

ISSN 1009-3044

Computer Knowledge and Technology 电脑知识与技术

Vol.9, No.9, March 2013.

E-mail: eduf@dnzs.net.cn

<http://www.dnzs.net.cn>

Tel:+86-551-65690963 65690964

# 基于 Kinect 的无监督图像识别

王丁杰,周昌乐

(厦门大学 智能系,福建 厦门 361005)

**摘要:** 现有的物体识别方法大都是有监督学习,而且主要处理的是 2D 图像数据,由于这类数据给出的背景信息和物体信息少于 2.5D 场景中的信息,传统有监督的学习方法对于非特定物体的识别效果还是不尽如人意,更别说基于无监督学习的识别系统了。随着 Kinect 识别的出现,获取场景的 2.5D 信息变得很容易。设计了一个全新的基于无监督物体识别系统。首先用平面拟合和空间聚合把物体从场景中分离以后。然后用几何基元拟合分割技术把物体分割为不同部件,在此基础上构建物体的图模型,最后把图模型嵌入为向量空间上的点集合,使用扩展陆地移动距离算法计算物体间的相似度。从实验结果来看效果不错。

**关键词:** 2.5D; 物体识别; kinect; 无监督; 几何基元拟合; 扩展陆地移动距离

**中图分类号:** TP18 **文献标识码:** A **文章编号:** 1009-3044(2013)09-2219-04

## Unsupervised Based Object Recognition with Kinect

WANG Ding-jie, ZHOU Chang-le

(Department of Artificial Intelligence, Xiamen University, Xiamen 361005, China)

**Abstract:** Most object recognition systems are based on supervised learning, and dealing with 2D image which gives little information about the background and foreground. Even state of art supervised object recognition can't promise good result, let alone unsupervised based system. With the advent of Kinect, it's more easy to get 2.5D data. This paper proposes an innovative unsupervised based object recognition system. The first step is to fit plane and use space based clustering to extract objects from the scene. The second step is segmenting objects based on geometry primitive fitting, constructing graph model for each object based on the segmentation result. Finally, embed the graph model to normal vector space and calculate the similarity between objects with extended Earth Mover's Distance. The experiment gives a good result.

**Key words:** 2.5D; object recognition; unsupervised; geometry primitive fitting; extended EMD

物体识别一直是机器人学科、计算机视觉和人机交互领域里的一个基础性工作。物体识别按表示方式主要可以分成两大类:基于部件模型的物体识别<sup>[3-4]</sup>和基于外观的物体识别<sup>[5-6]</sup>。从 2007 年开始,基于部件模型的物体识别技术日趋主流。

尽管基于部件模型的物体识别系统取得了很大的成功,但是绝大多数部件模型都有两个前提:一是的模型结构特征是事先指定的;二是训练样本中需要对物体的位置做出标记,并且告知物体的类型。虽然这种有监督的学习方式也是人类认知世界的一个过程,但是人类同样也可通过无监督学习的方式来认知这个世界。

随着 2010 年 Kinect 出现,获取场景的深度信息成为可能,物体和场景的分离变得更加容易。新技术的出现,给物体识别技术带来新的方法和视角。我们设计的物体识别系统就基于场景的 2.5D 信息。

我们设计了一个全新的无监督、增量式物体识别框架,利用场景中的点云信息来识别物体。它可被用于机器人中,用于探索未知的环境,模仿人类认识现实环境。

我们把场景中物体分为支撑体和被支撑体,比如桌面和桌面上的物体。主要步骤包括以下几个:

1) 物体提取。首先我们用 RANSAC 提取场景平面,删除场景平面中的支持平面,再用空间欧式距离聚类的方法,分割出被支撑物体。

2) 对每个被支撑物体用基于基元(三维球体,圆柱体,平面)的多层次分割方法建立物体模型。

3) 利用前面的分割结果构建物体的图模型<sup>[7]</sup>。

4) 用度量多维标度法把该图模型转化到二维向量空间中,获取每个物体的向量模型。

5) 用扩展陆地移动距离算法估计物体的向量模型间的距离,以此作为物体间的相似度。

收稿日期:2013-03-22

基金项目:国家自然科学基金(61273338)

作者简介:王丁杰(1985-),男,福建福清人,硕士,主要研究人工智能,图像处理;周昌乐,教授,博士。

本栏目责任编辑:唐一东

人工智能及识别技术 2219

### 1 物体提取

我们可以利用深度信息来把物体分成支撑体如桌面和挂着很多画的墙壁,被支撑体比如桌面上得物体和墙上的画。

我们用 Kinect 摄像头获取数据,并通过 OpenNI 接口得到数据的点云信息。然后用随机样本一致性算法从场景中提取所有的平面,对任意平面,用空洞的数量(因被支撑物和平面的分离产生)来判定该平面是否是支撑面。

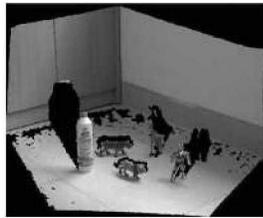


图1 一个简单的场景

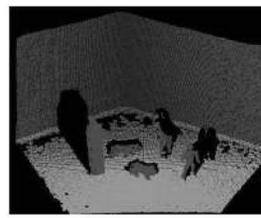


图2 深色部分为抽取到的物体,浅色部分为支撑平面

在支撑物被抽取后,会造成被支撑物在空间上的分离。这时我们可以用基于空间欧式距离的聚类方法,把和支撑面连接的每个物体提取出来,为进一步的处理做准备。图2中绿色部分为支撑平面,红色部分为提取到的在空间上分离的物体。

### 2 物体模型构建

#### 2.1 物体分割

基于部件的物体结构分解和人类看待物体的方式有很多共同之处。我们采用的分割方法基于 Marco 的层次几何基元拟合分割,分别用了三种基元:平面,球体,圆柱体。我们通过设置最终的分割部件的数目  $\beta$  来控制分割结果。



图3 玩具奥特曼,狮子,老虎及其对应的分割结果( $\beta=6$ )

#### 2.2 图模型构建

对于任意的物体  $x$ ,我们先用一个无向图来描述一个物体。 $x$  首先被逐层的分割成不同的部件。然后用[2]中的图结构来记录模型的拓扑结构和几何特征。图中的边表示部件间的父子关系或则兄弟关系。和节点相关的属性包括:该点和父节点的相对距离和方向,该节点的几何形状信息如外形类别,大小比例等。

由 EMD 的距离公式可知,权重接近的点之间如果相对距离越小,计算出来的 EMD 距离就越小。由于扩展 EMD 只对点集进行整体的移动,不会改变点集内部的相对位置。为了保证来自不同点集的权重相似的点尽可能的对齐,要求图模型到点集的嵌入过程要产生稳定的相对位置。否则 EMD 计算出来的距离不能稳定和真实的反映物体间的相似度。考虑如下两个图结构:

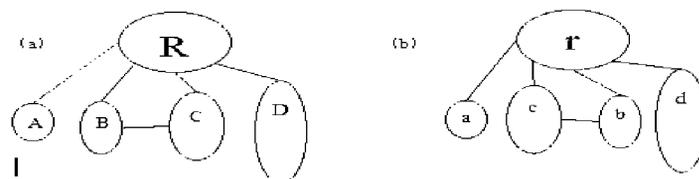


图4 两个相似物体的可能图模型

当它们被转化到向量空间后,一个好的对齐会是  $R-r, A-a, B-b, C-c, D-d$ 。如果我们设置上述图中的边权重为常数 1,那么在上述图在映射到向量空间后,可能有错位对齐。证明:假设图模型(b),映射到向量空间中的点集  $V$ ,点集的数量  $|V|=n+1$ 。对于特定点集中的某个点  $x$ ,它的相对距离向量  $d_x=(d_{x_1}, \dots, d_{x_n})$  中的每个元素表示该点到集合  $V$  上其它点  $y_i$  的距离,其中  $(1 \leq i \leq n)$ 。如果存在某个点  $z$ ,它的相对距离向量为  $d_z=(d_{z_1}, \dots, d_{z_n})$ ,且  $d_{z_k}=d_{x_k}(1 \leq k \leq n)$ 。那么  $d_x=d_z$ ,意味着这两个点在向量空间中的位置是可交换的,这是由图嵌入(见 4.1)过程决定的。所以对齐的时候就可能出现  $B-c, C-b$ 。因为  $c$  和  $b$  在向量空间中的相对距离向量相同。而这显然不是最佳的对齐方式。为了保证点集内部点间相对位置的更加稳定,对于图模型中的任意父节点  $p$ ,我们让父子边(部件的层次分割)的权重由 1 开始随子节点重量(其属性和基点属性的距离)的递增方向递增,增量为 1。兄弟边(部件间的相邻关系)的权重为常数 1。此外我们保证兄弟节点的重量不同。这样避免了同一个图中存在两个不同节点在嵌入到向量空间时由于位置可交换导致内部结构不稳定,从而可能发生错位对齐。

### 3 图嵌入和匹配

#### 3.1 图嵌入到向量空间

图嵌入是一种图匹配的有效手段。它能把图信息转化成向量空间上的点集合,其中每个点对应于图中的一个节点,点之间的

欧式距离对应于图中相应图节点间的距离,这里我们用最短路径表示图节点间的距离。考虑一个物体的图模型  $G$  有节点集  $\{v_i | i \leq j \leq n\}$ , 那么它的差异矩阵为  $\{D_{ij} = \text{short\_path}(v_i, v_j) | 1 \leq i, j \leq n\}$ 。我们用 Johnson 算法来计算节点对间的最短距离,其时间复杂度为  $O(2V \log V + VE)$ 。接着我们用度量多维标度法来把图嵌入到一个二维的向量空间中。对于任意两个图节点  $v_i$  和  $v_j$  间的最短路径  $d_{ij}$ , 其在向量空间中的对应节点间的欧式距离为  $f(\delta_{ij})$ , 度量多维标度法保证  $F_s$  最小。

$$F_s = \frac{\sum_{i < j} (f(\delta_{ij}) - d_{ij})^2}{\sum_{i < j} d_{ij}^2}$$

### 3.2 点集匹配

当图被嵌入到一个向量空间以后,图匹配问题就转化为向量空间中点集合的匹配问题。

陆地移动距离(EMD)源于运输问题,在图像处理中被广泛的用于解决这类带有权重的点集合间的匹配问题。两个点集间的移动距离越小,表示它们之间越相似。假设  $P = \{(p_1, w_{p_1}), \dots, (p_m, w_{p_m})\}$  为供应者,接收者  $Q = \{(q_1, w_{q_1}), \dots, (q_n, w_{q_n})\}$ , 其中  $m > n$ 。  $p_i$  和  $q_j$  表示点在向量空间中的坐标,  $w_{p_i}$  和  $w_{q_j}$  表示点的重量,这里我们用该点特征向量和基点向量的距离来表示,我们用根节点对应的特征向量最为基点向量。  $D = [d_{ij}]$  为距离矩阵,  $d_{ij}$  表示  $p_i$  到  $q_j$  的距离。我们要找出流  $F = [f_{ij}]$ , 使运输总代价  $EMD(P, Q)$  最小。这里  $f_{ij}$  表示  $p_i$  到  $q_j$  的流量。

距离公式为:

$$EMD(P, Q) = \frac{\sum_{i=1}^m \sum_{j=1}^n d_{ij} f_{ij}}{\sum_{i=1}^m \sum_{j=1}^n f_{ij}}$$

并且满足以下约束:

$$f_{ij} \geq 0 \quad 1 \leq i \leq m, 1 \leq j \leq n \quad (1)$$

$$\sum_{j=1}^n f_{ij} \leq w_{p_i} \quad 1 \leq i \leq m \quad (2)$$

$$\sum_{i=1}^m f_{ij} \leq w_{q_j} \quad 1 \leq j \leq n \quad (3)$$

$$\sum_{i=1}^m \sum_{j=1}^n f_{ij} = \min(\sum_{i=1}^m w_{p_i}, \sum_{j=1}^n w_{q_j}) \quad (4)$$

1)表示只能由供应者流向消费者。2)表示  $p_i$  的流出量不能超过该点的重量。3)  $q_j$  的接受量不能超过其容量限制。4)总的流量为两堆土中的最小值。

前面描述的 EMD 要求两个向量集合在空间上对齐。为了解决这个问题 Cohen 和 Guibas [11] 扩展了 EMD 算法,动态的在匹配的过程中对点集进行整体移动调整,从而找出最佳的匹配结果,我们称此为扩展 EMD 算法。Cohen 和 Guibas 用迭代的方式计算陆地移动距离的最小值。首先从一个最初的位置开始,扩展 EMD 计算最优的流向  $F_k$ , 在此流向的前提下计算一个更优化的空间坐标转化矩阵  $G_k$ , 应用此转化于其中一个点集,使其整体发生移动或旋转,进而使 EMD 距离减小。在  $G_k$  的前提下,计算出最优的  $F_{k+1}$ , 如此迭代,直到 EMD 值达到稳定值。图 5 是玩具狮子和老虎在向量空间中的扩展 EMD 匹配结果,我们可以看到两个点集在空间位置上有比较好的对齐结果,在此基础上计算出来的陆地移动距离能更好的体现两个点集间的距离,即图像间的相似度。

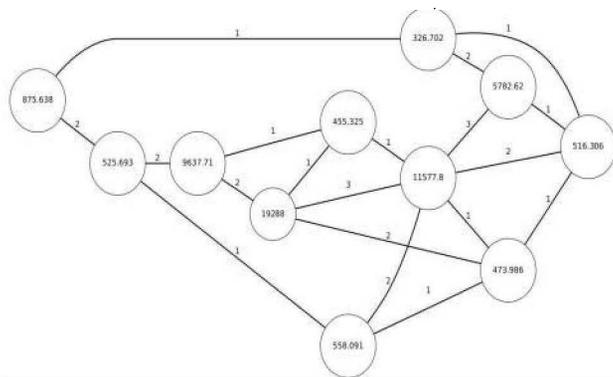


图 4 玩具狮子的图模型

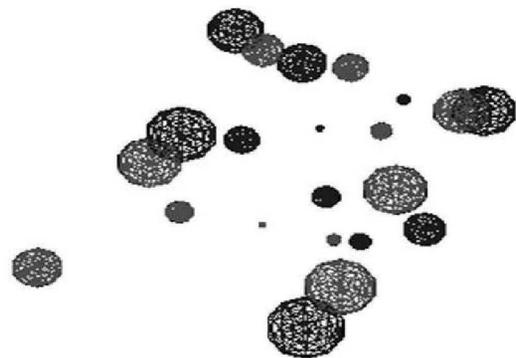


图 5 玩具狮子和老虎在向量空间中的扩展 EMD 匹配

## 5 实验结果

我们采集了每个物体 20 个视角的 2.5D 照片。设置分割结果部件参数  $\beta = 6$ 。考虑两个物体 A 和 B 间的距离,对于 A 的任何一个视角的 2.5D 照片,计算出它与 B 的所有视角照片间的距离,其中最小的那个为 A 到 B 的单视角最小距离。A 和 B 所有视角最小距离的和,以灰度度量的形式表示为图 6 灰度表中的格子。图 6 中任意一个格子对应于物体间的所有视角最小距离和。灰度越深的格子表示对应的两个物体越相似,白色的相反。为了了解物体视角对识别的影响,我们选择狮子奥特曼和狮子分别在在  $0^\circ, 30^\circ, 60^\circ$  度视角下的 2.5D 照片,来比较它们间相似度。由图 7 可看出,同类物体间视角差别越小,相似度越高。非同类物体间的视角差别对

识别的影响比较不稳定,这可能是因为从Kinect获取的数据有误差,不能准确反映物体的几何模型数据。



图6 五种物体间的相似度

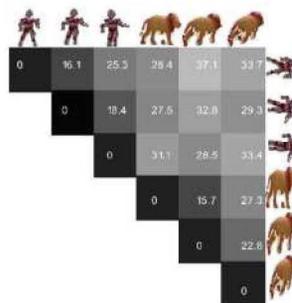


图7 三个视角的玩具奥特曼和狮子的相似度

## 6 结论

当今2D图像识别技术的识别率不高的很大一个原因是2D图片无法提供足够的背景和前景的信息。我们借助从Kinect获取的2.5D信息,设计了一个全新的无监督图像识别系统,该系统利用深度信息提取场景中的前景物体,然后再计算物体间的相似度。这样避免了背景信息对前景物体识别的干扰,有效的提高了识别率。该系统的缺点是假设所有的物体在空间上是分离的,这样如果有两个物体是相挨的,那么会被当做一个物体处理。从理论上说,无论是我们采用的图模型还是扩展EMD都有部分匹配的特性,对于物体的缩放,旋转也有一定的鲁棒性。如果部分匹配能取得好的效果,那么可以在一定程度上解决前面所述的不足。我们将在接下来的工作中探索Kinect的数据误差对鲁棒性和部分匹配的影响。

## 参考文献:

- [1] Shokoufandeh A, Dickinson S, Jönsson C, et al. On the representation and matching of qualitative shape at multiple scales[J]. Computer Vision—ECCV 2002, 2002: 6–14.
- [2] Dickinson S J, Pentland A P, Rosenfeld A. From volumes to views: An approach to 3-D object recognition[J]. CVGIP: Image Understanding, 1992, 55(2): 130–154.
- [3] Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part-based models[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010, 32(9): 1627–1645.
- [4] Leonardis A, Bischof H. Dealing with occlusions in the eigenspace approach[C]//Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on. IEEE, 1996: 453–458.
- [5] Turk M, Pentland A. Eigenfaces for recognition[J]. Journal of cognitive neuroscience, 1991, 3(1): 71–86.
- [6] Attene M, Falcidieno B, Spagnuolo M. Hierarchical mesh segmentation based on fitting primitives[J]. The Visual Computer, 2006, 22(3): 181–193.
- [7] Demirci M F, Shokoufandeh A, Keselman Y, et al. Object recognition as many-to-many feature matching[J]. International Journal of Computer Vision, 2006, 69(2): 203–222.
- [8] Cox M, Cox T. Multidimensional Scaling[M]. Chapman and Hall, London, 1994.
- [9] Rubner Y, Tomasi C, Guibas L J. A metric for distributions with applications to image databases[C]//Computer Vision, 1998. Sixth International Conference on. IEEE, 1998: 59–66.
- [10] Rubner Y, Tomasi C, Guibas L J. The earth mover's distance as a metric for image retrieval[J]. International Journal of Computer Vision, 2000, 40(2): 99–121.
- [11] Cohen S, Guibas L. The earth mover's distance under transformation sets[C]//Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on. IEEE, 1999, 2: 1076–1083.
- [12] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6): 381–395.