

大量类别下非纹理对象实时检测与识别

陶 军 刘建明 王明文 万剑怡

(江西师范大学计算机信息工程学院 南昌 330022)

摘 要 现有的对象检测方法主要针对特定对象,当类别比较多时,难以实现实时检测与识别。提出了一种基于 Objectness 和梯度方向模板的大量类别下非纹理对象的实时检测与识别算法。该方法首先通过计算图像 Objectness 值来评价待测图像中可能出现对象的区域,大量减少可能匹配的窗口。在此基础上,在可能出现对象的区域,采用基于模板主方向和查找表的模板匹配方法,实现大量类别下非纹理对象的实时检测与识别。该方法对非纹理物体的鲁棒性较好,同时在匹配的过程中也是方向无关的。

关键词 非纹理,模板匹配,鲁棒性,方向无关

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.10.065

Real-time Detection and Recognition for Large Numbers of Less-texture Objects

TAO Jun LIU Jian-ming WANG Ming-wen WAN Jian-yi

(College of Computer Information Engineering, Jiangxi Normal University, Nanchang 330022, China)

Abstract The existing objects detection methods can not achieve real-time detection and identification when the object classes are too many. To solve the problem, a real-time detection and recognition algorithm on many classes and texture-less objects was put forward. The new algorithm is based on Objectness and gradient direction template. Firstly, it evaluates the potential objects by computing its Objectness value, which can decrease many matching windows. Then in the area where the objects may appear, it detects and recognizes the texture-less objects in many classes using the template matching method based on the main direction of template and lookup table. The robustness of this algorithm to the texture-less object is better. And the algorithm is orientation independent in the process of matching.

Keywords Less-texture, Template matching, Robustness, Orientation independent

1 引言

在非纹理对象的识别方面,由于传统的基于统计学习技术的识别算法需要非常多的训练样本且复杂度往往很高,难以满足实时要求,因此基于模板实时匹配的识别算法依然是机器人识别中最吸引人的解决方案。模板匹配在检测跟踪应用中一直扮演重要角色。模板匹配不需要费时的训练阶段也能处理非纹理对象(这些对象往往难以采用基于特征点的方法进行检测)。

早期的模板算法采用了模板和图像轮廓间的 Chamfer 距离^[1]及其扩展^[2]来进行识别。Chamfer 距离的计算能够通过 DT 变换(Distance Transform)快速实现。文献[3]的方法基于 DT 变换,但它能够在满足实时匹配下实现缩小无关和透视变形无关。然而,这种算法要求对象轮廓必须是闭合的曲线,在时间应用中往往难以满足要求。所有基于 DT 变换的

方法都需要获取对象轮廓点,而其往往很难被准确获取,对象轮廓点的获取对光照的变化、噪声以及模糊等非常敏感。针对这个问题,文献[4]提出了一种能够克服上述问题的方法,通过考虑图像的相对梯度,计算模板梯度和图像梯度的点积作为相似度匹配。然而,这种方法对对象形状的细微变形非常敏感。梯度直方图(HOG)是另外一种非常流行的方法^[5],但是计算复杂度比较高。Ferrari 等^[6]提出了一种基于学习的对象识别算法。然而,这种方法的复杂度太高,无法满足实时的要求。文献[7]通过临近线段组来构建对象的局部表达(BOLD),具有旋转、平移以及缩放不变性。为了兼顾实时性,Stefan Hinterstoisser 教授^[8]提出了梯度响应图的方法,通过查找表找出相似性最高的区域来实现实时匹配。然而,该方法依然需要采用滑动窗口的方法逐个匹配,滑动窗口方法中测试对象可能存在的窗口区域往往有数百万个,当图像精度比较高且待检测识别的类别比较多时,难以达到实时性要求。

到稿日期:2014-06-05 返修日期:2014-07-19 本文受江西省教育厅青年科学基金项目:基于稀疏编码的艺术图像语义分类技术研究(GJJ12213),国家自然科学基金项目:大规模数据聚类的并行进化算法骨架研究(61163006),江西师范大学青年教师成长基金:基于稀疏编码的图像语义分类技术研究(3919)资助。

陶 军(1987-),男,硕士生,主要研究方向为图像处理,E-mail:bhtj727@126.com;刘建明(1981-),博士,讲师,主要研究方向为图像处理与模式识别、大数据与深度学习、文物数字化保护等,E-mail:stoneliu1981@qq.com(通信作者);王明文(1964-),男,博士,教授,主要研究方向为信息检索、数据挖掘、机器学习,E-mail:mwwang@jxnu.edu.cn;万剑怡(1974-),女,博士,教授,主要研究方向为并行分布式算法、大规模计算研究,E-mail:wanyany@yaho.com.cn。

本文在文献[8]方法的基础上,结合最新的任意对象可能区域的检测算法^[9],首先通过计算图像 Objectness 值来评价待测图像中可能出现对象的区域,以大量减少可能匹配的窗口。在此基础上,在可能出现对象的区域,采用基于模板主方向和查找表的模板匹配方法,实现大量类别下非纹理对象的实时检测与识别。

2 算法实现

首先通过 Objectness 算法^[9]实现对象区域的粗检测,从而得到可能存在对象的区域。在此基础上,利用基于梯度方向模板的快速匹配算法实现对象的检测与识别。匹配流程如图 1 所示。

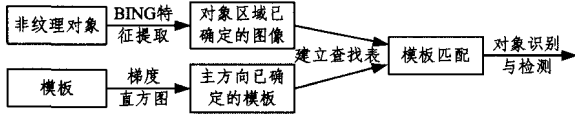


图 1 匹配流程

2.1 对象区域的粗检测

传统的对象匹配是通过滑动窗口的形式在待测图像中找到与模板对象最相近的区域。Cheng Ming-Ming 等人^[9]提出了一种简单而强大的“BING”特征方法,该方法通过使用 Objectness 值搜索图像中可能出现的对象区域,如图 2 所示。



图 2 图像中可能出现的对象

本文定义的 64 维线性模型可以近似表示成一系列向量:

$$\boldsymbol{w} \approx \sum_{j=1}^{N_w} \beta_j \boldsymbol{a}_j \quad (1)$$

其中, N_w 表示基向量的个数,基向量 $\boldsymbol{a}_j \in \{-1, 1\}^{64}$, $\beta_j \in \mathbb{R}$ 表示相应的系数,同时 $\boldsymbol{a}_j = \boldsymbol{a}_j^+ - \overline{\boldsymbol{a}_j^+}$,其中 $\boldsymbol{a}_j^+ \in \{0, 1\}^{64}$ 。

对于特征向量 \boldsymbol{b} 有:

$$\langle \boldsymbol{w}, \boldsymbol{b} \rangle \approx \sum_{j=1}^{N_w} \beta_j (2 \langle \boldsymbol{a}_j^+, \boldsymbol{b} \rangle - |\boldsymbol{b}|) \quad (2)$$

$$g_l = \sum_{k=1}^{N_g} 2^{8-k} b_{k,l} \quad (3)$$

其特征算法如表 1 所列。

表 1 BING 特征算法

窗口大小为 $W \times H$, 其 BING 特征的定义过程
输入: 二值梯度图像 $\tau_{W \times H}$
输出: BING 特征矩阵 $\boldsymbol{b}_{W \times H}$
初始化: $\boldsymbol{b}_{W \times H} = 0, \tau_{W \times H} = 0$
for (x, y) do
$r_{x,y} = (\tau_{x-1,y} << 1) \tau_{x,y}$
$\boldsymbol{b}_{x,y} = (\boldsymbol{b}_{x,y-1} << 8) r_{x,y}$
end for

其中, BING 特征 $\boldsymbol{b}_{x,y} \in \{0, 1\}^{8 \times 8}$, 行向量 $r_{x,y} \in \{0, 1\}^8$, 向量元素 $\tau_{x,y} \in \{0, 1\}$ 。

为了得到尽可能多的对象,必须扫描所有规模的对象

可能尺寸。在每个 64 维线性梯度模型中为每个对象定义一个分值,其线性公式表示如下:

$$s_i = \langle \boldsymbol{w}, g_i \rangle \quad (4)$$

$$l = (i, x, y) \quad (5)$$

对象的 Objectness 得分定义如下:

$$o_i = v_i \cdot s_i + t_i \quad (6)$$

式(4)、式(5)中 $s_i, g_i, l, i, (x, y)$ 分别表示得分、NG 特征值(64 维空间线性梯度)、位置、尺寸和各自对象窗口的位置。式(6)中 v_i 表示各自的学习系数, t_i 表示偏移量。

通过评价该 Objectness 特征值即可有效地捕获对象在图像中可能存在的位置。粗识别出对象之后,在这些区域里采用模板匹配的方式计算其相关性,以实现物体的精准识别。

2.2 基于梯度方向模板的快速匹配

模板匹配是数字图像处理的重要组成部分之一。把不同传感器或同一传感器在不同时间、不同成像条件下对同一景物获取的两幅或多幅图像在空间上对准,或根据已知模式到另一幅图中寻找相应模式的处理方法就叫做模板匹配。

2.2.1 模板主方向的选择

当前模板匹配过程中,通常模板数量比较多,没有进行很好的分类。为了缩短匹配时间,提升匹配精度,有必要对模板进行相应的处理。我们将模板按主方向来分类,首先对不同的模板计算其各个像素点的梯度方向,找出最多的梯度方向作为其主方向。其次在匹配过程中,只有当滑动窗口主方向与模板主方向一致时才计算其相似度。

具体做法:将模板在 360 度范围内平均分成 36 个角度范围,建立 36 个特征模板。实际上,我们只考虑了 0 到 180 度范围内。通过 Sobel 边缘检测计算像素点的梯度,引入直方图的方法,计算每个方向范围内的像素点的个数,将像素点个数最多的方向设定为该模板的主方向。

2.2.2 相似性度量

Steger 提出用梯度方向差的余弦值的绝对值来衡量相似性,其方法如下:

$$\epsilon_{Steger}(I, T, c) = \sum_{r \in P} |\cos(\text{ori}(O, r) - \text{ori}(I, c+r))| \quad (7)$$

其中, $\text{ori}(O, r)$ 是模板图像 O 在位置 r 处的梯度方向, $\text{ori}(I, c+r)$ 表示输入图像 I 在位置 $c+r$ 处的梯度方向。

上述方法有效地避免了微小形变带来的影响,但是在强背景条件下效果不是很明显,因此可以取最大的梯度方向作为主方向。式(7)可以改写成:

$$\epsilon(I, T, c) = \sum_{r \in P} (\max_{t \in R(c+r)} |\cos(\text{ori}(O, r) - \text{ori}(I, t))|) \quad (8)$$

其中, t 为形变范围内某点的位置,其中心为 $c+r$, 区间 $R(c+r) = [c+r-T/2, c+r+T/2] * [c+r-T/2, c+r+T/2]$, T 为形变范围内的最大尺寸。

2.2.3 梯度方向的选择

对 RGB 图像而言,为了提高复杂环境下的鲁棒性,取该图像在每个颜色通道的梯度方向的最大值作为最后的梯度方向。表示如下:

$$I_g(x) = \text{ori}(\hat{C}(x)) \quad (9)$$

$$\hat{C}(x) = \arg \max_{c \in \{R, G, B\}} \left\| \frac{\partial C}{\partial x} \right\| \quad (10)$$

其中, $I_g(x)$ 表示在位置 x 处的像素点在各个颜色通道上的梯度方向, $\hat{C}(x)$ 表示最大的梯度方向。

2.2.4 梯度传播方向

在梯度方向被确定之后,为了提高量化的抗噪性,不能只考虑单个像素的梯度,而应考虑其像素周围的其他像素(比如用 3×3 的模板时,考虑其周围的8个像素的梯度)。本文选用 3×3 模板,将某像素点与周围的8个像素点梯度方向的和作为该中心位置像素的传播方向,再对每个像素点的传播方向进行二值化操作,这样可以有效免量化误差。

同时对于每一次新模板的引进,为了防止评价的过操作,应该选取一个适当的阈值来二值化该梯度方向,低于阈值的梯度不用保留,定义 J 为量化后的传播方向。图3(a)是量化的梯度方向,图3(b)是传播方向,图3(c)是量化的传播方向。

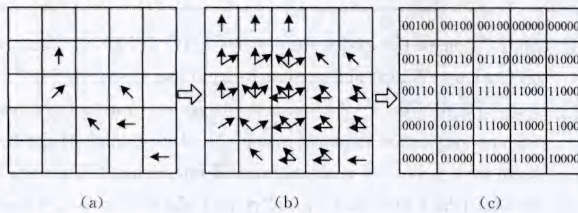


图3

2.2.5 梯度响应图的建立

当把传播方向二值化后,就可以建立一张查找表 T_i ,为快速模板匹配提供条件。查找表定义如下:

$$T_i[L] = \max_{i \in L} |\cos(i - l)| \quad (11)$$

其中, i 是量化方向的索引,为了计算简单,我们把它看作是相应弧度制的角度值; L 为 i 处像素点周围梯度方向的集合,实际上我们是使用二值化的整数值来表示。

对于每一个位置 c ,可以很容易计算不同方向的梯度值。

$$S_i(c) = T_i[J(c)] \quad (12)$$

此时式(8)可以改写为:

$$\epsilon(I, T, c) = \sum_{r \in P} S_{ori(O, r)}(c+r) \quad (13)$$

因为梯度图在各个模板之间是可以共享的,所以对于不同的模板,只要计算了梯度方向,就可以建立查找表,通过查找表可以找出相关性最大的位置,实现对象匹配。

3 实验结果和讨论

数据集选择 PASCAL VOC 2007, 根据 Cheng Ming-Ming 提出的方法,对象粗识别的准确率能达到 96.2%以上,而其识别时间只需要 0.003s。为了分析后续模板匹配的时间复杂度,设获得对象窗口个数为 $T=1000$ 。

每次滑动匹配的时间消耗为 O ,对于大小为 $M \times N$ 的图像,如果不用 Objectness 特征值的方法,其时间复杂度为 $M \times N \times O$;而对对象粗识别之后,其时间复杂度为 $T \times O$ 。通常情况下 $T \ll M \times N$,由此可见本文提出的算法实时性比较好。

为了验证本文提出算法的准确性,在 Microsoft Visual Studio 2010 环境下,模板采用 36 幅不同旋转方向的音箱图像,选取了 300 幅待测图片。实验比较了不同情况下不同模板匹配的准确率。

图4给出了模板对象为音箱的匹配效果,实验中设置区域大小固定。图4中,(a)-(d)分别代表了不同旋转角度下的匹配结果。图5是在不同方向下的音箱的模板。实验发现,其匹配效果基本令人满意。

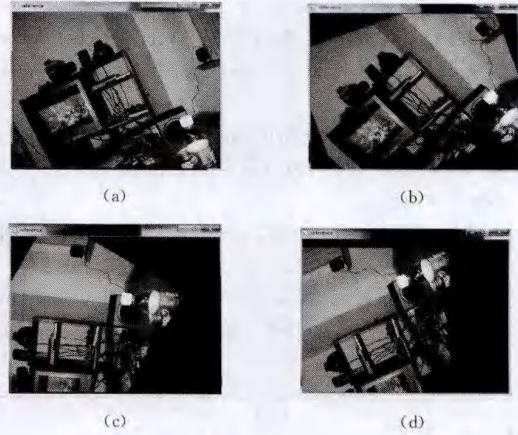


图4 匹配结果



图5 部分模板

实验中,待测对象分为为音箱、羊、汽车3种情况。其在不同方法下的准确率如表2所列。

表2 各种方法下匹配的准确率

模板对象	方法				
	LINE-2D	LINE-3D	Steger	DOT	LINE-MOD
音箱	85.7%	90.6%	92.5%	79.4%	97.5%
羊	89.8%	92.4%	95.8%	82.2%	98.8%
汽车	81.4%	88.5%	92.0%	80.0%	95.4%

从表2可以看出,本文提出的梯度响应图的方法的效果比单纯的 Steger 方法、LINE-2D、LINE-3D、DOT 方法更好;对于轮廓比较鲜明的对象,其识别率也更高。在同样的电脑配置情况下,本文提出的方法响应时间更短,实时性更好。

结束语 本文对现有的图像模板匹配做了简单的介绍,并指出了现有方法存在的一些不足,在此基础上对模板匹配方法做了一些改进。首先在预处理待测图像时,结合 Cheng Ming-Ming 等提出的“BING”特征方法,对对象可能出现的区域进行粗识别。同时在对模板处理时,引入主方向的概念,减少了匹配的复杂度。采用 Steger 提出的方法作为相似性度量方法,该方法量化简单,适于快速检测和匹配。本文建立了一张查找表来保存量化后的梯度信息,这样可以避免每次换模板时重新计算梯度。采用本文方法进行图像匹配能大大减少匹配的时间消耗,实时性更加显著。为了获得更好的方向无关性,同时实现旋转不变性,我们将在以后的工作中引入金字塔算法。

参考文献

- [1] Olson C F, Huttenlocher D P. Automatic Target Recognition by Matching Oriented Edge Pixels[J]. IEEE Transactions on Image

[2] Gavrilu D, Philomin V. Real-Time Object Detection for smart Vehicles[C]//IEEE international Conference on Computer Vision, 2010;533-542

[3] Holzer S, Hinterstoisser S, Ilic S, et al. Distance transform templates for object detection and pose estimation[C]// CVPR, 2009;1177-1184

[4] Steger C. Occlusion Clutter, and Illumination Invariant Object Recognition[C]//IAPRS, 2002

[5] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection[OL]. <http://lear.inrialpes.fr/people/triggs/pubs/Dalal-os.pdf>. cvpr

[6] Ferrari V, Jurie F, Schmid C. From images to shape models for object detection[J]. IJCV, 2009, 87(3):284-303

[7] Tombari F, Franchi A, Automation D, et al. BOLD features to detect texture-less objects[C]//2013 IEEE International Conference on Computer Vision (ICCV 2013), 2013;1265-1272

[8] Hinterstoisser S, Cagniart C, Ilic S, et al. Gradient response maps for real-time detection of textureless objects[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012;34(5):876-888

[9] Cheng M M, Zhang Z, Lin W Y, et al. Bing: Binarized normed gradients for objectness estimation at 300fps[C]//CVPR, 2014; 3286-3293

(上接第 300 页)

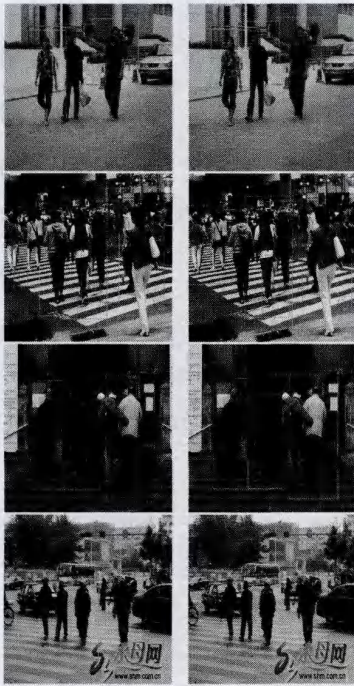


图 7 GPU 行人检测结果和 CPU 行人检测结果

结束语 图像金字塔和图像梯度的快速实现对于诸如 HOG、Shape Context、KLT 等视觉算法的快速实现是至关重要的,它们是视觉算法在进行譬如行人检测、字符识别、图像配准过程中常见的共性操作。合理高效地实现这两个操作对于包含它们的视觉算法是很有意义的。本文合理地设计和使用 GPU 的纹理内存、全局内存、共享内存,在 GPU 上分别进行图像梯度和图像金字塔的求取,力求找到快速处理这两种操作的方法。同时对传统的图像金字塔求取过程进行改进,使其更适合在 GPU 上实现,减少了图像金字塔需要耗费的存储空间。相对于 CPU 上的实现,随着图像分辨率的增加,在 GPU 上实现的加速比逐步增加。本文提出的并行计算方法对于局部特征在 GPU 上的实现也具有借鉴意义。

参考文献

[1] Dalal N, Triggs B. Histogram of oriented gradients for human Detection[C]//CVPR, 2005;886-893

[2] Belongie S, Malik J, Puzicha J. Shape Matching and object recognition Using Shape Contexts[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002, 24(4):509-522

[3] Shi J, Thomas C. Good feature to track[C]//IEEE Conference on Computer Vision and pattern Recognition, 1994;593-560

[4] Lucas B, Kanade T. An iterative image registration technique with an application to stereo vision[C]//Proceedings of the International Joint Conference on Artificial Intelligence, 1982;674-679

[5] Strengert M, Kraus M, Ertl T. Pyramid Methods in GPU-Based Image Processing[C]//Proceeding of Vision, Modeling, and Visualization 2006, 2006;169-176

[6] Dollár P, Appel R, Belongie S. Fast Feature pyramids for object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(8):1532-1545

[7] Nvidia. NVIDIA CUDA A Programming Guide version 4.0 [EB/OL]. <http://www.nvidia.com/object/cuda-cn>

[8] 赵杰伊,唐敏,童若锋.基于 CUDA 的细分曲面阴影体算法[J].浙江大学学报(工学版),2012,46(7):1301-1306

Zhao Jie-yi, Tang Min, Tong Ruo-feng. CUDA based shadow volume algorithm for subdivision surfaces [J]. Journal of Zhejiang University (Engineering Science), 2012, 46(7):1301-1306

[9] 唐家维,王晓峰.基于 GPU 的并行化 Apriori 算法的设计与实现[J].计算机科学,2014,41(10):238-243

Tang Jia-wei, Wang Xiao-feng. Design and Implementation of Apriori on GPU[J]. Computer Science, 2014, 41(10):238-243

[10] Lindeberg T. Scale-space theory in computer vision[M]. London; Kluwer Academic Publishers, 1994

[11] 王森,杨克俭.基于双线性插值的图像缩放算法的研究与实现[J].动化技术与应用,2008,27(7):44-46

Wang Sen, Yang Ke-jian. An Image Scaling Algorithm Based on Bilinear Interpolation with VC++ [J]. Techniques of Automation and Applications, 2008, 27(7):44-46

[12] 冯煌. GPU 图像处理的 FFT 和卷积算法及性能分析[J]. 计算机工程应用, 2008, 44(2):120-122

Feng Huang. Implementation and performance of FFT and convolution in image filtering on GPU[J]. Computer Engineering and Applications, 2008, 44(2):120-122

[13] 吴恩华. 图形处理器用于通用计算的技术、现状及其挑战[J]. 软件学报, 2004, 15(10):1493-1504

Wu En-hua. State of the Art and Future Challenge on General Purpose Computation by Graphics Processing Unit[J]. Journal of Software, 2004, 15(10):1493-1504

[14] 马安国,成玉,唐遇星,等. GPU 异构系统中的存储层次和负载均衡策略研究[J]. 国防科技大学学报, 2009, 31(5):38-43

Ma An-guo, Cheng Yu, Tang Yu-xing, et al. Research on Memory Hierarchy and Load Balance Strategy in Heterogeneous System Based on GPU[J]. Journal of National University of Defence Technology, 2009, 31(5):38-43