外,定义 q_j 与 $p_i = NN^a(q_j, P)$ 在框内的位置距离为 $r_j = d(q_j^l, p_i^l)$,距离越大,代表由 p_i 到 q_j 的形变程度越大。分别定义外观距离以及位置距离后,可以最终得到 DDIS 的解析形式,如式(2):

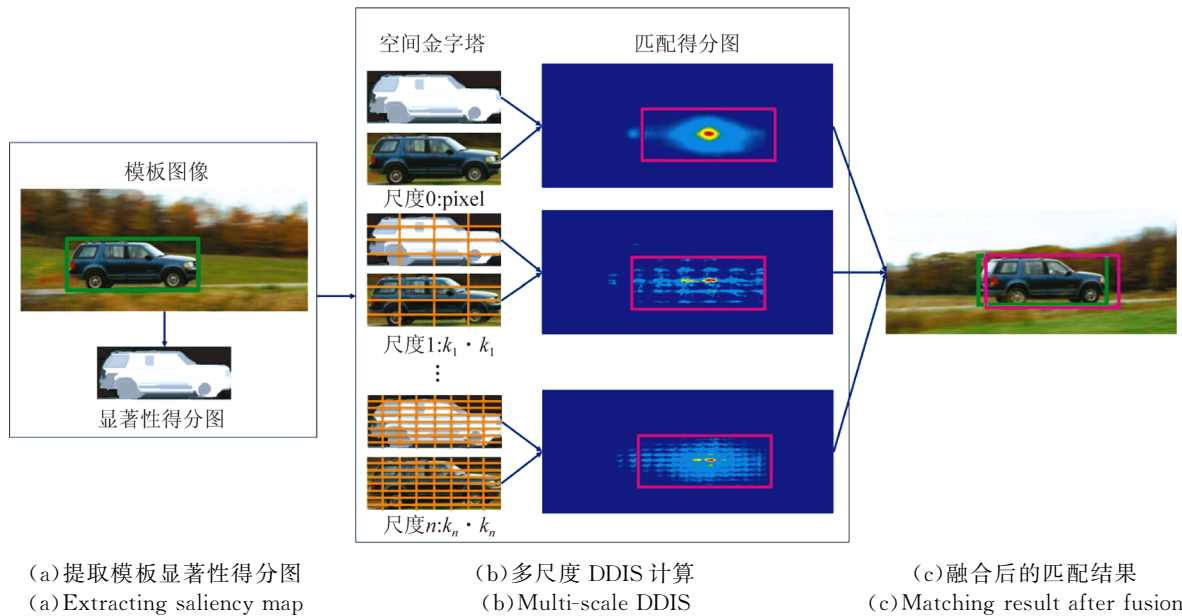
$$DDIS_{Q \rightarrow P} = c \sum_{q_j \in Q} \frac{1}{r_j + 1} \cdot \exp(1 - \kappa(NN^a(q_j, P))), \quad (2)$$

其中: $c = 1/\min\{M, N\}$ 为正则项,当目标框与模板框大小相同时, $M = N$ 。上式可以看作对目标框内的所有像素点 q_j 的贡献的求和。特别的,当 q_j 的最近邻是独一无二的时候,即 $\kappa(NN^a(q_j, P)) = 1$ 时,指数项达到最大值 1。当 q_j 的 NN^a 被其他的点共享时 $\kappa(NN^a(q_j, P)) > 1$,则指数项较小, q_j 的贡献也较小。另外, q_j 与其 NN^a 的空间位置距离越大,那么 r_j 越大,其贡献也越小。因此,DDIS 有如下几个性质:

- (1)它主要依赖于具有独特 NN^a 的目标点, 共享 NN^a 的目标点对于相似度的影响较低;
- (2)DDIS 不需要对形变有任何先验知识和

物理建模;

- (3)DDIS 建立在最近邻匹配之上, 而并非直接使用距离来度量相似度。



(a)提取模板显著性得分图
(a)Extracting saliency map

(b)多尺度 DDIS 计算
(b)Multi-scale DDIS

(c)融合后的匹配结果
(c)Matching result after fusion

图 2 算法框架
Fig. 2 Framework

DDIS 根据目标框内像素点与其最近邻的位置距离以及其最近邻被共享的程度决定该像素点对于相似度的贡献程度。事实上, 存在另外一个重要的因素, 即目标框内的像素点匹配到的最近邻的似物程度。当其最近邻属于物体的一部分时, 应当赋予高的权重, 反之则应赋予低权重。而图像区域的似物程度已经在图像显著性物体分割任务中有较好的研究, 因此本文采用显著性物体分割对模板图像进行处理, 并使用得到的显著性得分来对 DDIS 的结果进行进一步的加权。

图像显著性物体分割是一种旨在提取人类在识别图像中物体时的认知规律, 基于人的认知方法, 通过算法实现图像中“被关注物体”的语义分割。本文使用文献[10]提出的融合多区域显著图的显著性分割方法来进行显著性区域提取。该方法在不同的尺度上提取了区域, 对比区域属性和区域背景特征, 利用这些特征计算不同尺度的似物图, 并将不同尺度的似物图进行融合从而得到最终的似物图, 由该方法得到的一些似物图展示在图 3 中。由该方法得到的显著性得分图为灰度图, 灰度值越高代表越有可能是物体, 本文对得到

的灰度图进行线性归一化, 得到模板框内的显著性得分图。具体来说, 对于模板中的任意一点 p_i , 其显著性得分表示为 $s(p_i)$ 。

本文使用的显著性分割方法需要应用于背景比较单一或前景物体相对集中的图像上。而对于模板匹配来说, 整张图像的背景可能较为复杂, 前景中的物体也较多, 因此直接对全图进行显著性分割不会得到较好的分割结果。本文观察到, 模板框附近的区域背景通常较为单一, 能够很好的满足显著性分割的要求, 因此本文只对模板框附近的图像区域进行分割。具体做法为, 将模板框的大小扩大一倍, 在扩大后的图像区域上进行显著性分割, 从分割后的图像中再截取出模板部分作为模板的显著性得分图。对于显著性得分图进行归一化, 得到模板中每一个像素的似物性值。DDIS 方法中, 目标框内每一个像素点对于相似度的贡献与其在模板框内的最邻近点的受共享程度以及二者的位置差有关, 在加入显著性后, 其对相似度的贡献还要由显著性得分进行加权。加入显著性后的 DDIS 相似度定义为:

$$DDIS_{\text{salQ} \rightarrow P} = c \sum_{q_j \in Q} \frac{1}{r_j + 1} \cdot \exp(1 - \kappa(\text{NN}^a(q_j, P))) \cdot s(\text{NN}^a(q_j, P)), \quad (3)$$

这样,匹配到背景的目标框像素点对于相似度的贡献被抑制,从而提高 DDIS 对于背景变化及遮挡变化的鲁棒性。



图 3 显著性分割结果

Fig. 3 Saliency segmentation results

4 多尺度匹配及融合方法

本文使用空间金字塔模型进行多尺度匹配。空间金字塔模型^[11]是用于改进 bag-of-features^[12-15]的方法。其将图像分割成不同尺度的网格,在每一个格点中分别进行特征提取,从而加入了特征分布的空间信息。多尺度的特征提取被广泛用于其他任务中,例如 spatial pyramid pooling 方法^[16],使得在进行物体检测时一张图像只需要经过一次卷积网络就可以得到所有候选框的卷积特征,提升了识别速度。秦翰林^[17]等人提出多尺度的隐马尔科夫模型,利用剪切波变换提取各尺度的特征。王玮^[18]等人使用多尺度 LBP 特征描述子进行人脸识别任务等。

在原始的 DDIS 方法中,每一个匹配单位 q_j 均为像素点,这种匹配方式忽略了物体的局部轮廓及结构信息,导致有些结果明显地缺乏物体的局部轮廓和结构,如图 1(c)所示。因此,本文考虑以格点作为匹配单位的方法,即将 q_j 由像素点变为网格格点。将模板和目标框均分割成不同尺度的网格,在不同的尺度上进行相似度的度量,随后融合不同尺度的相似度。

假设 $DDIS_{\text{pixel}}$ 为像素尺度下的匹配得分,不同网格尺度下的匹配得分分别为 $DDIS_{k_1 \cdot k_1}, DDIS_{k_2 \cdot k_2}, \dots, DDIS_{k_n \cdot k_n}$, 其中 $k_i \cdot k_i$ 代表将模板和目标框划分为

$k_i \cdot k_i$ 的网格, n 为尺度总数。每个尺度下的匹配得分均由式(3)计算得到。在不同尺度下 q_j 所代表的意义不同,例如对于 $DDIS_{\text{pixel}}$, q_j 代表一个像素点,对于 $DDIS_{k_1 \cdot k_1}$, q_j 则代表 $k_1 \cdot k_1$ 个格点中的一个格点。融合后的匹配得分的计算方式为:

$$DDIS_{\text{fusion}} = \exp \left\{ \log(DDIS_{\text{pixel}} + 1) + \frac{\log(DDIS + 1)}{\text{weight}(k_1)} + \frac{\log(DDIS + 1)}{\text{weight}(k_2)} + \dots + \frac{\log(DDIS + 1)}{\text{weight}(k_n)} \right\}. \quad (4)$$

其中 $\text{weight}(k_i)$ 为各个尺度下的权重系数,该系数越大,则该尺度对于最终的融合结果影响越小。采用非线性融合方式,加 1 操作使得对数结果不小于 0,从而指数结果能够归一化。本文中网格尺度以及权重系数均为经过实验后人工选定的。

5 实验

5.1 数据集

本文使用的数据集是由用于物体跟踪任务的数据集 OTB(Object Tracking Benchmark)^[19]生成的,该数据集为彩色视频序列,其中目标物体种类较多,包含人、动物、车辆和日用品等等,场景也较复杂。目标物体一般会经历形变、旋转、遮挡和大小变化等,具有较大的难度。为便于比较,我们使用 DDIS 方法^[6]中生成的数据,其做法为从 OTB 中挑选出 35 个彩色视频序列,并从每个彩色视频序列中随机采集 3 对图像,这样共有 105 对模板匹配图像,每对图像之间的帧距为 20。

每一对图像中,第一张图像的包围框用于确定模板内容,即真值框,第二张是待匹配图像。模板匹配的精度度量方式为交并比,即:

$$\text{accuracy} = \frac{|R_{\text{est}} \cap R_{\text{cruth}}|}{|R_{\text{est}} \cup R_{\text{cruth}}|}, \quad (5)$$

其中 R_{est} 为估计的矩形框, R_{cruth} 为真值框。

表 1 不同方法的 AUC 结果

Tab. 1 AUC results of different model

方法	AUC
DDIS(GT)	0.640
sDDIS(ours)	0.655
DDIS-sp(ours)	0.660
sDDIS-sp(ours)	0.669

5.2 定量评价

首先,对于显著性区域提取以及多尺度匹配两个部分,比较单独使用某个部分以及将两者结合使用的性能。进行实验的方法有:原始 DDIS; 基于显著性区域提取的 DDIS(sDDIS),即使用像素级匹配以及显著性得分图;多尺度 DDIS(DDIS-sp),即使用多尺度匹配但不使用显著性得分图加权;多尺度显著性区域提取 DDIS(sDDIS-sp),即将多尺度匹配与显著性得分图结合。共 4 种方法。对于 DDIS 及 sDDIS,本文使用了和文献^[8]中相同的特征提取方式,将以每个像素为中心的 3×3 切片的 RGB 值向量化作为该像素的特征,切片之间相互重叠。对于 DDIS-sp 和 sDDIS-sp,本文使用了 5×5 与 10×10 两个网格尺度,并均将权重设置为 1,每个格点的特征为

格点内所有像素点的 RGB 值向量化后的结果。对于 sDDIS-sp,由于原始显著性得分图给每个像素都计算了分数,所以本文使用每个格点内部所有像素的显著性分数的均值作为该格点的显著性分数,对于网格尺度的匹配得分进行加权。

在采集的数据集上进行测试,并对每种方法做出 ROC 曲线如图 4。ROC 曲线代表了数据集中匹配精度大于阈值的样本的比例。其中(a)~(c)依次对比了 DDIS 和 sDDIS;DDIS 和 DDIS-sp;DDIS, sDDIS 和 sDDIS-sp。AUC 结果如表 1,AUC 为 ROC 曲线下的面积,代表了整体的正确率。可以看出,显著性区域提取的 DDIS 和多尺度 DDIS 较原始的 DDIS 度量方式均有所提高,而二者合并使用则能够进一步提高匹配准确率。

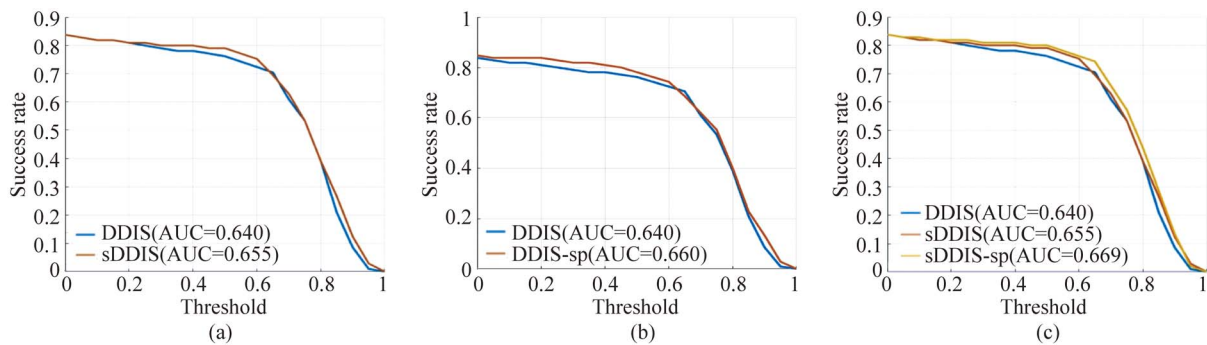


图 4 匹配正确率

Fig. 4 Template matching accuracy

随后,比较 sDDIS-sp 与目前存在的其他模板匹配算法的性能。进行比较的方法有:平方误差和(SSD),绝对值误差和(SAD),归一化相关匹配(NCC),颜色直方图匹配(HM),双向性相似度(BDS)^[20]。表 2 给出了各种方法对应的 AUC 结果。sDDIS-sp 的性能大幅度高于传统的模板匹配方法。

为了研究空间金字塔模型中尺度的大小以及数量对于匹配结果的影响,本文以 DDIS-sp 为例,对不同数量以及大小的金字塔模型进行实验,AUC 结果如表 3。由前 3 行可知,当网格尺度适中时匹配精度能够得到明显的提升,而过于粗略和细致的网格则不能带来明显地提升。这是由于如果网格过于粗略,那么 DDIS 的多样性就难以体现;而如果网格过于精细,那么就与像素级别的度量基本没有区别,仍然无法得到物体的轮廓及

结构信息。而对比后两行的结果,可知在不加入显著性的条件下,尺度的数量增加对于准确度的提升并不明显。

表 2 不同方法的 AUC 结果

Tab. 2 AUC results of different methods

方法	AUC
SSD	0.43
SAD	0.49
NCC	0.47
HM	0.41
BDS	0.50
sDDIS-sp(ours)	0.67

表 3 不同金字塔模型参数下的 AUC 结果

Tab. 3 AUC results under different grid levels

金字塔模型	AUC
pixel+5×5	0.658
pixel+8×8	0.630
pixel+10×10	0.633
pixel+5×5+10×10	0.660
pixel+8×8+10×10	0.631

5.3 定性评价

图 5(彩图见期刊电子版)展示了几组模板匹配结果,包含了遮挡,形变等复杂条件。由左至

右,(a)列绿色框标注出了模板内容,(b)列展示出了模板匹配结果,其中绿色框为真值,红色框为原始 DDIS 方法得到的结果,黄色框为 sDDIS-sp(见 4.2)方法得到的结果。(c)和(d)列分别为两种方法对应的匹配得分图。从上至下分析可以得到,第一行和第三行中 DDIS 未能考虑物体的局部轮廓及结构信息,因此匹配到了杂乱的背景区域;第四行中 DDIS 未能区分模板中的前景和背景物体,因此匹配到了背景;第三行中由于背景区域仍然较大,因此导致 DDIS 的结果偏向于背景区域。可以看出,本文的方法在对抗背景及考虑物体局部轮廓及结构信息上有更好的精确度。

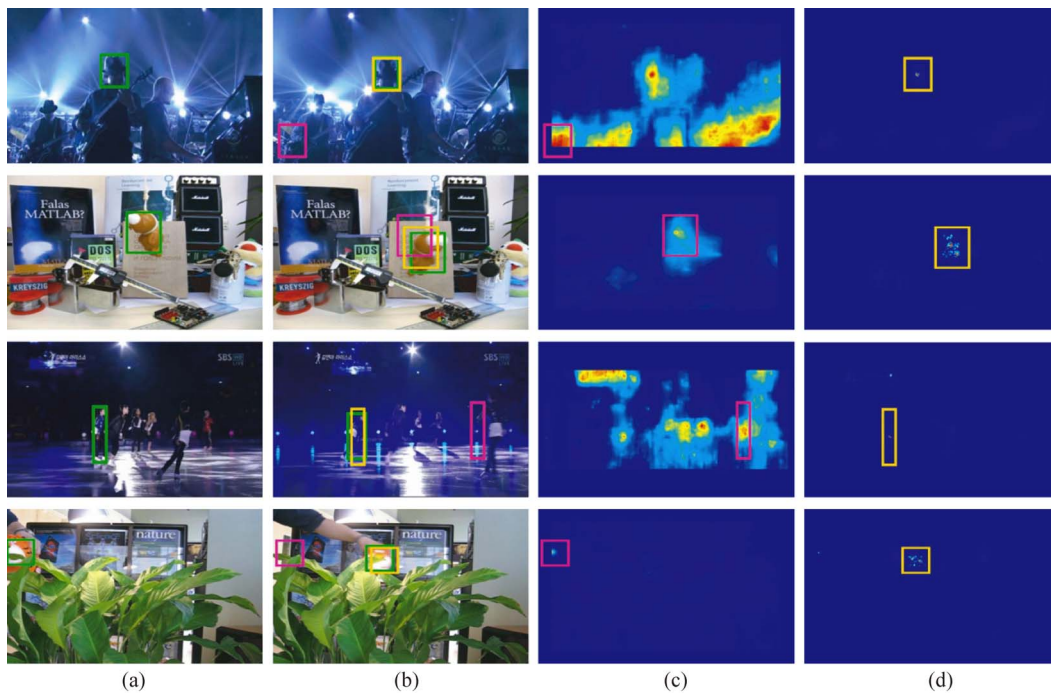


图 5 定性分析

Fig. 5 Qualitative analysis

6 结 论

本文分析了使用 DDIS 方法进行模板匹配时失败的原因有背景区域面积过大、存在较强遮挡以及未能利用物体轮廓信息等。为了改进 DDIS 方法的有效性,本文提出了多尺度显著性区域提取模型。该模型利用显著性区域分割方法提取模板中的显著区域,使用显著性得分图对 DDIS 方法产生的匹配得分图进行加权,从而抑制背景对于相似度的贡献。同时为了使模板匹配方法能够

利用多尺度信息,本文利用空间金字塔模型,将模板分割成不同尺度的网格,在不同尺度的网格下分别进行 DDIS 匹配,并融合多尺度的匹配得分图进而得到最终的匹配结果。多尺度显著性匹配模型与取得目前最好结果的 DDIS 方法相比,AUC 指标提升 2.9%。结果表明,显著性区域提取可以有效地增强模板匹配方法对于背景变化的抵抗能力,通过空间金字塔模型能够使模型获得物体的局部轮廓信息,从而本文提出的多尺度显著性区域提取模型在原始的 DDIS 相似性度量方法上有了明显的提升。

参考文献:

- [1] OUYANG W, TOMBARI F, MATTOCCIA S, *et al.*. Performance evaluation of full search equivalent pattern matching algorithms[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2011, 34(1):127-143.
- [2] ELBOHER E, WERMAN M. Asymmetric correlation: a noise robust similarity measure for template matching [J]. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 2013, 22(8):3062-3073.
- [3] HEL-OR Y, HEL-OR H, DAVID E. Matching by tone mapping: photometric invariant template matching [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2013, 36(2):317-330.
- [4] KORMAN S, REICHMAN D, TSUR G, *et al.*. FasT-match: fast affine template matching [C]. *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2013:2331-2338.
- [5] TIAN Y, NARASIMHAN S G. Globally optimal estimation of nonrigid image distortion [J]. *International Journal of Computer Vision*, 2012, 98(3):279-302.
- [6] DEKEL T, ORON S, RUBINSTEIN M, *et al.*. Best-buddies similarity for robust template matching [C]. *Computer Vision and Pattern Recognition*. IEEE, 2015:2021-2029.
- [7] ORON S, DEKEL T, XUE T, *et al.*. Best-buddies similarity-robust template matching using mutual nearest neighbors [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-14.
- [8] TALMI I, MECHREZ R, ZELNIK-MANOR L. Template matching with deformable diversity similarity [C]. *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017:1311-1319.
- [9] 邵枫,姜求平,蒋刚毅,等. 基于显著性分析的立体图像视觉舒适度预测 [J]. *光学精密工程*, 2014, 22(6):1631-1638.
- SHAO F, JIANG Q P, JIANG G Y, *et al.*. Prediction of visual discomfort of stereoscopic images based on saliency analysis [J]. *Opt. Precision Eng.*, 2014, 22(6):1631-1638. (in Chinese)
- [10] JIANG H, WANG J, YUAN Z, *et al.*. Salient object detection: a discriminative regional feature integration approach [C]. *Computer Vision and Pattern Recognition*. IEEE, 2013:2083-2090.
- [11] LAZEBNIK S, SCHMID C, PONCE J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories [C]. *Computer Vision and Pattern Recognition*, 2006 IEEE Computer Society Conference on. IEEE, 2006: 2169-2178.
- [12] GRAUMAN K, DARRELL T. Discriminative classification with sets of image features [C]. *International Conference on Computer Vision*, 2005.
- [13] WALLRAVEN C, CAPUTO B, GRAF A. Recognition with local features: the kernel recipe [C]. *IEEE International Conference on Computer Vision*. IEEE Computer Society, 2003:257.
- [14] WILLAMOWSKI J, ARREGUI D, CSURKA G, *et al.*. Categorizing nine visual classes using local appearance descriptors [J]. *Icpr Workshop on Learning for Adaptable Visual Systems*, 2004.
- [15] ZHANG J, MARSZALEK M, LAZEBNIK S, *et al.*. Local features and kernels for classification of texture and object categories: An in-depth study [R]. *Technical Report RR-5737*, INRIA Rhône-Alpes, 2005.
- [16] HE K, ZHANG X, REN S, *et al.*. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2014, 37(9): 1904-1916.
- [17] 秦翰林,周慧鑫,刘群昌,等. 采用多尺度隐式马尔可夫模型的红外图像背景抑制 [J]. *光学精密工程*, 2011, 19(8): 1950-1956.
- QIN H L, ZHOU H X, LIU Q CH, *et al.*. Suppression of infrared image background by multi-scale hidden Markov model [J]. *Opt. Precision Eng.*, 2011, 19(8): 1950-1956. (in Chinese)
- [18] 王玮,黄非非,李见为,等. 使用多尺度 LBP 特征描述与识别人脸 [J]. *光学精密工程*, 2008, 16(4):696-705.
- WANG W, HUANG F F, LI J W, *et al.*. Face description and recognition using multi-scale LBP feature [J]. *Opt. Precision Eng.*, 2008, 16(4): 696-705. (in Chinese)

- [19] WU Y, LIM J, YANG M H. Online object tracking: a benchmark [C]. *Computer Vision and Pattern Recognition. IEEE*, 2013:2411-2418.
- [20] SIMAKOV D, CASPI Y, SHECHTMAN E, *et al.*. Summarizing visual data using bidirectional similarity [C]. *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE*, 2008:1-8.

作者简介:

逯睿琦(1995—),男,山西临汾人,学士,2017年于清华大学获得学士学位,主要从事数字图像处理,物体识别与检测等方面的研究。E-mail: lurq17@mails. tsinghua. edu. cn

导师简介:

马惠敏(1972—),女,河南洛阳人,博士,副教授,博士生导师,2001年于北京理工大学获得博士学位,主要从事三维图像认知与仿真,视觉导航,智能无人系统等方面的研究。E-mail: mhmpub@tsinghua. edu. cn