

基于人造物体直线段结构特征的不变性识别

危 辉 裘 禛 宇

(复旦大学计算机科学技术学院认知算法模型实验室 上海 200433)

摘 要 传统的用假设验证法进行三维物体识别的方法需要通过一组非线性方程组求解从模型到场景的坐标系变换,具有非常高的复杂度.文中提出了一种基于能够表明物体几何构造的直线段特征的人造物体识别方法,将假设验证法中对于全局坐标系变换的求解分散在各个平面单应性变换的求解中,降低了求解的复杂度.该方法首先利用几何不变量预匹配特征点,进而假设并求出场景和模型平面之间的单应矩阵,随后通过模型与场景之间直线段特征匹配的结果进行验证.实验证明,该方法能够快速准确地识别含有较多共面直线段特征的人造物体.

关键词 人造物体识别;直线段特征;假设验证法;平面单应性;特征匹配

中图法分类号 TP18 **DOI号**: 10.3724/SP.J.1016.2010.01088

Line Feature Based Man-Made Object Recognition with Invariance

WEI Hui QIU Zhen-Yu

(Cognitive Model and Algorithm Laboratory, School of Computer Science and Technology, Fudan University, Shanghai 200433)

Abstract The traditional three-dimensional object recognition method based on hypothesis and test needs to solve the coordinate transformation matrix from scene to model through a group of non-linear equations. Therefore, it has a very high complexity. This paper presents a man-made object recognition method based on the geometric feature of line characteristics, and disperses the overall coordinate transformation calculation in every local Plane Homography calculation, reduces the complexity of the solution. Its process: firstly pre-match the feature points using geometric invariants, then assume and solve the Plane Homography matrix between scenes to model. After that it matches the line segments on the homography plane, and by this verifies the assumption. Experiments prove that this method can rapidly and accurately identify man-made objects with coplanar line features.

Keywords man-made object recognition; line features; hypothesis and test; plane homography; feature matching

1 引 言

三维物体识别是计算机视觉领域中最活跃的研究分支之一,在机器人控制、自动导航、自动检测、装配任务与医学图像的分析等大量现实领域发挥核心作用.如精确制导武器需要运用三维物体识别技术

识别并跟踪攻击目标,装配或者包装机器人需要运用三维物体识别技术对产品的姿态进行定位.在过去的半个多世纪中,三维物体识别一直受到各国视觉研究者的广泛关注,并提出了众多的理论和方法.现有的识别技术可以从以下两个角度进行分类^[1]:

一种分类方法是根据识别所采用的不同特征,如一些全局特征:颜色、纹理、致密;或者局部特征:

点、线段。每种特征都有它的局限性,因此它们的应用领域也各不相同^[1]。另一种分类方法是根据物体模型所采用的不同表示方法^[2],如基于表面特征的表达^[3-4]、基于体积特征的表达^[5]、基于表面不连续性的表达^[6,8]、基于定位的表达^[9]、基于整体的表达^[10-11]等。这些方法的主要区别在于存储信息的方式不同。但一般而言,这些信息都是从图像中提取得到的^[1]。

物体的形状特征不受光照、表面纹理特征等条件的影响,因此在三维物体识别中得到了广泛的应用。Arbter 等对傅立叶描绘子进行了扩展,而使其具有平移、旋转、尺度和剪切不变性^[12]; Weiss 提出了关于任意 3 点、共线 3 点、任意 4 点、一条二次曲线和两条二次直线等仿射不变量^[13]; Quan 等提出在已知点对的对应关系的情况下可由未标定的 3 幅图像中对应的 6 对点来计算出 3 个透视不变量^[14]; Zisserman 等提出对于具有重复结构、旋转对称、管状表面及多面体等类似物体,可由单幅图像计算其透视不变量^[15]。形状特征的匹配方法可以分为以下几类^[15-16]:

基于解释树的方法。基于解释树的方法常用于表示场景特征和模型特征间可能的匹配,且树上从根到叶子的每条路径都表示相关问题的一个可能解^[3]。该方法根据几何约束,通过遍历树来求得场景特征和模型特征之间的一致性解释,但是,一般而言,该方法需要一组图像,不能用于对独立单幅图像的解释。

假设验证法,又称为对准(alignment)。该方法通过对图像场景特征的提取以及原先对物体模型特征的假设,得到一组超限定线性方程或非线性方程。再根据此方程组的解得到物体参考坐标系到场景参考坐标系的变换。最后,将模型变换到场景中,利用场景图像验证假设的正确性^[17-18]。该方法在场景较为复杂的情况下具有非常高的复杂度。

姿态聚类法,又称为广义霍夫变换。它是根据局部几何特征约束来进行匹配。从每一个成功的特征匹配可以计算出一个几何变换,然后在符合该几何变换的模型的位姿空间中进行累加。对于所有模型的位姿空间中最大的聚类点就是与场景一致的假设^[19]。该方法对于噪声较敏感,且搜索的代价较高。

基于图的识别方法。一般指特征关系图,图中的节点对应于场景和模型中的主要特征,而节点之间的弧则表示两个特征之间的关系。场景和模型之间的匹配均使用图论的匹配技术实现,如同构、同态

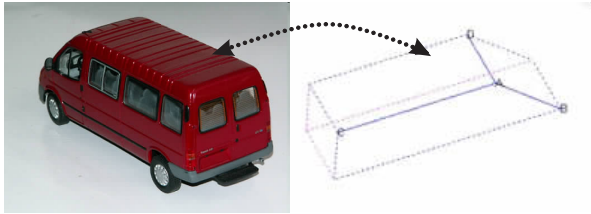
等^[20]。基于图的识别方法的主要困难在于难以处理物体遮挡问题和图匹配的大运算量问题。

常见的人造物体的几何构造一般由平面、圆面、直线等所组成,这些几何构造不随光照等环境因素而改变,反映了物体本身的特征。特别是含有较多共面直线段特征的人造物体,如大部分电器、家具、设备、器械、建筑、交通工具、军事装备等,在我们周围大量存在。针对此,本文提出了一种基于直线特征人造物体识别的新方法,通过对场景和模型图像中直线段的匹配识别人造物体。相对于基于特征点的识别,线段数量较少且具有较强的鲁棒性,因此在匹配过程中发生错误的概率较低。本方法采用假设验证法,在已知模型各个平面的直线特征及位置关系的情况下,通过对场景图像中不变性特征或者平面间所具有的不变性特征的分析,假设其中某些直线段的匹配,继而得到某一平面的匹配关系,并求出其单应矩阵。再将模型平面映射到场景,利用平面中直线段的匹配对假设进行验证。逐个识别所有平面。最后利用平面之间的不变性验证假设的正确性。相对于传统的假设验证法,本方法将假设验证法中对于全局坐标系变换的求解分散在各个平面单应性变换的求解中,降低了求解的复杂度,因此能够快速准确地识别物体。

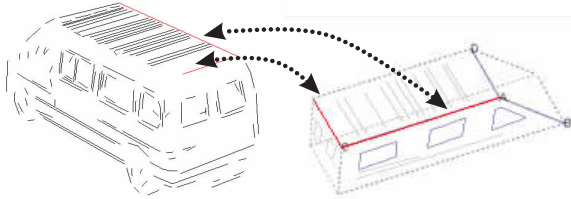
2 基本原理简述

整个识别算法的原理如图 1 所示。我们采用假设验证法,即我们已知某物体的一个或若干个标准图像,假设场景图像中含有某一待识别物体,我们试图在场景图像中找到该待识别物体的一些特征,以完成场景图像与标准模型的匹配。由于较点特征而言,边缘直线段是有明显几何意义的特征,数据整合程度较像素点大为提高,它们数量较少有利于加速匹配与组合,且具有较强的鲁棒性,因此我们利用边缘直线段作为主要特征对场景及标准模型之间的边缘直线段进行配准,如图 1(a)、(b)所示。配准的直线段越多则表明场景与标准模型越接近。

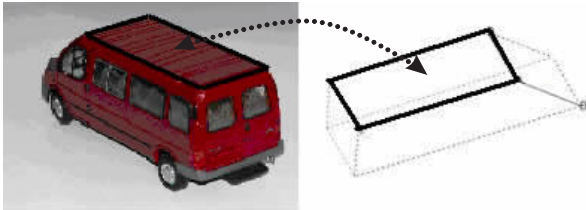
一般而言,人造物体具有多个结构面,如图 1(c)所示。利用这些平面在场景和模型中的单应性可以帮助完成配准,即场景图像中某一平面上的点在一个透视变换下能与模型图像中该平面上的点对应。对应关系称为单应矩阵,一般用 H 表示,是一个非奇异 3×3 的矩阵^[21]。若用 p_s 代表场景图像中的点,而 p_M 代表模型图像中的点,则有 $p_M \cong H p_s, \cong$



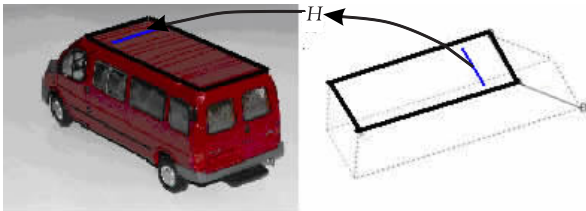
(a) 左边是待识别场景图像, 右边是标准模型, 我们希望完成他们之间可能的匹配



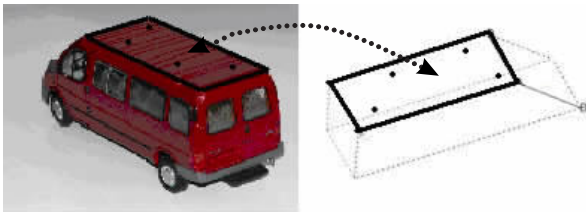
(b) 边缘直线段特征数量较少, 较点的匹配而言, 边缘直线段的匹配更易于实现



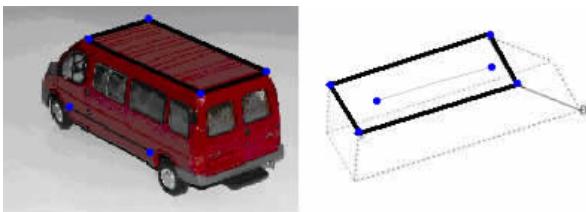
(c) 如果能确定物体的某个结构面的对应关系, 将有助于大大减少直线段特征的匹配数量



(d) 结构面的对应关系可以用单应矩阵 H 来表示. 若 H 被求出, 则我们可以将标准模型中的直线段映射到场景图像中, 帮助我们完成配准



(e) 求解单应矩阵 H 需要已知 4 个以上对应点, 且这四个点中任意三点不共线



(f) 透视投影的透视不变量, 如共面不变量或非共面不变量, 有助于我们找到点之间的对应关系, 从而求出单应矩阵. 图中为两相邻平面中 6 个点所组成的透视不变量

图 1 识别算法原理

表示在一个比例因子下的相等. 如果我们能够求出某一结构面在模型以及场景图像间的单应矩阵, 就能求出该结构面中每一条模型直线段在场景中的位置. 而这些位置能够帮助我们在场景中寻找合适的直线段进行匹配. 如图 1(d) 所示, 通过单应矩阵 H , 我们能够完成这个结构面上的每一条直线段的配准.

如何求取单应矩阵? 如图 1(e) 所示, 一般而言, 平面的单应矩阵可以由透视前后的两个平面上的 4 组对应点求出, 若有更多的点对, 则方程超定 (over-determined), 提供合理的最优解. 因此, 在求出平面的单应矩阵之前我们需要预先匹配两平面间 4 个或者 4 个以上的点^[21]. 但确定点的对应关系在通常情况下是困难的, 为了解决这个问题, 我们利用几何不变性来选取一些特殊点进行匹配.

几何不变性是物体在各种投影变换下保持不变的性质^[22], 在透视投影变换下保持不变的性质称为透视不变性, 更一般的, 在仿射变换下保持不变的性质称为仿射不变性. 最为基础的透视投影不变量是四点共线、四线共点的交比. 在此基础上衍生出其它诸如 5 条直线、相邻平面等构成的不变量, 众多学者对此进行了研究, 如 Zhu 等提出了一种相邻两平面共 6 点构成的特殊结构^[23], Song 求得的相邻两平面 5 条直线和 4 平面 6 条直线所具有的透视不变量^[24]等. 而仿射不变量一般有共线三点的单比、共点三线夹角之比、两条平行线段的长度比、共面封闭区域的面积之比等. 利用这些不变量可以确定某些特殊点或直线在场景中的位置, 如图 1(f) 所示. 但是, 一般而言, 完全利用几何不变量进行所有点或直线段之间的配准是非常复杂的, 参与不变量计算的直线段会产生众多排列组合. 因此, 我们仅利用几何不变量配准少量直线段, 满足平面单应矩阵计算的需要即可.

综上, 在识别过程中对于模型图像中的每一个结构面, 我们先利用几何不变量配准少量直线段, 以满足平面单应矩阵的计算, 然后再配准该平面中其余的直线段. 在配准过程完成之后, 再进行验证, 对匹配的代价进行估算, 以确定识别的正确性.

3 匹配与识别算法

3.1 边缘直线段表示与匹配代价

我们用一组直线段 $L = \{l_1, l_2, \dots, l_r\}$ 代表二维图像中人造物体的轮廓线, 其中的每一条直线段都可以表示为 $l_i = [x_i, y_i, \theta_i, d_i]^T$, 其中 x_i, y_i 表示直

线的中点, θ_i 表示直线的倾角, d_i 表示直线的长度. 现有已知的标准模型图像中的长直线段集 $L_{ML} = \{l_{M1}, l_{M2}, \dots, l_{Mn}\}$ 以及待识别的场景图像 Img_{Scene} . 我们可以通过边缘提取、直线链码跟踪、角点检测等步骤识别 Img_{Scene} 中的边缘直线段, 并形成场景图像的长直线段集 $L_{SL} = \{l_{S1}, l_{S2}, \dots, l_{Sm}\}$.

对于标准模型图像中的长直线段集 L_{ML} 中的每一条直线 l_{Mi} , 通过配准算法: $f_{match}(l_{Mi}) = l_{Sj} (1 \leq i \leq n, 1 \leq j \leq m)$ 与 L_{SL} 中的一条直线 l_{Sj} 形成匹配.

再通过代价估算函数 $f_{cost}(M, S) = \frac{1}{n} \sum_{i=1}^n \epsilon D(l_{Mi}, f_{match}(l_{Mi})) (1 \leq i \leq n)$, 估算模型与场景之间的平均匹配代价. 其中 $D(l_{Mi}, f_{match}(l_{Mi}))$ 为单个匹配的代价估算函数, ϵ 为加权参数. 若某一个模型图像 M 与场景图像 S 之间的平均匹配代价小于某一阈值 G , 则

该物体模型被识别. 配准算法 f_{match} 直接影响模型与场景图像之间的匹配, 是物体识别的关键.

3.2 定义模型图像

在识别之前, 需要对模型图像的结构进行描述以帮助之后的识别, 一个常用的对三维物体的表达方法是依据平面来组织特征. 本方法对于模型图像的描述由 3 个部分构成: 面结构关系、直线段特征、透视不变量. 依据可见面的相邻性将模型物体的所有平面够成图结构. 如图 2 所示, 可以将一个具有 3 个可见面的汽车模型图像(a)描述为(b)结构. 然后, 对于模型图像中长轮廓线集 L_{ML} 中的每一条直线段, 我们都需指明其所在的平面. 让每个平面 P_k 都分别维护一个直线组 L_{P_k} , 而 L_{P_k} 中的每一个元素都是指向 L_{ML} 中某一直线段, 代表该直线属于这个平面, 如图 2(c)所示.

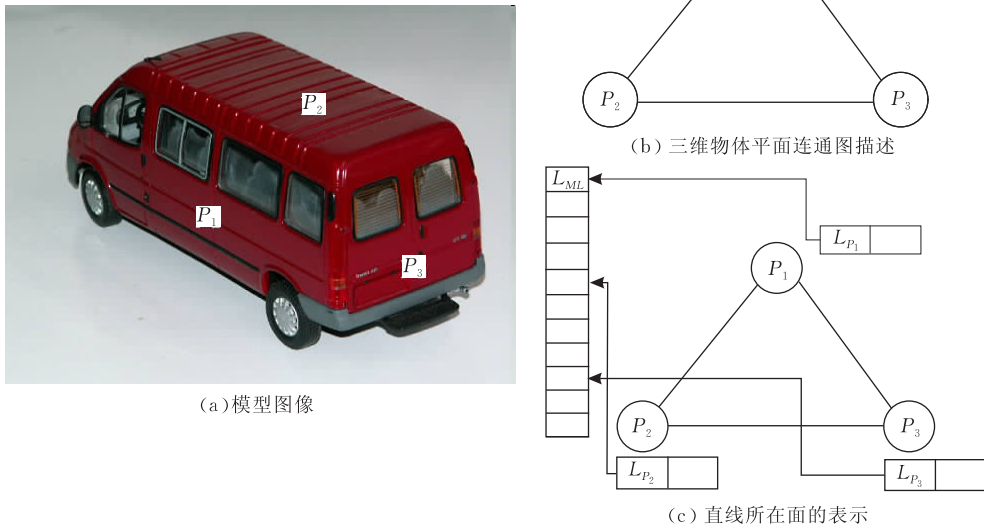


图 2 物体结构描述

假设一共有 K 个平面, 有 $L_{ML} = L_{P_1} \cup L_{P_2} \cup \dots \cup L_{P_K}$. L_{ML} 中的某些直线段可能同时拥有多个指针, 因为它可能是多个平面的交线. 除了描述了模型物体的结构之外, 我们还需给出模型中所含的不变量. 这些不变量可以被简单地分为两类: 第 1 类不变量 ($Inv I$) 为共面不变量, 如最简单的四点共线、四线共点, 所衍生的同一平面内 5 个点、同一平面内 5 条线等透视不变量以及大量仿射不变量如共线 3 点的单比、共点三线夹角之比、两条平行线段的长度比等; 第 2 类不变量 ($Inv II$) 为非共面不变量, 如 Zhu 等提出了一种相邻两平面共 6 点构成的特殊结构^[23]、Song 求得的相邻两平面 5 条直线和 4 平面

6 条直线所具有的透视不变量^[24]等. 对于第 1 类不变量, 可以用来直接匹配直线段, 但必须保证其共面性. 而第 2 类不变量可以表示两平面间的位置关系. 通过第 2 类不变量, 我们可以找出一个已识别平面的相邻平面.

对于第 1 类不变量 $Inv I$, 我们首先选用共线四点的交比值, 它由一个特征值以及指向同一平面内两条共线直线段的指针组成: $invI = [(l_1, l_2), ivI]^T$, 其中 $(l_1, l_2) = l_{ipair}$ 为组成交比的两条线段, ivI 表示两共线直线段所形成四点的交比值. 如果模型不含在同一平面内两条共线直线段这一特征, 可采用仿射不变量, 两条平行线段的长度比代替. 它

由一个特征值以及指向同一平面内两条平行直线段的指针组成: $invI' = [(l_1, l_2), ivI']^T$, 其中 l_1, l_2 为两条平行的直线段, ivI' 表示他们长度的单比值.

对于第 2 类不变量 $Inv II$, 我们选用相邻两平面中 6 个点, 即三条线段所形成的不变量, 它由一个特征值以及 3 个指针组成, 其中两个指针分别指向两个不同平面内的直线段, 另一个指向这两个平面的交线: $invII = [(l_1, l_2, l_0), ivII]$, $ivII$ 表示不同平面内两直线 l_1, l_2 以及两平面交线 l_0 所构成的六点的不变量的值, 如图 3 所示.

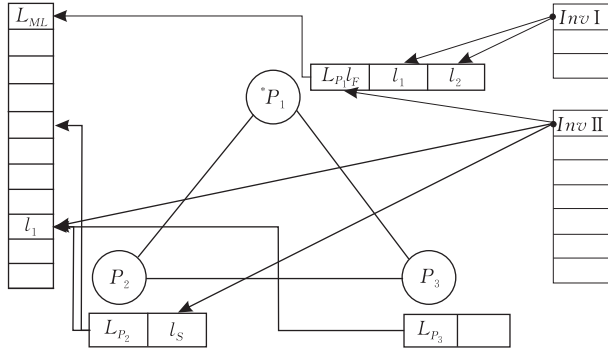


图 3 透视不变量描述

3.3 结构面的不变性匹配方法

不变性匹配旨在得到模型与场景图像间少量直线段间所存在的可能的匹配, 以帮助完成场景平面与模型平面间的映射. 共面不变性匹配及非共面不变性匹配旨在寻找可能的匹配, 即提出假设的过程, 单应平面匹配为验证过程.

3.3.1 $f_{match}^{Inv I}$ 共面不变性匹配

共面不变性匹配 $f_{match}^{Inv I}$ 通过在场景图像中寻找与标准模型中已知不变量相符合的特征完成匹配, 所得的匹配是一个集合. 当使用共线四点定义共面不变量时, 我们可以采用如下 3 个条件判断两直线段 l_1, l_2 共线 ($l_1, l_2 \in L_{SL}$):

(1) 两直线夹角小于 θ_μ 度, 或者在两直线长度比相差较大的情况下夹角约束适当放宽

$$|\theta_1 - \theta_2| < \theta_\mu \text{ 或} \\ [(d_1/d_2 < 0.15 \text{ 或 } d_2/d_1 < 0.15) \text{ 且 } |\theta_1 - \theta_2| < \epsilon_1 \theta_\mu] \quad (1)$$

(2) 任一直线的中点到另一直线的距离需小于一个阈值, 同理该条件在夹角足够小时适当放宽

$$(mid_1 + mid_2)/2 < mid_\mu \text{ 或} \\ (|\theta_1 - \theta_2| < \frac{1}{2} \theta_\mu \text{ 且 } (mid_1 + mid_2)/2 < \epsilon_2 mid_\mu) \quad (2)$$

mid_1, mid_2 分别表示一直线的中点到另一直线的

距离.

(3) 两直线最外端点的长度必须大于两直线长度和

$$d_{max} > d_1 + d_2 \quad (3)$$

d_{max} 表示两直线最外端点构成的直线 l_{max} 的长度.

场景图像的长直线段集 L_{SL} 中所有同时满足上述 3 个判据条件的共线组 $l'_{pair} = [l_1, l_2]$ 构成共线集 L_{pair} . 对于每一个共线组, 计算它们的交比, 得到不变量特征值 $ivalue(l'_{pair})$. 如果我们使用两平行线间单比的仿射不变量定义共面不变量, 我们只需参考两线夹角这一约束条件. 即在一组线段符合两线夹角小于 θ_μ 的条件下, 计算两线单比的不变量特征值 $ivalue(l'_{pair})$. 得到不变量特征值之后, 须与 $Inv I$ 中的值比较, 与特征值最为接近的 G_1 个共线组为可能的匹配, 用一个集合表示: $f_{match}^{Inv I}(l_{pair}) = \{l'_{pair} \mid l'_{pair} \in L_{pair} \wedge \min(|ivI - ivalue(l'_{pair})|)\}$. 将这些可能的候选匹配线段的端点与 l_{pair} 的端点一起加入配准点集 $Pair_k$. 当同一平面内有两组或者两组以上元素找到可能的匹配之后, 该平面进行单应性匹配. 事实上, 在定义模型图像时, 必定需要有一个平面拥有两个或者两个以上的共线不变量组以支持后续的识别, 我们将这个平面称为起始平面, 在图 3 用 $*P_1$ 表示.

3.3.2 f_{match}^u 单应平面匹配

单应平面匹配旨在利用平面的单应性配准模型图像中一个平面内的所有直线段. 求平面单应矩阵

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix},$$

需要预先配准 4 个或者 4 个以上的非共线点. 这些点的配准我们利用不变性匹配来完成.

在模型的某一平面 P_k 的配准点集 $Pair_k$ 中, 尽可能多地抽取 n 个已知可能的配准点, $n \geq 4$. 对于这 n 个点, 有 $p_M \cong H p_S$, 即

$$\lambda \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix},$$

可以得到两个有关平面单应矩阵 \mathbf{H} 的方程组^[21]:

$$\begin{cases} h_{11}x + h_{12}y + h_{13} = h_{31}xx' + h_{32}yy' + h_{33}x' \\ h_{21}x + h_{22}y + h_{23} = h_{31}xy' + h_{32}yy' + h_{33}y' \end{cases}$$

因此, 当具有 4 个或者 4 个以上的非共线点时, 我们可以在 $\|\mathbf{H}\| = 1$ 的约束下采用 SVD 法求解平面单应矩阵 \mathbf{H} .

在求出平面单应矩阵 \mathbf{H} 之后, 我们将该模型平

面中所有直线段集 L_{P_k} 按长度由长到短排序之后映射到场景图像中, 形成映射后的直线段集 $L_{S_k} = \mathbf{H}^{-1}L_{P_k}$. 对于映射后直线段集 L_{S_k} 中的每一条直线, 我们要依次在场景图像长直线集 L_{SL} 中找到适合的匹配, 使得它们的匹配代价 f_{cost} 最小.

我们从中点纵向距离偏差、横向距离偏差、直线长度偏差、角度偏差 4 个方面来估算两条直线段的匹配程度:

$$d_{\text{ver}}(l_1, l_2) = \frac{v_1 + v_2}{2},$$

$$d_{\text{hor}}(l_1, l_2) = \sqrt{d_{\text{mid}}^2(l_1, l_2) - d_{\text{ver}}^2(l_1, l_2)},$$

$$d_{\text{len}}(l_1, l_2) = |d_1 - d_2|,$$

$$d_{\theta}(l_1, l_2) = |\theta_1 - \theta_2|,$$

$$D(l_1, l_2) =$$

$$\sqrt{d_{\text{ver}}^2(l_1, l_2) + \frac{d_{\text{hor}}^2(l_1, l_2)}{W_{\text{hor}}^2} + \frac{d_{\text{len}}^2(l_1, l_2)}{W_{\text{len}}^2} + \frac{d_{\theta}^2(l_1, l_2)}{W_{\theta}^2}},$$

v_1 为直线段 l_1 中点到直线段 l_2 的距离, v_2 为 l_2 中点到 l_1 的距离, d_{ver} 为两直线段纵向距离偏差, d_{mid} 为两直线段中点的距离, d_{hor} 为两直线段横向距离的偏差, d_{len} 为长度偏差, d_{θ} 为角度偏差. W_{hor} 、 W_{len} 、 W_{θ} 为各项偏差的权重因子, 视情况而定. 一般而言, 纵向偏差为直线段的主要偏差即 $W_{\text{hor}} \gg 1$ 和 $W_{\text{len}} \gg 1$. 对于 L_{S_k} 中的每一条线, 依次从场景图像长直线集 L_{SL} 找出匹配代价最小的直线段, 匹配方式为

$$f_{\text{match}}^{\mathbf{H}}(l_{S_k}) = \begin{cases} \min_{l_{SL} \in L_{SL}} (D(l_{S_k}, l_{SL})), & \min_{l_{SL} \in L_{SL}} (D(l_{S_k}, l_{SL})) \leq G_3 \\ \min_{l_{SL} \in L_{SL}} (\min_{l_{SL} \in L_{SL}} (D(l_{S_k}, l_{SL})), 2G_3), & \\ \min_{l_{SL} \in L_{SL}} (D(l_{S_k}, l_{SL})), & \min_{l_{SL} \in L_{SL}} (D(l_{S_k}, l_{SL})) \geq G_3 \end{cases},$$

$l_{S_k} \in L_{S_k}$. G_3 为最大匹配阈值, 当 $\min_{l_{SL} \in L_{SL}} (D(l_{S_k}, l_{SL})) >$

G_3 时 $f_{\text{match}}^{\mathbf{H}}(l_{S_k})$ 取 $\min_{l_{SL} \in L_{SL}} (D(l_{S_k}, l_{SL}))$ 和 $2G_3$ 中的较小值. 当一条直线段的配准 $f_{\text{match}}^{\mathbf{H}}(l_{S_k})$ 完成之后, 将配准的 l_{SL} 从场景图像长直线集 L_{SL} 中取出, 并将它们放入集合 L'_{S_k} . 当所有直线都完成配准, 估算 L_{S_k} 与 L'_{S_k} 之间的平均匹配代价,

$$f_{\text{cost}}(L_{S_k}, L'_{S_k}) = \frac{1}{\text{num}} \sum_{i=1}^{\text{num}} \varepsilon f_{\text{match}}^{\mathbf{H}}(l_{S_k}),$$

如果该代价小于阈值 G_4 , 则该平面中所有直线段匹配成功. 如果该代价大于阈值 G_4 , 则考虑其它可能的 n 个配准点.

3.3.3 $f_{\text{match}}^{\text{Inv II}}$ 非共面不变性匹配

当一个平面匹配成功之后, 需要完成对其相邻平面直线段的匹配. 同样, 我们无须利用不变量去寻

找所有相邻平面中的直线, 因为只要获得 4 个非共线点就可以利用平面单应性完成该相邻平面内所有直线的配准. 在已知两相邻平面的公共边的情况下, 理论上我们只需要匹配一条直线段即可.

对于一个第 2 类不变量 $\text{invII} = [(l_1, l_2, l_0), \text{invII}]$, 其中 l_1 代表模型图像已知平面中的直线, l_0 代表模型图像中两平面的公共交线, 由于一个平面已经被识别, 其单应性矩阵 \mathbf{H}_1 已知, 因此 l_1, l_0 在场景中的像 $l'_1 = \mathbf{H}_1^{-1}(l_1)$, $l'_0 = \mathbf{H}_1^{-1}(l_0)$, 则

$$f_{\text{match}}^{\text{Inv II}}(l_2) = \{l'_2 | l'_2 \in L_{SL} \wedge \min_{G_2} (|\text{ivalue}(l'_1, l'_2, l'_0) - \text{invII}|)\},$$

即可能的匹配为与不变量都最为接近的 G_2 组直线段.

3.4 整体识别与算法描述

当模型的一个平面得到匹配后, 判断已识别的平面集是否满足识别模型所需要的条件, 如果是, 模型识别成功, 否则, 对其它相邻平面进行识别. 若所有相邻平面在识别后仍无法满足模型识别条件, 则识别失败. 模型识别条件即模型所具有的可见面的所有情况的集合, 一般与已识别平面的单应矩阵 \mathbf{H} 有关. 利用单应矩阵 \mathbf{H} 中反映变形变换 (deformation) 的 4 个量 $h_{11}, h_{12}, h_{21}, h_{22}$ 来表示可见平面的变形程度, 可以帮助构造模型识别条件 Cond .

我们可以用以下算法大致描述上述方法:

$$\text{Model} = \{P_K, L_{P_1}, \dots, L_{P_K}, \text{Inv I}, \text{Inv II}, \text{Cond}\}$$

输入: $\text{Img}_{\text{Scene}}, \text{Model}$

1. 利用直线检测算法找出场景图像 $\text{Img}_{\text{Scene}}$ 的长直线集 L_{SL} , 然后利用共线判据(1)(2)(3)找出共线组 L_{pair} , 若不具备共线条件, 则寻找平行线组 L_{pair} .

2. 对于某一模型平面中所有第 1 类不变量 Inv I , 找出 L_{pair} 中满足条件 $\min_{G_1} (|\text{ivI} - \text{ivalue}(l_{\text{pair}})|)$ 的 G_1 个候选直线对, 将它们的端点添加到该平面的匹配点集, 对于该平面, 完成步 3~4 操作.

3. 取该平面 P_k 的匹配点集 Pair_k 中尚未取过的非共线点所组成的最大子集, 如果点的数量小于 4, 转到步 6; 否则, 根据匹配点集 Pair_k 计算平面 P_k 的单应矩阵 \mathbf{H} .

4. 对于模型平面 L_{P_k} 中的每一条直线 l_{P_k} , 将其映射到场景图像: $l_{S_k} = \mathbf{H}^{-1}l_{P_k}$, 根据条件 $\min_{l_{SL} \in L_{SL}} (D(l_{S_k}, l_{SL}))$ 找场景中可能的匹配. 如果匹配程度满足 $f_{\text{cost}}(L_{S_k}, L'_{S_k}) < G_4$, 平面 P_k 匹配成功, 转到步 6; 否则, 转回到步 3.

5. 对于该相邻平面 P_j 中所有第 2 类不变量 Inv II , 找出 L_{SL} 中满足条件 $\min_{G_2} (|\text{ivalue}(l_1, l_2, l_0) - \text{invII}|)$ 的 G_2 个候选直线对, 将它们的端点添加到匹配点集 Pair_j , 对于该平面完成步 3~4 操作.

6. 如果匹配的平面集满足模型识别条件 Cond , 则识别

成功;否则,对于该匹配平面的所有相邻平面进行步 5 操作,若没有相邻平面,则识别失败.

根据算法描述,整个识别过程可以用图 4 描述.

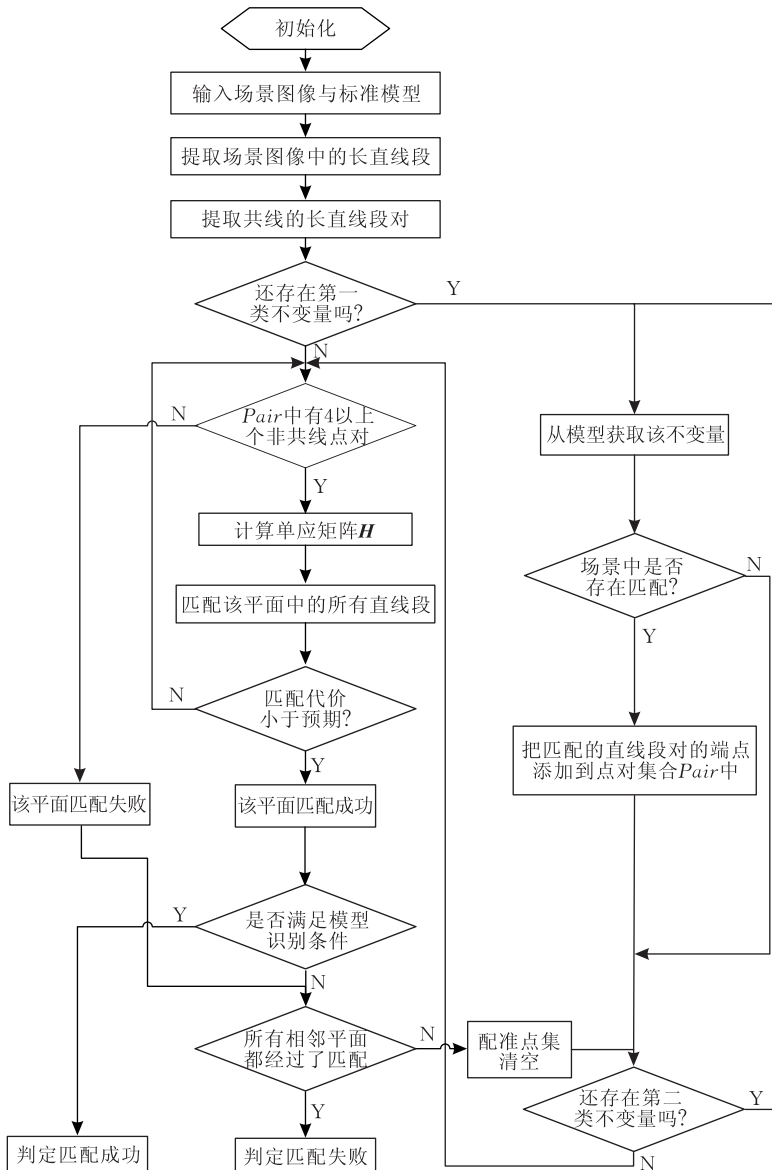


图 4 识别方法流程图

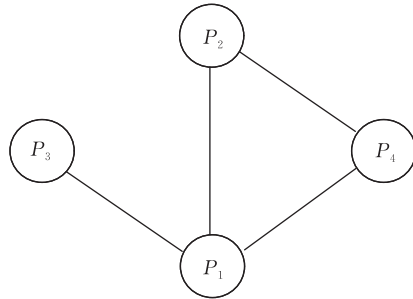
4 实 验

我们采用真实图像验证本算法,而实验的模型基于实际意义的考虑,选用与真实图像接近的汽车或者飞机模型.首先定义模型中各个平面间的对应关系,如图 5(b)所示.而后通过对于已有模型图像的人工采样和测量得到模型中每一个平面的直线段特征以及各类的不变量值.图 5(a1)~(a4)为全顺汽车模型各个平面直线段集的定义.图 5(c1)~(c3)列出了汽车模型中部分几何不变量的定义,可以看到,

模型侧面的窗框特征具有两组共线的直线段,能够提供第 1 类不变量.而第 2 类不变量的定义都是分别由两平面中各一条直线段以及一条两平面间交线组成的,他们在相交的两个平面间大量存在.对于汽车模型,可以给出一个较为简单的模型识别条件的定义:(1)有两个或两个以上的平面被识别;(2)有一个平面被识别,且这个平面的单应矩阵满足 $\sqrt{|(h_{11}-h_{22})(h_{12}-h_{21})|} < 0.5$. 条件(2)表示场景图像的视角接近于正对该平面.以上两个条件满足任意一个即可,更为复杂的模型识别条件也可以通过类似方法定义.



(a1) 模型 P_1 面的直线段模型描述



(b) 平面连通图描述



(a2) 模型 P_2 面的直线段模型描述



(c1) 部分几何不变量的表示



(a3) 模型 P_3 面的直线段模型描述



(c2) 部分几何不变量的表示



(a4) 模型 P_4 面的直线段模型描述



(c3) 部分几何不变量的表示

图 5 全顺汽车模型描述

图 6(a)为一幅场景图像,其中有一辆待检测的真实全顺汽车.采用文献[25]所提出的基于直线链码的直线段检测方法,并参考文献[26]设定直线段的最短链码数 $N_L = 10$,角点检测的最大链码曲率阈值为 $C_S = 0.4$,直线链码判别的直方图参数 $\sigma_1 = 0.85$ 、 $\sigma_2 = 0.2$,可自动得到场景中的长直线段集.取共线直线检测参数 $\theta_\mu = \pi/20$ 、 $\epsilon_1 = 1.2$ 、 $mid_\mu = 2.0$ 、 $\epsilon_2 = 0.5$;直线匹配权重参数 $W_{hor} = 3.0$ 、 $W_{len} = 10.0$ 、 $W_\theta = 2.0$;匹配阈值参数 $G_1 = 3$ 、 $G_2 = 15$ 、 $G_3 = 20.0$ 、 $G_4 = 25.0$,得到真实图像中各面的匹配结果,

如图 6(b)所示.

在匹配过程中,我们先利用共线不变量匹配汽车侧面共线的直线组,进而利用匹配的四点求出汽车模型侧面与场景侧面间的单应矩阵,如图 7(a1)所示,完成侧面从模型到场景的映射,然后匹配它们之间的直线段.当侧面完成匹配之后,利用侧面中已匹配的直线段以及两面间的交线,我们可以求出与已知第 2 类不变量对应的汽车模型前面的直线段,并将其与模型中的直线段对应.完成了一条前面直线段匹配之后,结合已知的前面与侧面交线,构成

4 个点便可以计算出汽车模型前面与场景前面之间的单应矩阵,完成映射并匹配,如图 7(b1)、图 7(b2)所示.图 7(c1)、(c2)为顶面的匹配结果.可以看出大多数模型侧面和顶面的直线段被成功匹配了,匹

配错误的直线段很少.而由于汽车前面是一个曲面,部分特征直线段并没有在场景图像中显示出来,因此相对而言匹配的直线段略少.



图 6 全顺真实场景图像以及自动直线段检测

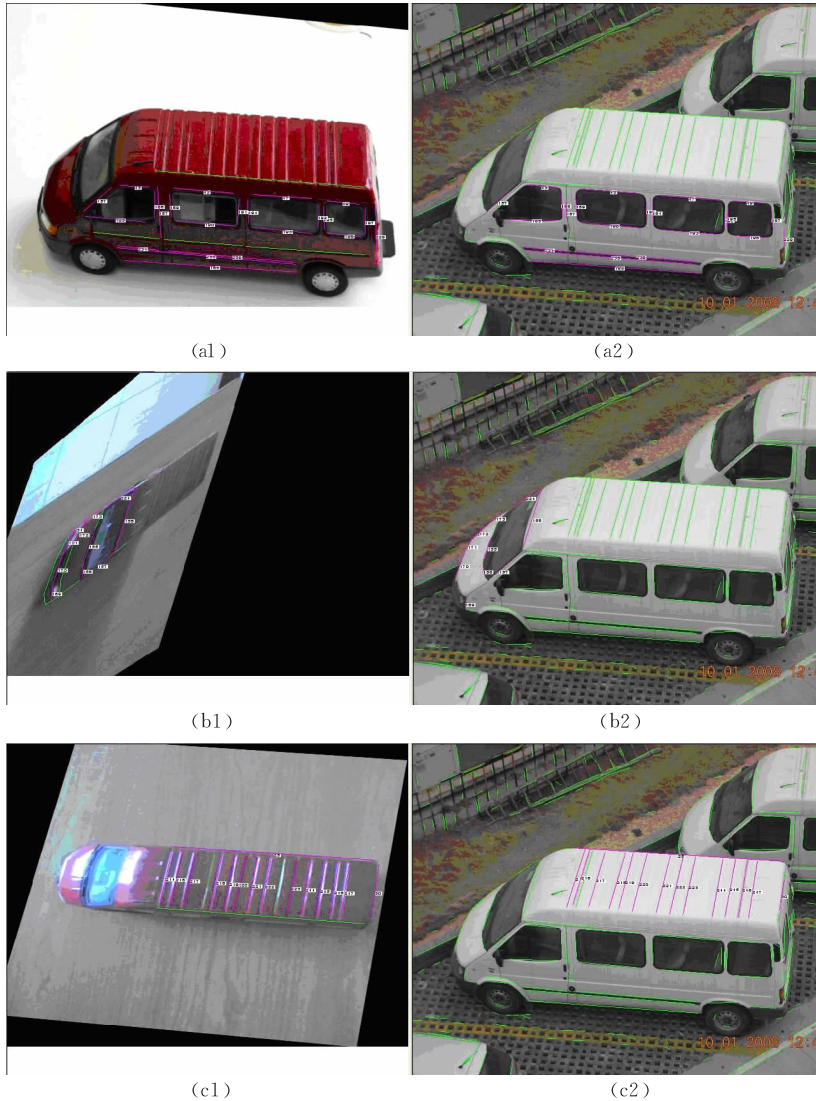


图 7 全顺汽车各面的匹配结果

图 8 为其它各类全顺汽车的匹配识别结果,场景图像大小都为 624×468 ,可以看出即使是右边

4 幅与模型稍有不同改进型全顺汽车,只要直线特征没有太大变化,本方法也能够较好地识别.



图 8 其它全顺汽车场景图像的匹配结果

对于一些基本共面的模型,本方法也同样适用,如附录中所示对飞机的不变性识别结果.我们还对其它类别的物体进行了不变性识别实验,由于它们

都具有比较显著的几何结构,因此第 1 类与第 2 类不变量易于确定,外形廓线分段也容易实现匹配.具体结果参见表 1 中的汇总.

表 1 对其它类别物体的匹配效果汇总

| 人造物体类别 | 目标物 | 模板数 | 所用透视不变量 | 容忍姿态差程度 | 识别成功率/% | 成功匹配到的可视轮廓线段数/% |
|--------|---------|-----|------------|---------|---------|-----------------|
| 建筑物 | 高层建筑 | 6 | 四点共线、两平面六点 | 较大 | 86 | >50 |
| 交通工具 | MPV 汽车 | 5 | 四点共线、两平面六点 | 中等 | 94 | >40 |
| 家具 | 多格书架 | 3 | 四点共线、两平面六点 | 较大 | 95 | >55 |
| 玩具 | 翻斗车 | 8 | 四点共线、两平面六点 | 较小 | 83 | <35 |
| 仪器设备 | 液晶显示器 | 3 | 四点共线、两平面六点 | 较大 | 92 | >60 |
| | 手枪 | 2 | 四点共线 | 较大 | 78 | >50 |
| 武器 | 坦克 | 5 | 两平面六点 | 较小 | 61 | <35 |
| | 战斗机 | 9 | 两平面六点 | 中等 | 65 | >40 |
| 室内场景 | 隔断形成的通道 | 8 | 四点共线、两平面六点 | 中等 | 75 | >40 |

一个人造物体的外部轮廓可以看成是由若干凸多边形组成,如三角形、四边形等.基于单应性矩阵映射的匹配方法对同一凸多边形在不同姿态下的图像的匹配满足一定的概率,这个概率在姿态相差较小时较大,姿态相差较大时较小,其概率分布是基于多边形类别、姿态差和模板多样性(大于等于 0 且小于等于 1)的函数 $Pr(polygon, \Delta\alpha, Diversity)$. 假设一个人造物体由凸多边形集合 $Sur = \{ polygon_i, 1 \leq i \leq K \}$ 组成其外部轮廓.由于成像使得 Sur 中的多边形不可能同时可见,因此可以得到一个不同投

影姿态下的侧面集合 $Side = \{ S_i | S_i \subseteq Sur, 1 \leq i \leq H \}$. 其中可用于对该物体进行识别的轮廓特征也构成一个集合 $Fea = \{ F_i | F_i \subseteq Sur, 1 \leq i \leq M \}$, 其中 F_i 是 Sur 的一个子集,表示基于其中的那些多边形已经足够辨识该物体了.我们容易发现在有些姿态下多边形的像与其模板的匹配效果好,也就容易识别;在另外一些姿态下匹配效果不好,当然也就识别困难.但在大多数情况下组成物体外部轮廓的多个多边形不可能同时处在不佳的匹配姿态下,因此能够获得可以接受的整体匹配概率 P_{match} . 按照或

事件的概率计算法则,

$$P_{match} = \sum_{i=1}^H \left[p(S_i) \times Pr(\exists F_j (F_j \in Fea \wedge F_j \subseteq S_i)) \times \prod_{polygon_k \in F_j} Pr(polygon_k, \Delta\alpha_k, diversity_k) \right].$$

由于简单凸多边形获得良好匹配概率的姿态范围比较大,同时在极端视角下观测物体的可能性很小,加上不同姿态下用于匹配的模板数量也比较充足,因此我们能够保证这个整体匹配概率较高.根据在汽车模型上的实验结果,此概率通常能够达到 0.7 以上.

5 讨 论

本文提出了一种基于直线段特征的三维物体识别方法,采用几何不变性特征结合平面单应性的方法,主要用于对人造物体的识别,取得了较好的实验效果,尤其是对于透视不变性特征较为明显且共面性较强的物体.而对于部分几何不变性特征不太明显的情况,本文将透视不变量扩展到仿射不变量,实现弱透视投影下的识别.相对于传统的假设验证法,本方法将对于全局坐标系变换的求解分散在各个平面单应性变换的求解中,降低了求解的复杂度,因此能够快速准确地识别物体,同时在物体共面性较差的情况下依然能够做到对物体局部的识别.本方法尚未能对曲线所构成的不变量进行求解,如利用平面内的圆形计算透视下的单应矩阵等,这将是今后工作的重点.

参 考 文 献

- [1] Vicente M A, Gil P, Reinoso O, Torres F. Objects recognition by means of projective invariants considering corner-points. *Journal of WSCG*, 2002, 10(3): 129-136
- [2] Paul J B, Ramesh C J. Three-dimensional object recognition. *ACM Computing Surveys*, 1985, 17(1): 75-145
- [3] Grimson W E L, Lozano-Perez T. Localizing overlapping parts by searching the interpretation tree. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1987, 9(4): 469-482
- [4] Stein F, Medioni G. Structural indexing: Efficient 3D object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, 14(2): 125-145
- [5] Best L, Magee M. Autonomous construction of three dimensional models from range data. *Pattern Recognition*, 1998, 31(2): 121-136
- [6] Chen S, Stockman G. Object wings-2 1/2 D primitives for 3D recognition//*Proceedings of the CVPR*. San Diego, CA, 1989: 535-540
- [7] Chen J L, Stockman G C. 3D free-form object recognition using indexing by contour features. *Computer Vision Image Understanding*, 1998, 71(3): 334-335
- [8] Campbell R J, Flynn P J. A survey of free-form object representation and recognition techniques. *Computer Vision Image Understanding*, 2001, 81(2): 166-210
- [9] Liang P, Taubes C H. Orientation-based differential geometric representations for computer vision applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1994, 16(3): 249-258
- [10] Santos E M, Gomes H M. Appearance-based object recognition using support vector machines. *Computer Graphics and Image Processing*, 2001: 399
- [11] Selinger A, Nelson R C. Appearance-based object recognition using multiple views//*Proceedings of the Computer Vision and Pattern Recognition*. Hawaii, 2001, 1: 905-911
- [12] Arbter K, Snyder W E, Burkhardt H. Application of affine-invariant Fourier descriptors recognition of 3D objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, 12(7): 640-647
- [13] Weiss I. Geometric invariants and object recognition. *International Journal of Computer Vision*, 1993, 10(3): 207-231
- [14] Quan L. Invariants of six points and projective reconstruction from three uncalibrated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, 1(1): 34-46
- [15] Zisserman A, Forsyth D A. 3D object recognition using invariance. *Artificial Intelligence*, 1995, 78(1-2): 239-288
- [16] Li Q, Zhou M L, Liu J. A review on 3D objects recognition. *Journal of Image and Graphics*, 2000, 5(12): 985-993
- [17] Ayache N, Faugeras O D. HYPER: A new approach for the recognition and positioning of two dimensional objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1996, 8(1): 44-54
- [18] Ullman S, Basri R. Recognition by linear combination of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991, 13(10): 992-1006
- [19] Mundy J L, Heller A J. The evolution and testing of a model-based object recognition system//*Proceedings of the 3rd International Conference on Computer Vision*. Osaka, Japan, 1990: 268-282
- [20] Kim D H, Yun D, Lee S U. Interactive 3-D shape retrieval system using the attributed relational graph//*Proceedings of the Computer Vision and Pattern Recognition Workshop*. Washington, DC, 2004: 147-154
- [21] Hartley R I, Zisserman A. *Multiple View Geometry in Computer Vision*. New York: Cambridge University Press, 2000
- [22] Forsyth D A, Mundy J L, Zisserman A, Coelho C, Heller A, Rothwell C. Invariant descriptors for 3-D object recognition and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, 13(10): 971-992
- [23] Zhu Y, Senevirame L D, Earles S W E. A new structure of invariant for 3D point sets from a single view//*Proceedings of the IEEE International Conference on Robotics and Automation*. Nagoya, Aichi, 1995, 2: 1726-1731

- [24] Song B S, Lee K M, Lee S U. Model-based object recognition using geometric invariants of points and lines. *Computer Vision and Image Understanding*, 2001, 84(3): 361-383
- [25] Freeman H, Davis L S. A corner finding algorithm for chain code curves. *IEEE Transactions on Computers*, 1977, 26:

297-303

- [26] Lu G Q, Xu H G, Li Y B. Line segment detection based on chain code detection. *Computer Engineering*, 2006, 32(14): 1-3



WEI Hui, born in 1971, Ph. D., professor. His research interests include artificial intelligence, cognitive science.

QIU Zhen-Yu, born in 1984, master. His research interests focus on computer vision.

Background

The study of this paper belongs to a field of object recognition with invariance in computer vision and pattern recognition. The evidences either from cognitive psychology or from physiological neuroscience, the object recognition is highly geometrical feature-based. But two big problems, one is feature selecting and the other is binding, block us to design machine vision system working purely in a space of geometrical features except that a larger amount of knowledge is indispensable. Due to these difficulties, more and more programs adopt some alternative methods, they can work to answer question 'is there any X (some special object) in a picture', but they usually fail to answer question 'where is it' or 'how many are they'. This is because image understanding relies more on geometrical features selection and combination, less on physical signals processing. The invariance of visual perception is not sensitive to geometrical features and their combination, but highly sensitive to even subtle change in pixel level. Thus using features with obvious geometrical formation to recognize object is a very good choice. As generally, the size of a geometrical feature is bigger than pixel, and it is often composed of hundreds of pixels. So the ap-

pearance of feature is influenced by object's pose, distance, structure and spatial position. This makes it is very difficult to isolate features from an image, and to locate their positions, and to define their relationships, and to produce a spatial distribution of them, and to distinguish figure from background. Feature-based perceptual invariance is challenging because several hard problems like integration, attention, knowledge-driven, get involved. At same time this method is also very promising because it consistent with essential mechanism of human vision. In theory we know little about how to select and combine geometrical features efficiently, and in practice we cannot cluster pixels into feature quickly. These two problems prevent computer vision based on geometrical feature to be popular. Line is a kind of general feature in manmade object. In this paper we utilize lines and geometrical invariant to recognize object with invariance. This work is a part of 973 Project with No. 2010CB327900 and Shanghai Science and Technology Development Funds Project with No. 08511501703, and this paper tries to achieve pose invariance. The main goal of these two projects is to develop new intelligent sensing algorithm.