

# 基于多视角距离图象的三维物体建模 及其在识别中的应用<sup>1)</sup>

王晓军 袁梅 吴立德

(复旦大学计算机科学系 上海 200433)

## 摘 要

模型的表示和构建是基于距离图象三维物体识别技术中的关键模块之一。针对已有方法存在的若干问题,提出一个新的综合多个视角距离图象的三维物体模型表示策略和增量式的模型习得算法,并将该模型表示用于三维物体识别中。实验结果验证了算法的有效性。

**关键词:** 多视角距离图象,三维物体建模,三维物体识别,微分和矩不变量,属性图表示。

## 1 概述

模型的表示和构建是基于距离图象三维物体识别中的关键之一。一般来说,模型的表示应该与场景中的物体表示相类似或易于相互转换。现有的模型表示可分成两类:一类是基于CAD的模型表示<sup>[1]</sup>,它以物体为中心,稳定性好,信息量丰富,但是从模型的CAD表示到用于识别的表示之间的转换并不是一件容易的事,而且CAD比较适合于工业部件等比较规则物体的建模,对于自然景物(比如树木、鸟等)的建模尚存在困难;还有一类模型表示是多视角的2D或2.5D表示<sup>[2]</sup>,即某个模型由一组2D或2.5D表示组成,其中每一个表示对应于从某个特定的视角所得到的这个模型的表示,它虽然较易获取,也能完成识别,但识别过程搜索空间大(每一模型有一组表示),效率不高。人类对物体的视觉形象并不等于从某一特定的投影方向看到的形象,对于物体的感知具有整体性,能把从各个不同视角得到的同一物体的不同感知结合起来,形成各物体的总体形象——模型<sup>[3]</sup>。根据这一原理,本文将各个视角的表示综合起来,形成一个统一的模型。

## 2 单个视角的表示

### 2.1 表面块基元的不变性质

根据人类视觉感知,物体总是被看作三维的,且其形状具有恒常不变性,因此选择表示时应该采用那些对视角变化不敏感的稳定特征;同时为了便于识别和定位,又应包含足够丰富的信息量。

1) 高校博士点基金和国家自然科学基金资助项目。  
本文于1993年11月30日收到

### 1) 微分不变量

广泛使用的高斯曲率和平均曲率的符号(简称 KH 符号)是一种微分不变量,它为表面基元提供了与视角无关的定性性质。根据 KH 符号的性质,可将表面基元分成八类<sup>[4]</sup>: (1)凸峰面,(2)凸脊面,(3)鞍脊面,(4)极小面,(5)鞍谷面,(6)凹脊面,(7)凹峰面,(8)平面。

### 2) 矩不变量

矩不变量最早是由 Hu 在 1962 年引入的<sup>[5]</sup>, Sadjadi 和 Hall 则将其从 2D 推广到 3D<sup>[6]</sup>。已有的方法是将 3D 矩应用于整幅图象,因此不够合理。本文则将 3D 矩不变量应用于分割后图象中的各表面块。由于高阶矩不变量(矩阶数 $>2$ )易受量化效应和测量误差的影响,因此选取 0 阶和 2 阶矩作为表面块的定量特征,它们在刚性变换下是不变的。

$$0 \text{ 阶矩 } I_0 = M_{000},$$

$$2 \text{ 阶矩 } I_1 = M_{200} + M_{020} + M_{002},$$

$$I_2 = M_{200}M_{020} + M_{200}M_{002} + M_{020}M_{002} - M_{101}^2 - M_{110}^2 - M_{011}^2,$$

$$I_3 = M_{200}M_{020}M_{002} - M_{002}M_{110}^2 + 2M_{110} \times M_{101}M_{011} - M_{020}M_{101}^2 - M_{200}M_{011}^2.$$

式中  $M_{pqr} = \sum (X_i - \bar{X})^p (Y_i - \bar{Y})^q (Z_i - \bar{Z})^r$ 。

## 2.2 单视角的属性图表示

以面为基元的三维物体可以用一种属性图  $G = \langle V, E \rangle$  来表示,其中  $V = \{v_1, v_2, \dots, v_n\}$ , 是结点的集合,每个结点对应着一个面;  $E = \{l_{ij} \mid l_{ij} = (v_i, v_j)\}$ , 是边集,  $l_{ij}$  表示面之间的关系。

### 1) 结点属性

(1) 度数  $d$ ; (2) 面的平均法向  $N$ , 质心  $C$ ; (3) 面的定性属性  $a \in \{0, 1, 2, 3, 4, 5, 6, 7\}$ , 其中 0—7 分别对应于表面片的 8 种类型; (4) 面的定量属性  $I_0, I_1, I_2, I_3$ 。

### 2) 关系属性

$G$  中的每个关系,即  $E$  中的每条边  $l_{ij} = (v_i, v_j)$ , 表示  $v_i$  与  $v_j$  相邻,即  $v_i$  与  $v_j$  之间至少存在一条公共的折边。

对于场景,假定各遮掩物体彼此不接触,这样场景  $S$  可用若干个连通子图表示,  $S = \{G_1, G_2, \dots, G_r\}$ , 其中每个  $G_i$  对应着一个连通子图。

## 3 基于多视角距离图象的模型表示和习得

### 3.1 模型表示的形式描述

与上节的物体单视角属性图表示类似,基于多视角距离图象的模型也是一种属性图。它以某一视角为基准,先得到这一基准视角下的表示作为基础,然后再与其它视角的表示通过学习过程进行合并而得。其形式描述如下:

令  $G = \{G_0, G_1, \dots, G_n\}$  是模型物体在各个视角下的表示集合;令  $(R_i, T_i)$  表示视角  $i$  到视角 0 的变换,其中  $R_i$  为旋转矩阵,  $T_i$  为平移向量,  $G'_i = f(G_i)$  是  $G_i$  经  $(R_i, T_i)$  变换得到的图表示,  $G'_i = \langle V'_i, E'_i \rangle$ 。易知  $\forall v' \in V'_i, \exists v \in V_i$ , 使得  $v'$  的属性满足

(1)  $d' = d$ ; (2)  $N' = RiN$ ,  $C' = RiC + T$ ; (3)  $a' = a$ ; (4)  $I'_j = I_j$ ,  $j = 0, 1, 2, 3$ .

$\forall l' = (v'_p, v'_q) \in E'_i$ ,  $\exists l = (v_p, v_q) \in E_i$ , 其中  $v'_p, v'_q$  分别与  $v_p, v_q$  对应. 显然  $G'_0 = G_0$ .

令  $M$  为模型表示, 则形式地有

$$M = (V_M, E_M) = \cup f(G_i) = \cup G'_i,$$

其中  $V_M = \cup V'_i$ ,  $E_M = \cup E'_i$ .

易知,  $|V_M| \leq \sum |V'_i|$ ,  $|E_M| \leq \sum |E'_i|$ .

这种模型表示揭示了模型物体在不同视角下, 各种表示之间的内在联系, 而且一个模型物体只对应于一个模型表示, 从而有效地控制了识别过程中的搜索空间, 提高了识别效率.

### 3.2 基于表面块对应的三维变换参数求解

令  $S, S'$  分别表示视角  $i$  和视角  $0$  下模型物体的对应 3D 表面块,  $S = \{(P_i, N_i) | i = 1, 2, \dots, n\}$ ;  $P_i$  是点的三维坐标,  $N_i$  是  $P_i$  点的法向};  $S' = \{(P'_i, N'_i) | i = 1, 2, \dots, n\}$ . 假设  $S$  与  $S'$  上各点的对应已知, 则

$$N'_i = RN_i, \quad (1)$$

求平均得 
$$\bar{N}' = R\bar{N}, \quad (2)$$

其中  $\bar{N} = \sum N_i/n$ .

若有  $L(L > 1)$  个面对应, 可利用最小平方求解  $R$

$$\min \sum_{j=1}^L \|\bar{N}'(j) - R\bar{N}(j)\|^2 + \sum_{i < j} \|C'(i, j) - RC(i, j)\|^2, \quad (3)$$

式中  $C(i, j) = C(i) - C(j)$ ; 在非退化情况下, 可用四元数方法线性求得  $R$  的唯一解<sup>[7]</sup>.

若  $L = 1$ , 则令

$$Q_i = P_i - C, \quad Q'_i = P'_i - C', \quad (4)$$

有 
$$Q'_i = RQ_i, \quad (5)$$

叉乘(1)式得

$$Q'_i \times N'_i = R(Q_i \times N_i). \quad (6)$$

求平均得

$$U' = RU. \quad (7)$$

式中  $U = \sum_{i=1}^n (Q_i \times N_i)/n$ . 相应地,  $R$  可由下述最小平方问题求得

$$\min \{\|\bar{N}' - R\bar{N}\|^2 + \|U' - RU\|^2\}.$$

进而可求得

$$T = 1/L \sum_{j=1}^L [C'(j) - RC(j)]. \quad (8)$$

需要强调的是, 在上述变换参数求取式(3), (8)中, 仅使用了表面块的整体信息, 而没有用到点对应的信息(对于实际距离图象而言, 精确点对应的获取是困难的). 当然, 在公式推导中作了点对点存在假设, 这实际上意味着用于变换参数求取的各表面块不能有遮掩存在, 在没有额外信息可以使用的情况下, 可使用下述方法来判别:

对于表面块对应  $(S, S')$ , 若存在  $i(0 \leq i \leq 3), I_i \cong I'_i$ , 则  $(S, S')$  不能用于变换参数的求取.

### 3.3 增量式的模型习得算法

根据图的属性和物体的刚性约束, 可以将各个视角下的表示综合起来, 得到模型的整体表示. 以下给出一个增量式的模型习得算法.

输入:  $G = \{G_0, G_1, \dots, G_n\}$ , 其中  $G_k(0 \leq k \leq n)$  对应于第  $k$  个视角下的图表示;  
输出: 模型的图表示  $M = (V_M, E_M)$ ;

过程:

步骤 1.  $M := G_0, k := 0, G := G - \{G_0\}$ ;

步骤 2. if  $G = \text{EMPTY}$  then 算法结束 else  $k := k + 1$ ;

步骤 3. 求出  $G_k$  与  $M$  中结点的对应关系

$$\text{Match-Set} := \{\langle v_{i1}^{(k)}, v_{i1}^{(M)} \rangle, \langle v_{i2}^{(k)}, v_{i2}^{(M)} \rangle, \dots\},$$

步骤 4. 根据  $\text{Match-Set}$ , 利用式(3), (8)求得变换参数  $(R(k), T(k))$ ;

步骤 5. 修改  $V_M$ . 对  $G_k$  中的每一个结点  $v^{(k)}$

if  $\text{Match-Set}$  中不存在结点对  $\langle v^{(k)}, v^{(M)} \rangle$  then

begin

由(3.1)节修改  $v^{(k)}$  的属性, 并记  $v'_{(k)} =$  修改属性后的  $v^{(k)}$ ,

$$V_M := V_M + \{v'\}, \text{Match-set} := \text{Match-set} + \{\langle v^{(k)}, v'_{(k)} \rangle\}$$

end;

步骤 6. 修改  $E_M$ . 对于  $G_k$  的边集  $E_k$  的每个元素  $l(p, q) = (v_p, v_q)$ , 根据  $\text{Match-Set}$  找到相应的  $v'_p, v'_q$ ,

if NOT (边  $(v'_p, v'_q) \in G_M$  的边集  $E_M$ ) then

begin

$$l'(p, q) = (v'_p, v'_q), E_M := E_M + \{l'(p, q)\}$$

end;

步骤 7.  $G := G - \{G_k\}$ , goto step 2..

## 4 应用于三维物体的识别

对于一幅场景, 在获取了属性图表示  $GS = \{GO_1, GO_2, \dots, GO_N\}$  (每个连通子图对应着一个物体)后, 识别和定位就是要在模型库中寻找与场景中物体  $GO_i$  相匹配的模型表示  $GM_i(i = 1, 2, \dots, N)$ , 并求取模型到物体的变换参数  $\langle R_i, T_i \rangle$ . 其实质是在满足刚性约束的条件下, 寻找  $GO_i$  的最大子图与  $GM_i$  的最大子图同构.

### 4.1 识别和定位方法

识别和定位过程包括如下三个模块:

#### 1) 粗匹配模块

这是选择候选模型的过程. 对于场景中的某个物体  $GO$ , 如果模型  $GM$  满足下列条件就把它选为  $GO$  的候选模型:

- (1)  $GO$  的  $KH$  符号种类  $\leq GM$  的  $KH$  符号种类;
- (2)  $GO$  的三维面积之和  $\leq GM$  的三维面积之和;
- (3)  $GO$  的结点度数之和  $\leq GM$  的结点度数之和;
- (4)  $GO$  的结点数  $\leq GM$  的结点个数.

## 2) 细匹配模块

这是求取物体  $S$  和候选模型  $M$  之间最大同构子图的过程, 包括两个子过程:

(1) 寻找可能的结点匹配对. 若  $S$  中的任一结点  $v^S$  与  $M$  中的结点  $v^M$  具有相同的定性性质和相似的定量属性, 则  $\langle v^S, v^M \rangle$  是一候选对应;

(2) 求取子图同构. 采取深度优先的树搜索方法, 对  $S$  中的每个结点寻找与之匹配的  $M$  中的结点(允许空匹配存在), 并根据刚性约束和关系约束来剪除不可能的匹配路径. 该过程的输出是匹配树中最长的一条满足刚性约束和关系约束的匹配路径, 以及相应的变换参数.

## 3) 决策模块

设物体  $S$  有  $p$  个候选模型  $M_1, M_2, \dots, M_p$ , 利用细匹配模块可获得  $S$  的  $p$  个最大同构子图, 记为  $SS_1, SS_2, \dots, SS_p$ ; 则当  $M_q$  满足  $|SS_q| = \max\{|SS_1|, |SS_2|, \dots, |SS_p|\}$  时选取  $M_q$  作为  $S$  的最佳匹配, 其中  $|SS_i|$  表示子图  $SS_i$  中结点的个数.

## 4.2 识别和定位实验

已完成一个基于距离图象的三维物体建模和识别实验系统, 下面介绍一个实验结果, 其中定位信息  $R$  和  $T$  分别表示将基准视角下的模型表示变换到待识别物体的旋转矩阵和平移向量.

图 1(a) 是一个由两个物体所组成的场景, 图 1(b) 是场景的关系图表示, 图 1(c), (d) 分别表示基准视角下的模型经过  $R, T$  变换后与待识别物体 1 和待识别物体 2 的重合情况, 不难看出挖孔台体与其对应模型的吻合比较精确, 而飞机的定位精度则不够理想. 这主要是由于挖孔台体的输入数据是生成的距离图象, 而飞机的输入数据是实际的距离图象, 所以后者存在着较大的测距和量化误差. 这些误差极可能破坏刚性约束的条件, 使物体的运动变成非严格的刚体变换, 而且由于在场景合成过程中对实际图象作了压缩, 更增大了其输入数据的误差, 从而降低了飞机的定位精度. 在识别过程的最后, 由于在模型库中找不到与物体 3 相匹配的模型, 所以需要根据物体 3 中的面与变换后模型中某些面的匹配情况将它归入物体 2. 以下是识别的结果:

物体 1 (挖孔台体)

$$R = \begin{pmatrix} 0.94966 & -0.30894 & 0.05198 & -0.14979 \\ 0.29888 & 0.94317 & 0.14523 & -77.49356 \\ -0.09390 & -0.12238 & 0.98803 & 29.41078 \end{pmatrix}, \quad T =$$

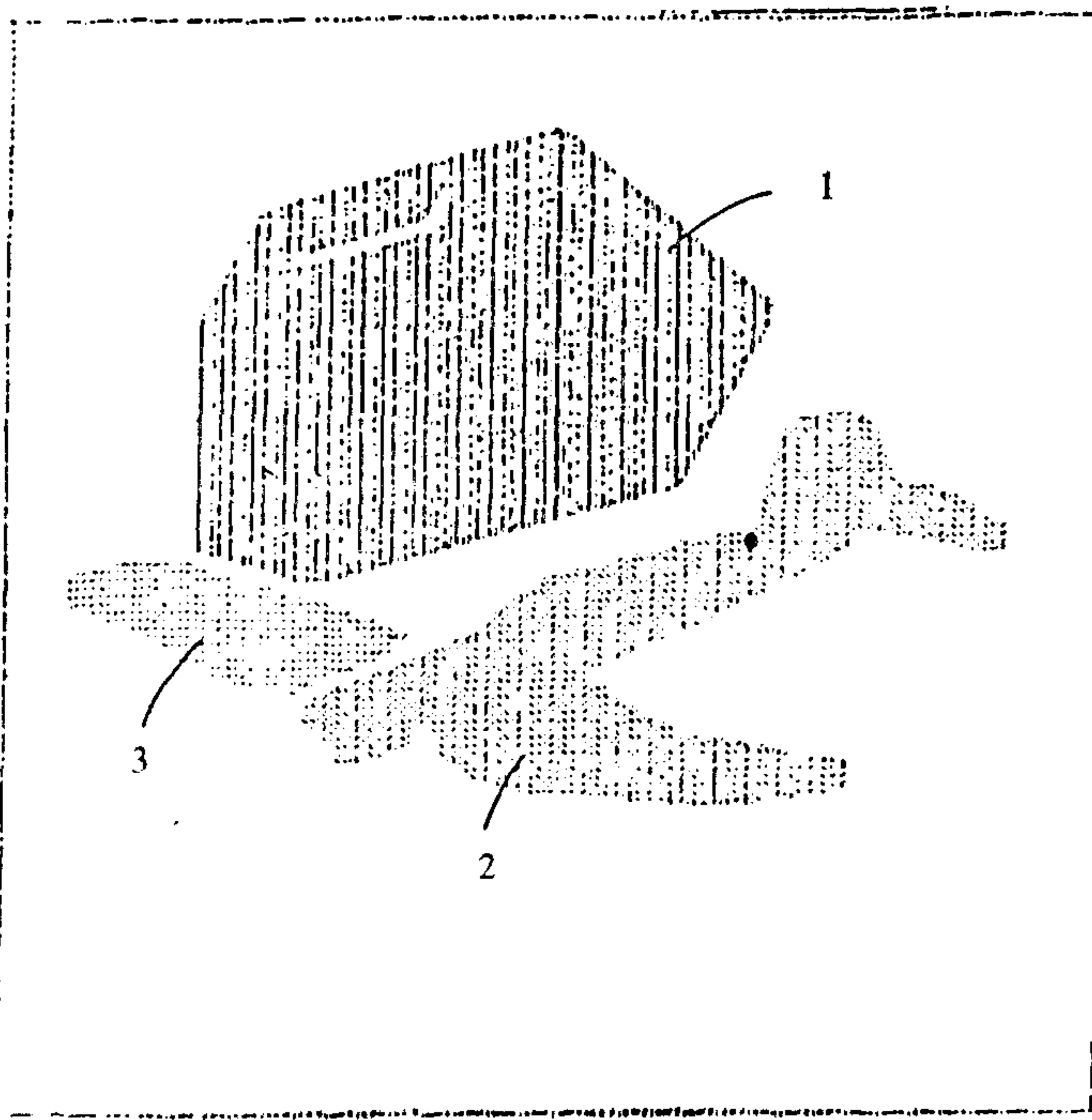
物体 2 和物体 3 (飞机)

$$R = \begin{pmatrix} 0.87366 & -0.25661 & -0.41338 & 405.53149 \\ 0.42484 & -0.01176 & 0.90519 & -535.32778 \\ -0.23714 & -0.96644 & 0.09874 & 1312.93564 \end{pmatrix}, \quad T =$$

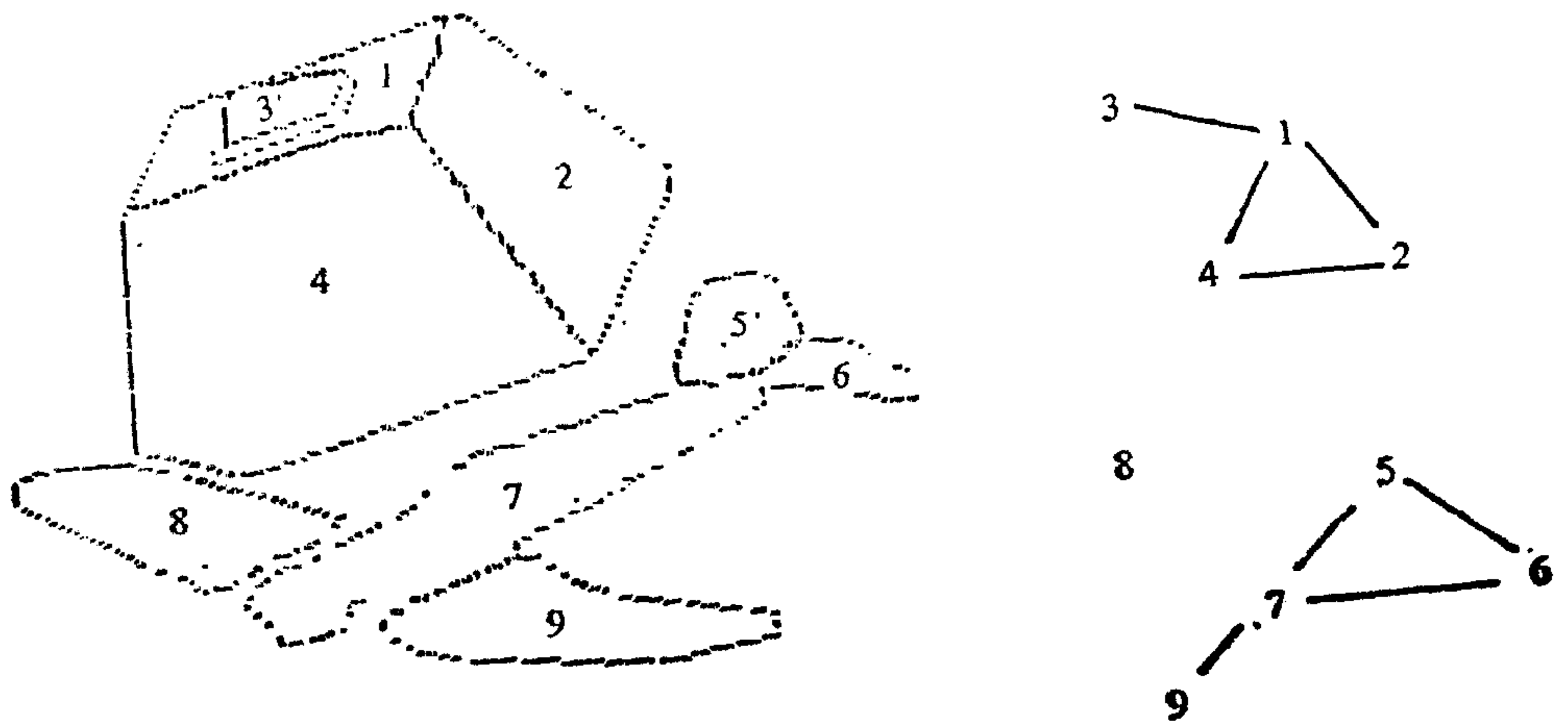
## 4.3 方法性能比较

FAN 等在文献[2]中提出的一种基于面基元的距离图象识别定位方法, 是同类算法中较有代表性的. 与它相比, 本文方法在以下方面有更好的性能:

1) 文献[2]中将表面块的 3D 可见面积(即 0 阶 3D 矩)作为表面基元的定量性质, 本文除了 0 阶 3D 矩外, 还引入了 3 个 2 阶 3D 矩不变量, 从而为面基元的粗匹配引入了强有力的定量约束信息.



(a)



(b)

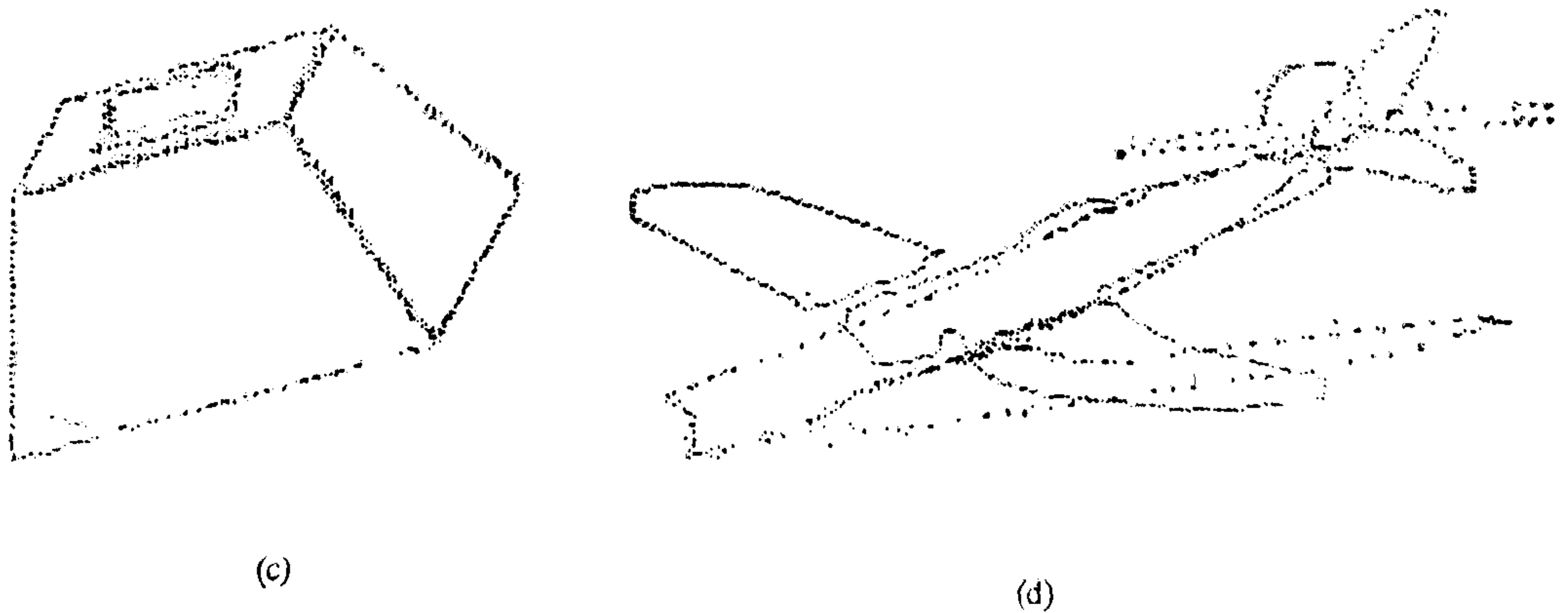


图 1

2) 文献[2]中的系统,使用一组不同视角的属性图表示来描述每个模型,这样模型库的存贮空间较大,而且在进行物体识别和定位时,须对每个模型的所有视角的表示都搜索完后才能确定其匹配,因而开销较大。本文提出的模型综合表示(称为 IMD)策略,将模型的多个视角的图表示合并成一个,从而使得一个模型物体对应于一个模型表示。这既压缩了模型库的存贮空间,也降低了匹配的开销。

3) 计算定位变换的方法是新的。

致谢 本文所用距离图象由南加州大学 G. Medioni 教授提供,特此致谢。

### 参 考 文 献

- [1] Kim W and Kak AC. 3-D Object recognition using bipartite matching embeded in discrete relaxation. *IEEE T-PAMI*, 1991, **13**(3):224—251.
- [2] Fan T, Medioni G and Nevatia R. Recognizing 3D objects using surface descriptions. *IEEE T-PAMI*, 1989, **11**(11): 1140—1157.
- [3] Marr D, Vision WH. Freeman and Company. San Francisco, 1982.
- [4] Besl P J and Jain R C. Invariant surface characteristics for 3D object recognition in range images. *CVGIP*, 1986, 33: 33—80.
- [5] Hu M K. Visual pattern recognition by moment invariants. *IEEE T-IT*, 1962, **8**:179—187.
- [6] Sadjadi F A et al. Three dimensional moment invariants. *IEEE T-PAMI*. 1980, **2**:127—136.
- [7] Horn B K P. Robot Vision. MIT Press, 1986.

# ON THE CONSTRUCTION OF INTERGRATED MODEL DESCRIPTION FROM MULTIPLE VIEW RANGE IMAGES WITH APPLICATION IN THREE-DIMENSIONAL OBJECT RECOGNITION

WANG XIAOJUN    YUAN MEI    WU LIDE

(*Computer Science Department, Fudan University Shanghai 200433*)

## ABSTRACT

This paper addresses the problem of model representation and construction for three-dimensional object recognition from range images. First a new model representation scheme called intergrated model description (IMD) is proposed; then an incremental learning algorithm is developed for constructing the IMD from multiple view range images; finally the recognition algorithm is presented based on the IMD model base. Experimental results are given to demonstrate its effectiveness.

**Key words:** Range image, object modelling, object recognition, invariants, attributed graph.

**王晓军** 照片、简介见本刊第 18 卷第 4 期。

**袁 梅** 复旦大学计算机科学系助教，硕士。主要研究领域是模式识别，图象处理和计算机视觉。

**吴立德** 照片、简介见本刊第 18 卷第 4 期。