

单幅数字图像多尺度空间下的场景深度估计

陆振杰, 宋 进

(上海大学 机电工程与自动化学院, 上海 200072)

摘 要: 双目立体视觉方法已广泛应用于移动机器人导航领域, 通过该方法得到的视差图实现了对场景深度的有效估计, 然而, 双目立体视觉方法需要对图像做匹配, 计算量大, 不适合动态场景的深度信息的获得。为了避免图像匹配的计算, 以得到真实场景的深度估计, 文中提出了一种基于 Markov 随机场模型对单幅数字图像特征建模来获得场景深度信息的方法。实验证明, 通过单幅数字图像获得的场景深度可以有效地估计真实场景中摄像机与场景目标之间的距离, 并且, 随着尺度空间的变大, 可以有效减小其所获得的深度值误差。

关键词: 双目立体视觉; 机器人导航; 单幅数字图像; Markov 随机场; 场景深度估计; 多尺度空间

中图分类号: TP31

文献标识码: A

文章编号: 1673-629X(2013)01-0051-03

doi: 10.3969/j.issn.1673-629X.2013.01.013

Scene Depth Estimation for Single Digital Image in Multi-scaled Space

LU Zhen-jie, SONG Jin

(College of Mechanical and Automation, Shanghai University, Shanghai 200072, China)

Abstract: Binocular stereo vision method has been widely used in mobile robot navigation area, the parallax diagram get by this method realizes the effective estimation for scene depth, however, binocular stereo vision method needs to do the image matching with large amounts of calculation, and is not suitable for dynamic scene depth information gain. In order to avoid the calculation of image matching, getting real scene depth estimation, present a method based on Markov random field model for singles digital image feature modeling to get scene depth information. The experiment proved the scene depth through the single digital image can effectively estimate the distance between camera and target scene in real scene, and with the variable scale space, can effectively reduce its received depth value error.

Key words: binocular vision; robot navigation; single digital image; Markov random field; scene depth estimation; multi-scaled space

0 引 言

三维场景映射到二维数字图像上时, 丢失了场景中目标物体的深度信息, 而真实场景的深度信息一直是导航机器人需要获得的关键数据^[1], 在已获取的数字图像基础上, 现有计算机视觉方法已能通过对图像信息的分析和计算来有效估计场景的深度, 以还原三维场景的深度信息。对通过双目立体视觉方法获得场景视差图的研究已成为热点, 然而目前典型的双目视觉系统存在成本高、计算量大、不适合动态场景重建等问题^[2]。因此, 试图寻找一种仅用单幅数字图像来估计真实场景深度的方法, 以避免双目视觉系统中的特征匹配和计算等问题。

除了双目视觉方法, 其他一些场景深度估计的计

算机视觉方法有利用激光测距仪获得场景深度图方法, 通过运动分析, 由序列图像估计三维物体运动参数与几何形状的方法; 通过对图像做散焦处理来估计场景深度的方法等^[3]。文中提出一种仅利用单幅数字图像来估计场景深度的方法; 主要方法是在对已获取图像做分割处理的基础上; 通过对分割图像建立 Markov 随机场概率模型, 建模必须要充分考虑图像本身的特征属性, 比如阴影、遮挡、颜色、纹理、共面和共线等^[4], 采用机器学习理论中线性规划学习算法和统计学理论中最大似然估计的参数估计方法来实现对场景深度的有效估计。过程需要考虑图像的局部和全局特征, 并利用一些场景的先验知识, 通过机器学习以及概率学一些理论来实现。这种方法的缺点是误差比较大, 且影响误差的因素比较多。因此, 如何来降低该方法的误差也是亟待考虑的问题。

1 图像特征

在对图像建立 Markov 随机场模型之前首先需要对图像进行分割, 将图像分割成许多单元块^[5], 然后用

收稿日期: 2012-04-18; 修回日期: 2012-07-22

基金项目: 上海市科研重大项目 (10dz1203500)

作者简介: 陆振杰 (1986-), 男, 上海人, 硕士研究生, 研究方向为进行模式识别、图像匹配、机器学习方面的研究; 宋 进, 副教授, 研究方向为自动化检测、智能机器人视觉导航、图像处理方面的研究。

估计每个单元块深度的方法来得到场景的深度图,因此,所分割的单元块即为最小单位。这里直接使用一种过分割的图像分割算法^[6],该算法中的单元块称为超像素,并且假设超像素中的像素位于或近似位于同一平面,这种假设合理考虑了真实场景的先验知识,因此,对图像进行过分割处理相比其他一些分割算法,对估计场景深度更有效。

为了得到图像的概率模型,需要为每一超像素的特征建立特征向量,并考虑其与邻近超像素特征间的关系。这里考虑两类特征,一类为绝对深度特征,用于估计场景的绝对深度;另一类为关联深度特征,用于估计相邻超像素的相对深度。

选取的深度特征需要捕捉三种视觉线索:纹理变化、纹理梯度和颜色。图像的纹理信息一般可以在亮度通道中得到,因此用罗斯掩膜^[7]对图像的亮度通道做卷积,如图 1 所示前 9 个掩膜,这些卷积过滤器可以用于检测纹理的能量并可对图像做局部均匀化、边缘检测和点检测等^[8];后 6 个掩膜可以通过对图像亮度通道做卷积来计算纹理的梯度估计;颜色的线索可以用图 1 所示第一个掩膜对图像的颜色通道做卷积来得到。



图 1 用于获取图像特征的掩膜

1.1 绝对深度特征

过分割处理后,假设图像 $I(x, y)$ 被分为 n 个超像素单元,通过 $F_n (n = 1 \cdots 17)$ 个掩膜与图像做卷积(9 个罗斯掩膜、6 个纹理梯度掩膜和 2 个颜色掩膜),得到每个超像素单元的能量 $E_i(n) = \sum_{(x,y) \in \text{patch}(i)} |I * F_n|^k$, 其中 $k \in \{1, 2\}$, 分别表示所有单元的能量和及所有单元能量的平方和,因此,每个超像素单元可以定义 34 个特征,得到 34 维的特征向量,如图 2 所示。

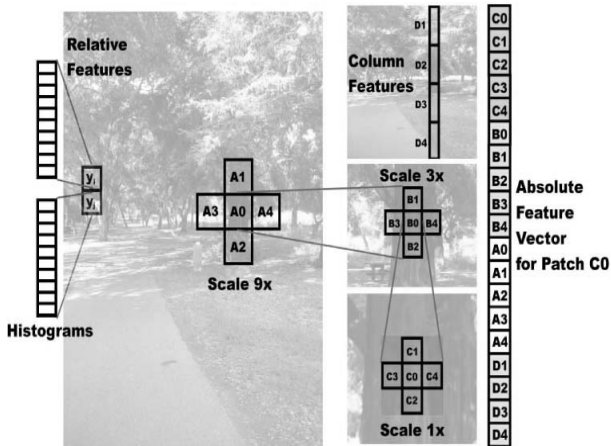


图 2 用于估计场景深度的绝对深度特征

为了准确地估计整个场景中每个超像素单元块的

深度值,以上定义的局部特征向量还无法满足整个场景的需求,因此需要考虑场景的一些全局特性,因此,有必要在不同尺度空间下对相邻超像素定义特征向量,所以,当 $\text{scale} = 3$ 且考虑超像素 i 的 4-邻近单元时,有 $34 \times 5 \times 3 = 510$ 个特征。之所以要选择不同尺度空间是因为场景中的物体在不同的分辨率下会有变化,选用不同尺度空间可以捕捉到这些变化,另外,计算多尺度空间下的特征也有助于解释物体不同的相对尺寸,以获得更好的全局特性。

为了准确地估计单元块的深度值,在不同尺度空间下,同样计算与其相邻的单元块。最后,许多结构会表现出一些竖直的结构,例如树木和楼房生长于地面而不会悬浮在空中,因此,为每个超像素单元定义其所在列的 4 个竖直特征,最终,每个超像素单元的特征向量 x 为 $19 \times 34 = 646$ 维。

1.2 关联深度特征

使用一个不同的特征向量描述关联深度特征,计算 17 个滤波器输出的 10 柱直方图,可以得到任一超像素单元 i 在特定尺度 s 下的 170 个特征 y_{is} ,这些特征可以用来估计场景中不同位置的深度关系,这种估计需要很少的用来估计绝对深度的全局信息,但需要该超像素单元更详细的局部信息^[9]。比如,不需要大量的全局特征就可以估计两个拥有近似特征的超像素单元属于同一个物体,因此,它们的深度值是近似的。因此,超像素单元 i 和 j 在尺度空间 s 下的关联深度特征 y_{ijs} 可以被定义为它们直方图的代数差 $y_{ijs} = y_{is} - y_{js}$ 。

2 Markov 随机场概率模型

由于图像的局部特征,即绝对深度特征不足以准确地估计场景深度,因此,需要考虑空间场景中更全局的特征,即关联深度特征。尽管要估计图像上最小单元块的深度很大程度上依赖该单元块的特征向量,但该单元块的相邻最小单元块一定程度上也可以帮助我们更好地估计深度值,因此,需要充分考虑图像的相邻最小单元间的关系,因为这种相邻关系包含了关于场景的部分先验知识,比如颜色特征相近的邻近单元块可以认为处在一个平面上^[10]。这种典型的相邻关系即 Markov 随机场理论中的 4-邻近关系,为获得多尺度空间的深度关系,将图像的尺度空间定义为 $s = 1, 2, 3$,如图 3 所示。

Patric Perez 发表的 Markov Random Field and Images 中,对 Markov 随机场模型及其与二维图像的关系做了具体阐述,证明 Markov 随机场模型是二维图像特征较理想的模型。因此,在众多典型的 Markov 随机场模型中,选择用概率模型中最常用的联合高斯概率 Markov 随机场模型来建立概率模型。

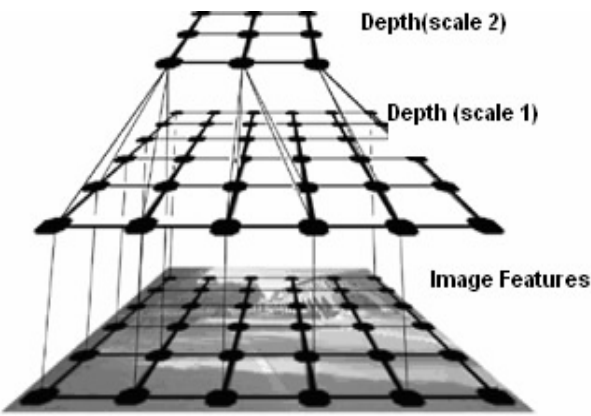


图3 图像不同尺度空间下的像素结构

$$P_c(d \mid X; \theta, \sigma) = \frac{1}{Z_c} \exp \left(- \sum_{i=1}^M \frac{(d_i(1) - x_i^T \theta_r)^2}{2\sigma_{lr}^2} - \sum_{s=1}^3 \sum_{i=1}^M \sum_{j \in N_i(s)} \frac{(d_i(1) - x_i^T \theta_r)^2}{2\sigma_{lr}^2} \right) \quad (1)$$

其中 M 为最低尺度空间下最小单元块的总数, Z 为模型的归一化常数, x_i 为单元块 i 的绝对深度特征向量, θ 和 σ 为需要学习的模型参数向量,事实上,因为图像是由水平安装的摄像机拍摄的,因此,图像的不同排有不同的静态特性,比如,蓝色的超像素单元在场景图像较上方时可能是蓝天,场景图像的较下方时可能是海水,所以对图像中的每一排像素使用不同的参数 $(\theta_r, \sigma_{lr}, \sigma_{lr})^{[11]}$ 。

在考虑不同尺度空间下的深度时,给出严格的约束条件 $d_i(s+1) = (1/5) \sum_{j \in N_i(s) \cup |i|} d_j(s)$,即较高尺度空间下某单元块的深度为较低尺度空间下该单元块与其邻近单元块深度的平均,其中 d_i 为单元块 i 的深度值, $N_i(s)$ 为单元块 i 的4邻近单元块。

3 实验

所有实验训练图像数据来自 <http://make3d.cs.cornell.edu/data.html>,共534张2272×1704分辨率的数字图像,其中400张用于训练,134张用于测试,这些图像数据分为森林、校园和室内场景。这些数据来自三维激光测距仪(SICK LMS-291)的扫描数据,它的最远测距为80m,测距误差为35mm。实验计算机为Windows XP Professional系统,基本硬件配置为Intel(R) Xeon(TM) CPU 2.8GHz 2.79GHz 2.00GB of RAM,图4所示为其中一张数字图像及其用SICK测距仪获得的场景深度图。

图5所示为一些场景训练之后所得到的场景深度图,其图像模型 $scale = 1$,分别训练不同尺度空间下的图像模型为 $scale = 1, 2, 3$,定义深度值误差 $error = | \frac{d_{laser} - d_{estimate}}{d_{laser}} |$,可以得到表1所示的误差数据。

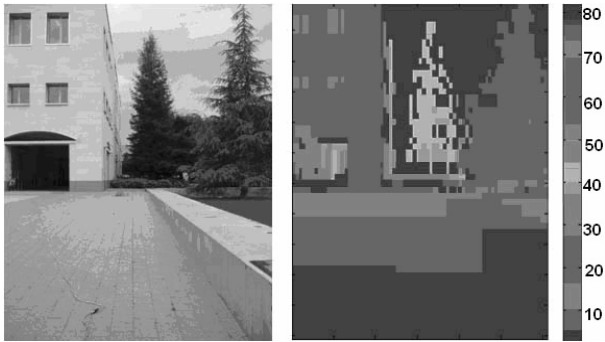


图4 真实场景及其深度图(SICK)

表1 不同场景和尺度空间下的深度值误差

IMAGE MODEL	FOREST	CAMPUS	INDOOR	ALL
Gaussian (s=1)	0.189	0.192	0.178	0.184
Gaussian (s=2)	0.179	0.173	0.166	0.175
Gaussian (s=3)	0.154	0.156	0.142	0.146

上述实验数据表明,在各种不同场景的数字图像的训练后,得到的深度值误差在0.17左右,证明该方法可以有效地对单幅数字图像中的场景做深度值估计。另外,场景深度值的误差随着尺度空间的变大而减小,因此,在讨论图像模型的时候,适当地考虑尺度空间 $scale$ 可以缩小场景深度值的误差。



图5 数据训练后的真实场景及其深度图 (scale = 1)
(下转第58页)

录数 = $19/20 = 95.0\%$

本实验选取的 Web 页面相对规整,可能会使抽取精度相对偏高,但是实验结果已经足以表明文中所设计模型是合理有效的。

4 结束语

文中提出了一种新的 Web 信息抽取的方法,即 WIEBC 方法,并设计了一个完整的系统模型。与已有模型的不同之处在于模型中增加了页面主体提取模块,引入了页面分类方法,并且利用正则表达式自动生成各类抽取规则,能够动态生成数据库中的表格,提高了抽取方法的通用性和准确性。最后用实验证实了 WIEBC 方法的可行性。文中下一步工作是将提出的模型以及其中的算法加以改进和完善,并推广到更复杂网页的信息抽取中。

参考文献:

- [1] 陈 静,朱巧明,贡正仙. 基于 Ontology 的信息抽取研究综述[J]. 计算机技术与发展,2007,17(10):84-86.
- [2] 周合明,奚建清. 基于模板的 Web 信息提取系统的设计与实现[J]. 计算机技术与发展,2011,21(11):105-108.
- [3] Xie Tao, Shi Shengsheng, Quan Fuliang, et al. Research on Complex Structure-oriented Accurate Web Information Extraction Rules[C]//International Conference on Progress in

Informatics and Computing. [s. l.]:[s. n.],2010.

- [4] 周 登,戴玉刚,付 涛. 基于树结构的 Web 信息抽取[J]. 计算机技术与发展,2009,19(9):38-41.
- [5] 李效东,顾毓清. 基于 DOM 的 Web 信息抽取[J]. 计算机学报,2002,25(5):526-533.
- [6] Liu Yaqing, Chen Rong, Yang Hong. Web Information Extraction Based on Hierarchical Model[C]//Second International Conference on Computational Intelligence and Software Engineering. [s. l.]:[s. n.],2009.
- [7] 曾维佳. 基于 HTML Parser 的 Web 信息提取系统的设计和实现[J]. 电脑知识与技术,2007,7(4):970-971.
- [8] HTMLParser[EB/OL]. 2006-09-17[2012-03-12]. <http://htmlparser.sourceforge.net>.
- [9] 任仲晟,薛永生. 基于页面标签的 Web 结构化数据抽取[J]. 计算机科学,2007,34(10):133-136.
- [10] Lin Shan, Hu Yanzhong. An Approach of Extracting Web Information Based on HTMLParser[C]//Second International Conference on Information Technology and Computer Science. [s. l.]:[s. n.],2010.
- [11] 蔡捷飞,陈泓泓,梁志宏,等. 主题型网页发现以及网页内信息块发现[EB/OL]. 2010. <http://www.doc88.com/p-64281396657.html>.
- [12] Watt A. 正则表达式入门经典[M]. 北京:清华大学出版社,2008.
- [13] 李 丹. 基于序列比对的动态 Web 信息抽取算法研究[D]. 长春:吉林大学,2009.

(上接第 53 页)

4 结束语

文中讨论了一种利用 Markov 随机场下构建的多变量高斯模型来对过分割后的数字图像的特征向量进行建模,通过对模型的训练学习得到模型的参数向量,实现了单幅数字图像在多尺度空间下的场景深度的估计,在此基础上统计了该方法在一些场景下的误差。

实验证明,该方法可以有效地估计场景的深度值,且证明了随着尺度空间的变大,会减小其所获得的深度值误差。理论上,改进图像的概率模型可以缩小误差,因此,之后在对模型的改进上仍有许多不足需要改进。

参考文献:

- [1] 朱伟利,朱 枫,郝英明. 基于单幅建筑物图像的三维信息提取[J]. 仪器仪表学报,2008,33(29):33-40.
- [2] 吴凤和,张晓峰,施法中. 单幅图像三维表面重建算法的研究与实现[J]. 计算机应用,2009,15(12):56-62.
- [3] 赵冬斌,陈 强,陈善本. 由单幅图像恢复物体三维形状的应用研究[J]. 光学技术,2001,11(9):78-82.
- [4] 高月芳,罗 飞,曹建忠. 由单幅二维灰度图像重构物体表

面形状[J]. 计算机应用,2007,36(7):15-21.

- [5] 李 波,王祥凤,李本山. 基于单幅图像的三维重建技术[J]. 信息与电子工程,2006,4(4):71-77.
- [6] 孙宇阳. 基于单幅图像的三维重建技术综述[J]. 北方工业大学学报,2001,23(1):9-13.
- [7] Sudderth E B, Torralba A, Freeman W T, et al. Depth from familiar objects: A hierarchical model for 3D scenes[J]. Computer Vision and Pattern Recognition (CVPR), 2006, 24(3): 229-237.
- [8] Saxena A, Schulte J, Ng A Y. Depth estimation using monocular and stereo cues[C]//International Joint Conference on Artificial Intelligence (IJCAI). [s. l.]:[s. n.],2007:112-118.
- [9] Felzenszwalb P F, Huttenlocher D P. Efficient graph-based image segmentation[J]. International Journal of Computer Vision, 2004, 59(2):167-181.
- [10] Saxena A, Chung S H, Ng A Y. Learning depth from single monocular images[J]. Neural Information Processing System, 2005, 18(8):66-78.
- [11] Davies E R. Laws' texture energy in TEXTURE[M]//Machine vision: theory, algorithms, practicalities. San Diego: Academic Press, 1997.

单幅数字图像多尺度空间下的场景深度估计

作者: [陆振杰](#), [宋进](#)
作者单位: [上海大学 机电工程与自动化学院, 上海 200072](#)
刊名: [计算机技术与发展](#)
英文刊名: [Computer Technology and Development](#)
年, 卷(期): 2013(1)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_wjtz201301015.aspx