

# 基于双向融合机制的建筑目标检测方法<sup>\*</sup>

金泰松<sup>1</sup>, 汪哲慎<sup>2</sup>, 叶聪颖<sup>2</sup>, 李翠华<sup>2</sup>

(1. 北京理工大学 计算机科学与技术学院 计算机科学工程系, 北京 100081; 2. 厦门大学 信息科学与技术学院 计算机科学系, 福建 厦门 361005)

**摘要:** 模拟人类视觉中有意识主动寻找与无意识被动受吸引相交互的视觉过程, 提出了一种基于双向融合机制的建筑目标检测方法。该方法综合了基于自底向上数据驱动的视觉显著性模式和自顶向下基于环境感知的目标搜索模式检测图像中的建筑目标区域。利用拍摄的自然图像进行实验表明: 该方法能较好地检测出图像中的建筑目标, 具有处理速度快、准确性高的特点, 能够满足处理复杂场景图像的实时性要求。

**关键词:** 双向融合; 建筑; 显著性; 竖直线**中图分类号:** TP391 **文献标志码:** A **文章编号:** 1001-3695(2008)02-0455-03

## Approach to building extraction based on two-way fusion mechanism

J N Tai-song<sup>1</sup>, WANG Zhe-shen<sup>2</sup>, YE Cong-ying<sup>2</sup>, LI Cui-hua<sup>2</sup>

(1. Dept of Computer Science &amp; Engineering, School of Computer Science &amp; Technology, Beijing Institute of Technology, Beijing 100081, China; 2. Dept of Computer Science, School of Information Science &amp; Technology, Xiamen University, Xiamen Fujian 361005, China)

**Abstract:** This paper proposed an approach called two-way fusion mechanism, which detected the interested building in the image. It imitated the detection process of human vision, in which searching subjectively and being attracted objectively happened at the same time. Through the bottom-up methods which utilized visual saliency model and the top-down methods based on environmental perception technology, the regions where building exist were marked in the image. Experiment on natural images validates that this mechanism has competitive accuracy in building detection in complex scene and favorable speed, and satisfies the requirement of processing the natural image.

**Key words:** two-way fusion; building; saliency; vertical line segment

选择图像目标区域的过程一般包含两种机制: a) 基于自底向上数据驱动的视觉显著性模式。它模拟人类无意识地被场景中的凸显区域所吸引的过程。b) 基于自顶向下的任务驱动的主动搜索模式。它模拟人们利用先验知识和经验, 有意识地寻找目标的过程。

传统的目标检测与识别方法通常采用自顶向下的方式, 利用目标的先验知识, 试图从整个场景中寻找符合特征要求的目标, 忽视了快速泛读的重要性, 使得大量的计算资源浪费在与目标无关的细节上。研究人员也提出了基于自顶向下任务引导的计算模型<sup>[1,2]</sup>模拟人眼的任务搜索功能。然而以往提出的计算模型一般只利用目标本身的知识, 目标周围的环境知识利用得不多。事实上, 目标的环境知识对识别结果也起着重要作用, 过去这方面的研究多见于场景分析、机器人自主导航控制领域<sup>[3,4]</sup>。随着视觉心理学研究的深入, 计算机视觉中也引入了选择性注意机制, 提出了基于自底向上策略的视觉显著性模型。相关的计算模型很多<sup>[5,6]</sup>, 最新提出的计算模型<sup>[6]</sup>对自然图像也取得了很好的效果。然而选择性注意机制中自顶向下的方式常常被忽略, 原因在于如何使先验知识参与到注意的计算方面存在较大的困难<sup>[7]</sup>。

在复杂场景图像中检测建筑目标, 无论是采用自顶向下的检测方式, 还是根据目标在场景中客观视觉显著性的不同进行

检测均不符合人类视觉系统特性。因此, 本文提出了基于自顶向下与自底向上相结合的双向融合策略的检测方法, 试图克服两种方式各自孤立使用的缺陷, 达到更好的目标检测效果。

### 基于环境感知的搜索模式

在建筑目标检测中, 建筑周围的环境信息对判断建筑很有帮助。然而已往的研究中, 环境信息如植被区、树木、道路等一般均被当做目标的干扰信息被排除; 若能对这些环境信息适当处理, 将有助于检测图像中的建筑目标。

一般说来, 建筑目标的纹理、细节等信息均比较显著的, 传统的基于像素的影像处理方式的效率以及其所能获得的结果信息都是十分有限的。因此本文利用基元<sup>[8]</sup>作为处理的基本单位, 通过将一幅图像分割成不同基元组成的区域; 然后根据识别对象的不同, 将图像中基元分为目标基元和环境基元两大类。

图像中不同基元之间具有一定的位置关系, 如拓扑关系、重合关系、遮挡关系等。例如建筑的一部分被树木遮挡; 树木一般生长在公路的两侧; 车辆、停车场一般位于建筑附近。这些目标和环境的相互关系对目标检测起着重要作用, 特别是在复杂场景图像中, 目标本身的信息不足以完成识别的任务, 此时利用目标和环境的关系, 如遮挡关系等有着一定的启发

收稿日期: 2006-10-10; 修回日期: 2007-03-20 基金项目: 国家创新研究群体资助项目 (60024301); 国家自然科学基金资助项目 (60175008); 厦门大学“985”二期信息创新平台资助项目

作者简介: 金泰松 (1978-), 男, 博士研究生, 主要研究方向为图像处理与模式识别 (jtsong\_2005@yahoo.com); 汪哲慎, 硕士; 叶聪颖, 硕士; 李翠华 (1960-), 男, 教授, 博导。

作用。本文充分利用环境信息对目标的启示作用,在自顶向下任务驱动搜索模式中采用基于环境感知的方法确定图像中的建筑目标区域。

1)提取某种环境基元 建筑图像中的环境基元多种多样,如车辆、道路、树木等均是环境基元,而且不同图像中的环境信息对目标的启示作用各不相同,这取决于已有的先验知识,并且根据先验知识的不同,提取图像中的环境基元。

2)提取图像中的建筑目标区域 本文研究自然图像中的主要建筑具有较长的外墙体直线段并且是基本正立。对建筑目标的边缘图进行分析发现:竖直线是建筑目标存在的明显特征。因此本文提取图像的竖直线特征,并分析如图 1 所示的竖直线的空间位置关系,进行竖直线分组,每个组均对应着图像中存在建筑目标的区域。

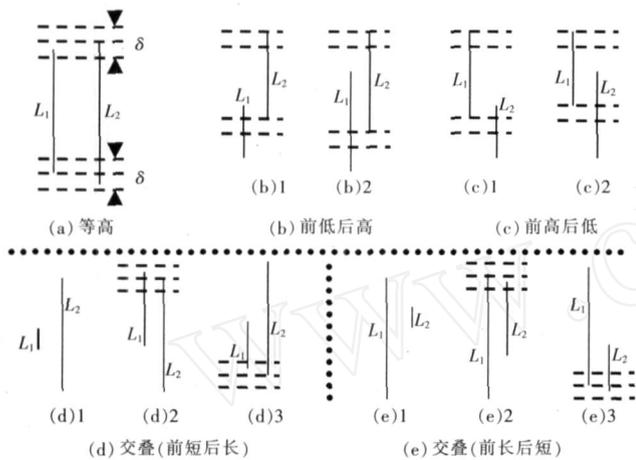


图 1 两条竖直线之间的关系模式

最后,分析提取的环境基元与图像中建筑目标区域的关系,如位置关系、拓扑关系等,修正得到的目标区域。例如已知位于道路两侧的建筑目标,利用道路和建筑的位置关系,确定图像中的建筑目标区域。

### 自底向上数据驱动的视觉显著性模型

在自然图像中,建筑目标相比天空、树林、山脉等自然背景往往更容易吸引视觉的注意,因此通过视觉注意力机制进行计算可以找出图像中视觉焦点集中的区域。这种方法能够有效排除自然景物的干扰,缩小搜索范围,对建筑目标识别等视觉处理任务是非常有意义的。

本文在文献 [9] 中的模型基础上针对建筑图像的特点进行了改进,并应用于建筑图像的显著性计算。文献 [9] 中的模型提取的特征包括了方向信息,而对于建筑图像来说,人工建筑的方向一般均比较规则,相互之间差异小;反倒是一些自然景物具有不规则和差异较大的方向特征,因此方向信息不适合被用在显著性建筑区域检测中。为了克服这个问题,笔者摒弃了方向信息这个特征,而以亮度和颜色信息为主进行显著性计算;同时为更好地反映颜色信息的显著性,选择在 HS 颜色空间进行特征提取,然后利用文献 [10] 中的方法提取建筑图像的显著图。该方法从输入的 RGB 图像中提取 H、S、I 三个分量的信息后,分别构造它们的 Gaussian 金字塔。在三个特征金字塔的基础上,接着进行视觉差异程度的计算和显著图的合成。计算每个特征的 center-surround 的特征差异图,对不同尺度下的特征差异图进行多尺度融合,最终得到建筑图像的显著图。

### 双向融合的目标检测机制

在两种建筑目标检测策略的基础上,本文提出的基于双向融合策略的建筑目标检测方法。该方法融合了基于环境感知的任务搜索模式和视觉显著性模型两种策略:a)基于环境感知的目标检测过程属于自顶向下任务驱动的主动搜索模式,即利用目标的先验知识和经验(如目标和环境的位置关系),在图像中有意识地寻找目标;b)显著性计算过程属于自底向上数据驱动的视觉显著性模式,即基于目标本身醒目的设计原则,计算出客观图像中感兴趣目标区域。

对建筑图像来说,基于双向融合策略的目标检测方法首先利用先验知识,特别是目标和环境的关系快速定位出建筑图像中的建筑目标的区域,获得建筑目标的分割图。在分割图上对利用视觉注意力模型获得各区域的显著性效果。根据分割区域显著性效果的不同,提取图像中超过阈值的区域,确定为建筑目标的候选区域并进行验证,获得图像中建筑目标的位置。

### 实验结果和分析

本文利用基于双向融合的建筑目标检测方法实现了一个实验系统。该系统首先利用文献 [10] 中的竖直线分组算法获得图像中建筑目标的区域,并根据树木、道路、草地的纹理不同提取图像中的环境基元,对每一幅图像提取何种基元由输入指定;然后,将图像中建筑区域中的点到包含环境基元最小矩形中心的距离定义为邻近关系,并去除不满足邻近关系的建筑目标区域。同时利用文献 [11] 中的视觉注意力的算法提取各分割区域的显著性图。最后,将区域中各个像素的显著性值与设定阈值进行比较,将高于设定阈值的像素确定为图像中的建筑目标区域。

实验中,将实地拍摄的大量建筑图像作为测试集图像进行了实验。图像是在不同光照条件下拍摄的大小为 640 × 480, JPG 格式的自然图像。以下是一个典型图像的实验结果。已有先验知识是建筑目标位于草地上,因此提取图像中草地信息作为环境基元。

图 2 是利用环境感知方法对原图进行分割的结果。其中利用草地的纹理特征提取图像底部的草地作为环境基元,并根据建筑目标位于草地之上的知识,将图像分割为上下两个部分。同时根据图像中竖直线的分组关系,将图像中草地以上部分分割两个部分。如图 2(c) 所示。从 (c) 可以发现:建筑目标位于图像中一个分割区域内,但不能精确地确定它在图像中的位置。



图 2 利用环境感知方法得到的分割图

接下来,对得到的分割图利用视觉注意力方法获得的显著图进行分析。图 3 是分割图的显著图效果。从显著图上看出:与其他区域相比,建筑目标区域的显著性比较高,算法中设定阈值参数为 0.3,提取显著性值高于阈值的像素值作为候选的建筑目标区域,并根据建筑目标的形状信息检测出建筑目标在图像中的位置,如图 4 所示。

本文以平均运行时间作为算法时间性能的评价标准,利用两种方法建立了建筑目标识别系统:a)本文基于双向融合的区域检测方法,称为算法 1;b)文献 [11]中的区域检测方法,称为算法 2。

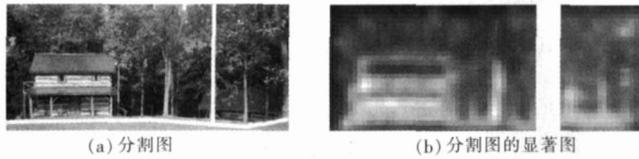


图 3 利用显著图模型得到的效果图

算法 1 中主要包括两个算法步骤:a)提取图像的显著图,称为算法步 1;b)利用环境感知方法获得图像中的建筑目标区域,称为算法步 2。利用测试集图像进行了实验,实验环境是主频 3 GHz,1 GB 内存。表 1 为实验结果。



图 4 检测出的建筑目标

表 1 算法的平均运行时间

算法	平均运行时间/ms
算法步 1	200
算法步 2	50
算法 2	700

从表 1 看出,建筑物检测算法主要耗时在于获取图像的显著图。其中:本文给出的算法中获取图像显著图的算法步平均耗时是 0.2 s;文献 [11]中的计算图像显著图的算法平均耗时是 0.7 s。这是因为本文算法利用已有先验知识对图像中建筑目标区域进行了分割,并计算分割图的显著性。因此与算法 2 相比,本文算法的时间复杂度有了明显的降低,满足处理自然图像的实时性要求。

### 结束语

本文提出了一种基于自顶向下和自底向上检测策略相结合的建筑目标检测方法。在基于自顶向下的主动搜索过程中,利用环境感知技术寻找图像中的建筑目标区域。在基于自底

向上的检测过程中,采用视觉显著性模型获得图像中不同区域的显著性图;然后融合两种区域检测策略提出了一种建筑目标检测方法。实验表明,该方法取得了较好的检测效果和效率。下一步,笔者要更深入地研究纹理、颜色等信息在分割和识别中的应用,探索自底向上和自顶向下两种搜索策略更好的结合方式。此外,仅仅基于可见光图像也是不够的,还应该通过多传感器信息融合技术,结合其他如红外、DEM 等数据信息,达到更好的分割和识别效果以及让该方法在更加复杂的自然环境和天气条件下拥有更好的适应性。

### 参考文献:

- [1] ITTL, KOCH C. Computational modeling of visual attention [J]. *Nature Reviews Neuroscience*, 2001, 2 (3): 194-203.
- [2] NAVALPAKKAM V, ITTL. Modeling the influence of task on attention [J]. *Vision Research*, 2005, 45 (2): 205-231.
- [3] 邹小兵,蔡自兴.基于传感器信息的环境非光滑建模与路径规划 [J]. *自然科学进展*, 2002, 12 (11): 1188-1192.
- [4] SALICHS M A, MORENO L. Navigation of mobile robots: open questions [J]. *Robotica*, 2000, 18 (2): 227-234.
- [5] PETERS R J, IYER A, ITTL, *et al*. Components of bottom-up gaze allocation in natural images [J]. *Vision Research*, 2005, 45 (8): 2397-2416.
- [6] MEUR O L, CALLET P L, BARBA D, *et al*. A coherent computational approach to model bottom-up visual attention [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28 (5): 802-817.
- [7] 桑农,李正龙,张天序.人类视觉注意机制在目标检测中的应用 [J]. *红外与激光工程*, 2004, 33 (1): 38-42.
- [8] 明冬萍,骆剑承,周成虎,等.高分辨率遥感影像信息提取及块状基元特征提取 [J]. *数据采集与处理*, 2005, 20 (1): 34-39.
- [9] ITTL. Visual attention and target detection in cluttered natural scenes [J]. *Optical Engineering*, 2001, 40 (9): 1784-1793.
- [10] 叶聪颖.建筑目标识别中检测与分割技术的研究 [D]. 厦门:厦门大学,2005:36-47.
- [11] 叶聪颖,李翠华.基于 HSI 的视觉注意力模型及其在船只检测中的应用 [J]. *厦门大学学报:自然科学版*, 2005, 44 (4): 484-488.

(上接第 451 页)是图像  $x, y$  在  $(i, j)$  位置的像素值。因为  $C_k$  是融合后图像与原始高分辨图像经高通滤波后的相关系数,所以,相关系数越大,表示更多地保持了原高分辨图像中的高频细节信息。

表 1 和 2 为融合评价指标的各项数据比较。从表中数据可以看出,HSI 方法的高频信息得到了很好的保留,但是光谱信息丢失比较严重;WT 方法的光谱差异较小,但是高频细节信息丢失较多;本文方法的光谱差异较小,与原始高分辨图像的高频成分间的相关系数较大,可见本文方法在保留光谱信息和提高分辨率两个方面综合而言是最优的。

表 1 与原始多光谱图像的光谱差异

项目	通道		
	$D_R$	$D_G$	$D_B$
HIS 方法	30.420 0	24.937 6	21.867 8
WT 方法	14.428 0	14.768 9	13.002 3
本文方法	12.485 8	12.805 4	10.883 3

表 2 与原始高分辨图像高频成分间的相关系数

项目	通道		
	$C_R$	$C_G$	$C_B$
HIS 方法	0.993 5	0.994 5	0.975 3
WT 方法	0.886 8	0.892 3	0.871 8
本文方法	0.991 5	0.993 0	0.973 5

### 结束语

本文提出的基于 HSI 和小波变换的低分辨多光谱和高分辨全色图像的融合方法通过对高分辨图像的低频进行低通滤波的方法,将高分辨图像低频中的高频信息融入到多光谱图像

的低频中。试验结果表明,该方法在保留多光谱图的光谱特性和高分辨率的空间信息两方面是有效的。

### 参考文献:

- [1] CLARKE J J, YULLE A L. Data fusion for sensory information processing system [M]. Norwell, MA: Kluwer, 1990: 1-15.
- [2] HALL D L. Mathematical techniques in multi-sensor data fusion [M]. Boston: Artech House, 1992: 20-59.
- [3] 倪国强.多波段图像融合算法研究及其新发展 ( ) [J]. *光电子信息技术*, 2001, 14 (5): 11-17.
- [4] 倪国强.多波段图像融合算法研究及其新发展 ( ) [J]. *光电子信息技术*, 2001, 14 (6): 1-17.
- [5] CHAVEZ P S Jr, SDES S C, ANDERSON J A. Comparison of three different methods to merge multi-resolution and multi-spectral data: TM & SPOT pan [J]. *Photogramm Eng Remote Sensing*, 1991, 57 (3): 295-303.
- [6] 李弼程,魏俊,彭天强.基于 HSI 变换与小波变换的遥感图像融合 [J]. *数据采集与处理*, 2003, 18 (3): 268-272.
- [7] SHETTIGARA V K. A generalized component, substitution technique for spatial enhancement of multi-spectral images using a high-resolution data set [J]. *Photogramm Eng Remote Sensing*, 1992, 58 (5): 561-567.
- [8] DAVD A Y. Image merging and data fusion by means of the discrete two-dimensional wavelet transform [J]. *J Opt Soc Am A*, 1995, 12 (9): 1834-1841.