

摄像机位姿的加权线性算法*

杨 森, 吴福朝⁺

(中国科学院 自动化研究所 模式识别国家重点实验室, 北京 100190)

Weighted Linear Methods for the Camera Pose Estimation

YANG Sen, WU Fu-Chao⁺

(National Laboratory of Pattern Recognition, Institute of Automation, The Chinese Academy of Sciences, Beijing 100190, China)

+ Corresponding author: E-mail: fcwu@nlpr.ia.ac.cn

Yang S, Wu FC. Weighted linear methods for the camera pose estimation. Journal of Software, 2011, 22(10): 2476-2487. <http://www.jos.org.cn/1000-9825/3916.htm>

Abstract: This paper presents a novel weighted linear method for the camera pose estimation. The key idea of this method is to replace the algebraic error in the classic linear method with the weighted algebraic error that closes the geometric error. The method provides a linear solution whose accuracy is close to the accuracy of an ML estimation. Based on the DLT (direct linear transformation) algorithm and EPnP algorithm, the weighted DLT (WDLT) and weighted EPnP (WEPnP) algorithms are obtained by using the weighted linear technique. Experimental results with simulative data and real images show that the WDLT and WEPnP algorithms remarkably outperform the DLT and EPnP algorithms and in the case of small depth ratio, both of them outperform the Lu's nonlinear algorithm.

Key words: camera pose; DLT (direct linear transformation); EPnP; WDLT (weighted DLT); WEPnP; ML estimation

摘 要: 针对摄像机位姿问题提出了一种加权线性方法,其关键思想是通过加权使经典线性方法的代数误差近似于重投影算法的几何误差,从而达到接近于最大似然估计(Levenberg-Marquardt 简称 ML)的精度.通过对经典 DLT(direct linear transformation)算法和 EPnP 算法使用加权的方法,给出了加权 DLT 算法(WDLT)和加权 EPnP 算法(WEPnP).大量模拟数据和真实图像实验结果均表明,WDLT 和 WEPnP 算法不仅能提高 DLT 和 EPnP 算法的精度,而且在深度较小的情况下优于 Lu 的非线性算法.

关键词: 摄像机姿态;DLT(direct linear transformation);EPnP;WDLT(weighted DLT);WEPnP;最大似然估计

中图法分类号: TP391 **文献标识码:** A

给定摄像机内参数,从空间点与它的图像点对应估计摄像机的旋转 R 和位置 c ,通常称为摄像机位姿估计问题,也称为 PnP 问题.PnP 问题是计算机视觉中的经典问题,在计算机视觉和图形学领域有很多的重要应用,因此一直受到人们的关注.求解 PnP 问题的方法一般可以分为两类:一类方法是线性算法^[1-5],算法速度较快,但因未使用内在的旋转约束而得不到较高的精度;另一类为非线性迭代算法^[6-9],对某种目标函数进行迭代优化,通常能够获得较高的精度但计算量较大.

* 基金项目: 国家自然科学基金(61075038, 60835003)

收稿时间: 2009-09-27; 修改时间: 2010-03-04; 定稿时间: 2010-07-06

解决 PnP 问题所需要的最少点对应数为 3(P3P),它最多有 4 个解^[10,11].Gao 等人^[12]通过吴方法^[13]给出了 P3P 的完全解分类.Quan 和 Lan 对于 P4P 和 P5P 问题给出了线性 SVD 算法^[2],他们把 PnP 问题转化为多个 P3P 问题,根据 P3P 问题之间的关系线性求解 PnP,其算法时间复杂度较高.在 $n \geq 6$ 的情况下,经典的线性算法是 DLT(direct linear transformation)算法^[1],它忽略了旋转矩阵的正交约束,将投影矩阵 12 个变量看作独立变量而得到线性约束.Fiore 在文献[3]中提出了一种时间复杂度为 $O(n^2)$ 的线性算法,但是对噪声较为敏感^[5].最近,Moreno 和 Lepetit 提出了一种高效、高精度线性算法 EPnP^[4],其基本思想是把空间点都通过 4 个虚拟的控制点所建立的坐标系来表示,其中 4 个控制点的选取分别为空间点集的中心以及 3 个主方向上的单位点.然后根据图像点的信息求解 4 个控制点的摄像机坐标系下的坐标,进而得到旋转矩阵和摄像机坐标.EPnP 被认为是一种高效的线性算法,其时间复杂度为 $O(n)$,但实验结果表明,当深度幅度变化较为剧烈时(即深度比较小时),EPnP 算法难以获得较好的估计结果.

非线性迭代算法^[6-9]是最小化某种代价函数为目标,Lu 方法^[6]是基于空间距离最小化的迭代算法,相比较其他迭代算法,其速度较快,但在图像点噪声服从高斯分布下,其精度劣于重投影误差最小化算法^[9].如果图像点噪声服从高斯分布,则重投影误差最小化方法在最大似然意义下是最优的.非线性迭代算法虽然能够达到很高的精度,但对初值却非常依赖,其可能得到局部极小值,而非全局极小值,并且计算量大大高于线性算法.Hartley 和 Kahl^[7]最近提出一种利用无穷范数的全局优化算法,他们利用四元数^[14],以一个半径 π 的球为旋转参数空间,并通过遍历旋转空间和分支定界算法寻找全局最优解.但是这种算法运算量巨大,并且实验结果表明,在高斯噪声假设情况下,其精度低于重投影误差最小化方法的精度.

本文提出了一种新的线性算法,该算法除了继承 DLT 算法速度快的优点以外,与现有的其他线性算法相比能够获得较高精度,并且对深度变化较为稳定,在深度变化幅度较大的情况下仍然可以获得较好的结果.本文算法的主要思想是使用加权技术,使线性算法中的代数距离误差近似于重投影算法的几何距离误差,从而达到接近最大似然算法的精度.

本文第 1 节给出加权求精算法的基本思想.第 2 节和第 3 节分别给出加权 DLT 算法(WDLT)和加权 EPnP 算法(WEPnP).第 4 节是模拟数据与真实图像的实验对比.第 5 节总结全文.

1 加权的基本思想

给定一组空间点与图像点对应 $\{x_i = (\tilde{x}_i^T, 1)^T \leftrightarrow u_i = (\tilde{u}_i^T, 1)^T \mid i = 1, 2, \dots, N\}$,其中 x_i 表示空间点在世界坐标系下的齐次坐标, $\tilde{x}_i = (x_i, y_i, z_i)^T$ 是该空间点的非齐次坐标, u_i 表示图像点的齐次坐标, $\tilde{u}_i = (u_i, v_i)^T$ 是它的非齐次坐标.记 $P=(A, b)$ 为投影矩阵,如果 R 为摄像机坐标系关于世界坐标系的旋转,并且摄像机中心在世界坐标系中的齐次坐标为 $c = (\tilde{c}, 1)^T$, 则 $A = KR, b = -KR\tilde{c}$.即,投影矩阵能够表示为 $P = KR(I, -\tilde{c})$.其中: K 是摄影机内参数矩阵, $R(I, -\tilde{c})$ 是外参数矩阵,即摄像机的姿态.本文仅考虑摄像机姿态估计问题,即摄像机内参数是已知的,不失一般性,假定内参数 $K=I$.此时,投影矩阵则可以被表示为 $P = R(I, -\tilde{c})$.空间点的齐次坐标 x_i 和图像点齐次坐标 u_i 在针孔模型下存在映射关系,其表示为投影方程:

$$s_i u_i = P x_i = (R, -R\tilde{c}) x_i, i = 1, 2, \dots, N \tag{1.1}$$

姿态估计问题是根据图像点和空间点的对应信息估计旋转矩阵 R 和摄像机位置 c ,使之尽可能地满足投影方程(1.1).

假定空间点 x_i 在摄像机坐标系下的非齐次坐标为 $\tilde{x}_i^c = (x_i^c, y_i^c, z_i^c)^T$,它在图像上的真值为 \hat{u}_i, u_i 是 \hat{u}_i 的测量点.令 $\tilde{x}_i'^c = z_i^c u_i$,则代数误差为 $d_A = \|x_i'^c - \tilde{x}_i^c\|$,因此,重投影算法的几何误差为(如图 1 所示)

$$d_G = \|u_i - \hat{u}_i\| = \left\| u_i - \frac{R(\tilde{x}_i - \tilde{c})}{R^3(\tilde{x}_i - \tilde{c})} \right\| = \frac{1}{z_i^c} \| \tilde{x}_i'^c - \tilde{x}_i^c \| \tag{1.2}$$

其中, R^3 为旋转矩阵 R 的第 3 行.不难看出,代数误差 d_A 与几何误差 d_G 成正比例关系,即 $d_A = z_i^c d_G$.

假设图像点的噪声服从高斯分布,那么几何误差最小化估计在最大似然意义下是最优的.但在最小化过程

中,因分母 $R^3(\tilde{x}_i - \tilde{c})$ 存在未知项而无法进行线性求解.相对于几何误差最小化的方法,最小化代数误差虽然可以线性求解,但是对深度变化较为敏感,如果深度变化剧烈,最小化代数误差往往得不到很好的估计.根据代数误差和几何误差的关系,如果几何误差相同,空间点的深度越大,其代数误差越大;深度越小,其代数误差越小.如果在使用最小二乘算法的过程中忽略了以上误差关系,那么优化时会对于大的误差项有所偏重,因而在一定程度上忽略了小误差项.因此,对于深度较大的数据项应赋予较小的权值,深度较小的数据项赋予较大的权值,使得每个数据项的代数误差相对于整体误差有同等贡献,线性算法的估计精度就能被提高.这就是本文所提加权线性算法的基本思想.

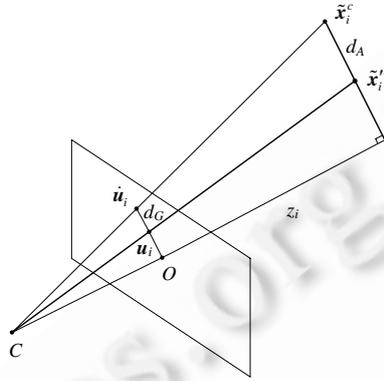


Fig.1 Geometric error and algebraic error
图 1 代数误差和几何误差

在下面的两节中,我们将给出两种具体的加权线性算法,它们分别是加权 DLT 算法和加权 EPnP 算法.

2 加权 DLT 算法

2.1 DLT 算法

由向量 u_i 左叉乘投影方程(1.1),我们得到:

$$[u_i]_{\times} P x_i = 0, i=1, 2, \dots, N \tag{2.1}$$

其中,

$$[u_i]_{\times} = \begin{bmatrix} 0 & -1 & v_i \\ 1 & 0 & -u_i \\ -v_i & u_i & 0 \end{bmatrix} \tag{2.2}$$

由于反对称矩阵 $[u_i]_{\times}$ 的秩为 2,所以对于每个点对应,方程(2.1)只能提供两个独立方程^[9].如果忽略旋转矩阵 R 的正交约束,将投影矩阵 P 的每个元素作为独立变量,通过等式(2.1)、等式(2.2)可以推导出这两个独立方程:

$$\begin{pmatrix} 0 & -x_i^T & v_i x_i^T \\ x_i^T & 0 & -u_i x_i^T \end{pmatrix} \begin{pmatrix} P^{1T} \\ P^{2T} \\ P^{3T} \end{pmatrix} = 0 \tag{2.3}$$

其中, P^i 是投影矩阵 P 的第 i 行.因此,使用 $N(\geq 6)$ 个点对应,我们得到投影矩阵 P 的线性方程:

$$M p = 0, p = (P^1, P^2, P^3)^T \tag{2.4}$$

其中,矩阵 M 是由 N 个点对应提供的 $2N$ 个线性方程所组成的 $2N \times 12$ 的矩阵,即

$$M = \begin{pmatrix} 0 & -\mathbf{x}_1^T & v_1 \mathbf{x}_1^T \\ \mathbf{x}_1^T & 0 & -u_1 \mathbf{x}_1^T \\ \vdots & \vdots & \vdots \\ 0 & -\mathbf{x}_N^T & v_N \mathbf{x}_N^T \\ \mathbf{x}_N^T & 0 & -u_N \mathbf{x}_N^T \end{pmatrix} \quad (2.5)$$

DLT 算法分为下述两个步骤:

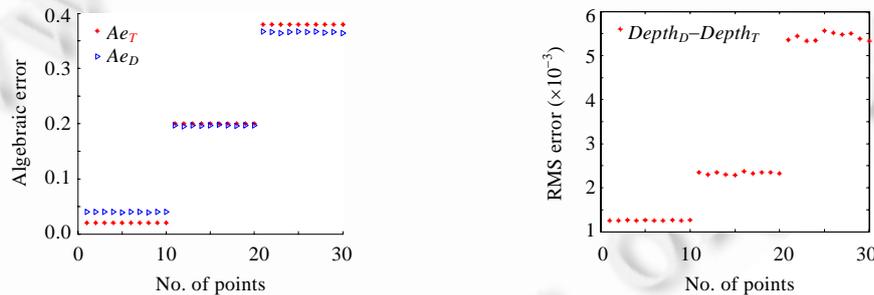
- (1) 在约束条件 $\|p\|=1$ 下,求线性方程(2.4)的最小二乘解,它是矩阵 M 的最小奇异值的右奇异向量,得到投影矩阵 P 的最小二乘估计,记为 $P_{LS}=(A,b)$;
- (2) 置 $P_{LS}=(\hat{A},\hat{b})$, 其中, $\hat{A} = \text{sign}(\det(A))(\det(A))^{-1/3} A$, $\hat{b} = \text{sign}(\det(A))(\det(A))^{-1/3} b$, 则摄像机的位置为 $\tilde{c} = -\hat{A}^{-1}\hat{b}$. 摄像机的旋转 R 是 \hat{A} 的最佳旋转近似,即 $R=UV^T$,其中, U,V 分别是 \hat{A} 的左、右奇异向量矩阵.

2.2 加权DLT算法

不难看出,DLT 算法的最小二乘估计是最小化代数距离 $\|[\mathbf{u}_i]_x P \mathbf{x}_i\|$ 的平方和,即 $\|s_i \mathbf{u}_i - P \mathbf{x}_i\|$ 的平方和.因此,在 DLT 算法步骤 1 中的最小二乘估计 P_{LS} 可被等价表述为

$$P_{LS} = \arg \min_{\|p\|_F=1} \sum_{i=1}^N \|P^3 \mathbf{x}_i \cdot \mathbf{u}_i - P \mathbf{x}_i\|^2 \quad (2.6)$$

其中, $P^3 \mathbf{x}_i = z_i^c$ 是空间点 \mathbf{x}_i 的深度, $\|\cdot\|_F$ 是矩阵的 F-范数.因此,DLT 算法忽略了各数据代数误差之间的关系,必然对大误差项有所偏重,而在一定程度上忽略了小误差项,从而使估计的代数误差严重偏离真实的代数误差,在小误差项上尤为明显,如图 2(a)所示(其中 30 个点对应(分为 3 组,每组深度分别为 10,100,190),图像点添加 1 像素的等距噪声,图 2(b)与此相同).好的估计的代数误差应和真实的代数误差相一致,但从图 2(a)可以看出,DLT 算法估计的代数误差严重偏离了真实代数误差(小深度所对应 DLT 估计误差大于真实误差,大深度所对应的 DLT 估计误差小于真实误差),因此,DLT 算法的估计 P_{LS} 不是一种好的估计.



(a) Algebraic errors of DLT (Ae_D) and true algebraic errors (Ae_T)
 (a) DLT 的代数误差和真实代数误差
 (b) RMS of depth error
 (b) 深度误差的均方根

Fig.2 Mean of algebraic errors and RMS of depth error of 10 000 independent experiments

图 2 10 000 次独立实验的代数误差的均值和深度误差的均方根

DLT 算法是一种最基本的算法,虽然在数据噪声存在的情况下它不能提供高精度的摄像机位姿估计,但是通过大量模拟数据实验我们发现,DLT 算法所估计的空间点的深度与真实深度具有一致性.即真实深度较大,DLT 所估计的深度也较大,如图 2(b)所示.根据加权的基本思想,如果将 DLT 所估计深度的倒数作为对应数据项的权值,则通过最小化加权代数误差的平方和,就能保证每个数据项对整体误差贡献一致.因此,我们提出下面的加权 DLT 算法:

- 深度的最小二乘估计:由 DLT 算法步骤 1 计算投影矩阵 $P_{LS}=(A,b)$,然后通过公式 $s_i^* = P_{LS}^3 \mathbf{x}_i$ 得到空间

点 \mathbf{x}_i 的深度估计;

- 投影矩阵的加权最小二乘估计:求解 $\mathbf{P}_{WLS} = \arg \min_{\|\mathbf{P}\|_F=1} \sum_{i=1}^N \|(s_i^*)^{-1}(\mathbf{P}^3 \mathbf{x}_i \cdot \mathbf{u}_i - \mathbf{P} \mathbf{x}_i)\|^2$, 得到投影矩阵 \mathbf{P}_{WLS} . 这可以通过下述方法来实现:令 $\mathbf{M}_w = \text{diag}^{-1}(s_1^* \mathbf{I}_2, s_2^* \mathbf{I}_2, \dots, s_N^* \mathbf{I}_2) \mathbf{M}$, 然后对矩阵 \mathbf{M}_w 进行 SVD 分解, 得到投影矩阵 \mathbf{P}_{WLS} ;
- 摄像机位姿估计:以 \mathbf{P}_{WLS} 替换 DLT 算法中的 \mathbf{P}_L , 用 DLT 算法步骤 2 计算摄像机位姿.

加权 DLT 算法与 DLT 算法一样, 都可以通过 SVD 分解技术来实现. 在计算量方面, 加权 DLT 算法是 DLT 算法的两倍. 但大量的实验结果表明, 加权 DLT 算法能够有效提升 DLT 算法的估计精度.

3 加权 EPnP 算法

3.1 EPnP 算法

EPnP 算法的基本思想是:首先选取 4 个非共面的虚拟点作为控制点, 然后根据空间点和控制点的位置关系以及空间点图像, 求解控制点在摄像机坐标系下的坐标, 从而得到摄像机位姿. 为了加以区别, 我们用右上标 w 来表示世界坐标系下的坐标, 右上标 c 表示摄像机坐标系下的坐标. 假定 $\{\tilde{\mathbf{z}}_i^w : i=1, 2, 3, 4\}$ 是 4 个控制点在世界坐标系下的非齐次坐标, 这 4 个控制点通常被选择作为空间点集 $\{\mathbf{x}_i^w : i=1, 2, \dots, N\}$ 的重心和 3 个主方向的单位点, 令 $\{\tilde{\mathbf{z}}_i^c : i=1, 2, 3, 4\}$ 是它们在摄像机坐标系下的非齐次坐标, $\mathbf{x}_i^c = (\tilde{\mathbf{x}}_i^{cT}, 1)^T$ 为空间点在摄像机坐标系下的齐次坐标, 则空间点 $\mathbf{x}_i^w = (\tilde{\mathbf{x}}_i^{wT}, 1)^T$ 与控制点世界坐标 $\{\tilde{\mathbf{z}}_i^w : i=1, 2, 3, 4\}$ 存在下述关系:

$$\tilde{\mathbf{x}}_i^w = \sum_{j=1}^4 a_{ij} \tilde{\mathbf{z}}_j^w, i=1, 2, \dots, N \quad (3.1)$$

其中, 向量 $(a_{i1}, a_{i2}, a_{i3}, a_{i4})^T$ 为空间点 $\tilde{\mathbf{x}}_i^w$ 在以控制点 $\{\tilde{\mathbf{z}}_i^w : i=1, 2, 3, 4\}$ 为基的欧式空间中的坐标, 即

$$(a_{i1}, a_{i2}, a_{i3}, a_{i4})^T = [\mathbf{z}_1^w, \mathbf{z}_2^w, \mathbf{z}_3^w, \mathbf{z}_4^w]^{-1} \tilde{\mathbf{x}}_i^w, i=1, 2, \dots, N \quad (3.2)$$

根据线性关系在欧式变换下的不变性,

$$\tilde{\mathbf{x}}_i^c = \sum_{j=1}^4 a_{ij} \tilde{\mathbf{z}}_j^c, i=1, 2, \dots, N \quad (3.3)$$

于是,

$$s_i \mathbf{u}_i = \tilde{\mathbf{x}}_i^c = \sum_{j=1}^4 a_{ij} \tilde{\mathbf{z}}_j^c, i=1, 2, \dots, N \quad (3.4)$$

其中, s_i 表示空间点 $\mathbf{x}_i^w = (\tilde{\mathbf{x}}_i^{wT}, 1)^T$ 的投影深度. 令 $\tilde{\mathbf{z}}_i^c = (x_i^c, y_i^c, z_i^c)^T$, 则有

$$s_i = \sum_{j=1}^4 a_{ij} z_j^c, i=1, 2, \dots, N \quad (3.5)$$

因此, 方程(3.4)可以重写为

$$\sum_{i=1}^N \sum_{j=1}^4 a_{ij} (z_j^c \mathbf{u}_i - \tilde{\mathbf{z}}_j^c) = 0, i=1, 2, \dots, N \quad (3.6)$$

令 $\mathbf{z} = (\tilde{\mathbf{z}}_1^{cT}, \tilde{\mathbf{z}}_2^{cT}, \tilde{\mathbf{z}}_3^{cT}, \tilde{\mathbf{z}}_4^{cT})^T \in R^{12}$, 则方程(3.6)的矩阵形式为

$$\mathbf{M} \mathbf{z} = 0 \quad (3.7)$$

其中,

$$\mathbf{M} = \begin{pmatrix} a_{11} & 0 & -a_{11}u_1 & a_{12} & 0 & -a_{12}u_1 & a_{13} & 0 & -a_{13}u_1 & a_{14} & 0 & -a_{14}u_1 \\ 0 & a_{11} & -a_{11}v_1 & 0 & a_{12} & -a_{12}v_1 & 0 & a_{13} & -a_{13}v_1 & 0 & a_{14} & -a_{14}v_1 \\ \vdots & \vdots \\ a_{N1} & 0 & -a_{N1}u_N & a_{N2} & 0 & -a_{N2}u_N & a_{N3} & 0 & -a_{N3}u_N & a_{N4} & 0 & -a_{N4}u_N \\ 0 & a_{N1} & -a_{N1}v_N & 0 & a_{N2} & -a_{N2}v_N & 0 & a_{N3} & -a_{N3}v_N & 0 & a_{N4} & -a_{N4}v_N \end{pmatrix} \quad (3.8)$$

因此, z 的解是矩阵 M 的核空间, 对于透视模型矩阵 M 的零空间的维数为 1. 在图像数据有噪声时, z 的最小二乘解 z_{LS} 是矩阵 M 的最小奇异值的右奇异向量. 由于方程(3.6)是一个齐次线性方程, 因此, z_{LS} 与控制点的摄像机坐标向量相差一个常数因子 s , 这个常数因子 s 可以通过欧氏变换的保距性来确定.

根据控制点的世界坐标 \tilde{z}_i^w 及其摄像机坐标 \tilde{z}_i^c 以及方程(3.1)、方程(3.3), 得到空间点的世界坐标 \tilde{x}_i^w 和它的摄像机坐标 \tilde{x}_i^c , 即得到摄像机位姿的线性方程组如下所示:

$$R^T \tilde{x}_i^c + \tilde{c} = \tilde{x}_i^w, i = 1, 2, \dots, N \tag{3.9}$$

根据此方程组, 通过使用 Horn 绝对定位算法^[14]得到摄像机姿态估计 (R_{LS}, \tilde{c}_{LS}) . 总结上述讨论, EPnP 算法可被如下概括:

- 选择空间点集 $\{x_i^w : i = 1, 2, \dots, N\}$ 的重心和 3 个主方向的单位点作为 4 个控制点 $\{\tilde{z}_i^w : i = 1, 2, 3, 4\}$, 根据公式(3.8)构造矩阵 M , 通过求解 M 的最小奇异值的右奇异向量得到 4 个控制点在摄像机坐标系下的坐标;
- 通过方程(3.1)、方程(3.3)得到空间点的世界坐标 \tilde{x}_i^w 和它的摄像机坐标 \tilde{x}_i^c , 通过使用 Horn 绝对定位算法^[14]得到摄像机姿态估计 (R_{LS}, \tilde{c}_{LS}) .

EPnP 算法是一种精度较高的快速线性算法^[4]. 由于控制点坐标是相互独立的, 所以求解控制点在摄像机坐标系下的坐标避免了未知量之间的交互. 因此, EPnP 算法在一定程度上提高了 DLT 算法的精度. 但从实验结果(如后文图 3(b)所示)可以看出, 在小深度比时, 即深度变化剧烈的情况下, EPnP 算法的估计精度相比 ML 算法在大深度比的情况下明显要差, 因此对深度变化剧烈的数据, EPnP 算法不能提供好的估计. 将本文的加权基本思想应用于 EPnP 算法, 能够显著地提高 EPnP 算法的估计精度, 特别是对深度变化较大的数据.

3.2 加权EPnP算法(WEPnP)

由方程(3.7)求出的最小二乘解 z_{LS} 等价于下述最小二乘问题:

$$z_{LS} = \arg \min_{\|z\|_r=1} \sum_{i=1}^N \left\| \sum_{j=1}^4 a_{ij} (z_j^c u_i - \tilde{x}_i^c) \right\|^2 \tag{3.10}$$

所以, 在 EPnP 算法第 1 步中, 最小二乘解 z_{LS} 与 DLT 算法一样, 也是最小化代数距离的平方和. EPnP 算法的第 2 步是绝对定位问题, 在最小化过程中同样需要施加约束: 深度越大的空间点的代数误差越大; 深度越小的空间点的代数误差越小, 即施行加权绝对定位算法才能获得更好的估计. 因此, 本文所提加权基本思想也适用于 EPnP 算法. 下面仅就加权绝对定位算法给出详尽描述:

记 \tilde{x}_i^c 为空间点摄像机坐标系下的坐标, \tilde{x}_i^w 为空间点世界坐标系下的坐标, 满足:

$$\tilde{x}_i^c = R(\tilde{x}_i^w - \tilde{c}), i = 1, 2, \dots, N \tag{3.11}$$

则有,

$$\tilde{x}_i^c - \bar{x}^c = R(\tilde{x}_i^w - \bar{x}^w), i = 1, 2, \dots, N \tag{3.12}$$

其中,

$$\bar{x}^c = \frac{1}{N} \sum_{i=1}^N \tilde{x}_i^c, \bar{x}^w = \frac{1}{N} \sum_{i=1}^N \tilde{x}_i^w \tag{3.13}$$

在等式(3.12)左、右两边分别除以对应空间点的深度 $s_i^* = z_i^c$, 即

$$\frac{1}{s_i^*} (\tilde{x}_i^c - \bar{x}^c) = \frac{1}{s_i^*} R(\tilde{x}_i^w - \bar{x}^w), i = 1, 2, \dots, N \tag{3.14}$$

通过以下等式构造矩阵 D ,

$$D = \sum_{i=1}^N \frac{1}{s_i^{*2}} (\tilde{x}_i^c - \bar{x}^c)(\tilde{x}_i^w - \bar{x}^w)^T \tag{3.15}$$

根据 Horn 的绝对定位算法^[14]对矩阵 D 进行 SVD 分解, $D = U \Sigma V^T$, 获得旋转矩阵 R 的估计:

$$R = UV^T \tag{3.16}$$

摄像机位置可通过公式(3.11)得到:

$$\tilde{c} = \bar{x}^w - R^T \bar{x}^c \quad (3.17)$$

综上所述,WEPnP 算法有下述具体步骤:

- 初始估计:根据系数矩阵 $[a_{ij}]$ 和图像点构造矩阵 M 对矩阵 M 进行 SVD 分解,并由奇异值的右奇异向量的线性组合表示控制点坐标 z_{MLS} .根据控制点之间的距离不变性,得到控制点最小二乘估计 z_{MLS} .然后,根据等式(3.5)可以估计出每个点的深度值 s_i^* ;
- 加权求精:用系数矩阵 $[a_{ij}]$ 的第 i 行除以对应深度 s_i^* ,代数误差最小二乘经过加权后变成

$$\min \sum_{i=1}^N \left\| \frac{1}{s_i^*} (z_i^c u_i - x_i^c) \right\|^2.$$

可以看出,经过加权后的误差函数近似重投影误差.然后,使用新的系数矩阵再次构造矩阵 M ,对矩阵 M 进行 SVD 分解,得到求精后的投影矩阵 z_{MLS} ;

- 加权绝对定位:利用欧氏变换的保距性恢复出控制点在摄像机坐标系下的坐标;然后求取全部空间点的摄像机坐标 x_i^c ,并计算空间点集的中心在摄像机坐标系和世界坐标系下的坐标 \bar{x}^c 和 \bar{x}^w ;最后,根据等式(3.15)构造矩阵 D ,通过 SVD 算法得到旋转矩阵 R 和摄像机坐标 c .

4 实验结果

为了验证加权线性算法的有效性,本节在模拟数据和真实图像上比较以下算法的估计精度:

- DLT:经典 DLT 算法^[1];
- EPnP:Moreno 和 Lepetit 提出的 EPnP 算法^[4];
- ML:重投影误差最小化迭代算法^[9];
- Lu:Lu 的空间距离平方和最小化迭代算法^[6];
- WDLT:本文提出的加权 DLT 算法;
- WEPnP:本文提出的加权 EPnP 算法.

4.1 模拟数据

在模拟数据实验中,假定内参数 K 为单位矩阵.图像点平均分布在 $[-1,1] \times [-1,1]$ 的范围内,空间点的最大深度为 150,深度比($\min(s_i)/\max(s_i)$)的范围是 0.1~0.8.在深度和图像点已知的情况下,根据空间点在摄像机坐标系下的坐标 \tilde{x}_i^c 与深度和图像点的关系 $\tilde{x}_i^c = s_i u_i$,得到坐标真值 \tilde{x}_i^c .给定旋转矩阵真实值 R_{true} 和光心坐标 c_{true} ,根据等式(3.11)计算空间点在世界坐标系下的世界坐标真值 x_i .实验中使用旋转矩阵误差($\|R - R_{true}\|_F$)和光心误差($\|c - c_{true}\|$)来评估各种算法的估计精度.值得指出的是,ML 算法和 Lu 算法都是非线性算法,并且 ML 算法在高斯噪声下是最优的,我们不能指望存在一种线性算法能够优于 ML 算法.如果一种线性算法的估计精度接近于 ML 算法的估计精度,我们认为它是一种好的线性算法.

- 深度比实验

深度比($\min(s_i)/\max(s_i)$)的范围为 $[0.1,0.8]$,点数对应数 $N=80$.在每次实验中,随机选取图像点、深度、旋转矩阵和光心坐标,并在图像点添加均值为 0 标准差为 1 个像素的高斯噪声.图 3 是各种算法进行 10 000 次独立随机实验的旋转矩阵和光心的误差均方根.

- * 从图 3(a)可以看出,WDLT 有效地提高了 DLT 的估计精度.对于旋转矩阵,当深度比小于 0.35 时,WDLT 的估计精度明显优于 Lu 的非线性算法的估计精度;对于摄像机位置,当深度比小于 0.15 时,WDLT 的估计精度也明显优于 Lu 的非线性算法的估计精度.对于小深度比,WDLT 算法接近于 ML 算法.
- * 从图 3(b)可以看出,WEPnP 也有效地提高了 EPnP 的估计精度.在更大深度比的范围内(旋转矩阵,深度比小于 0.55;摄像机位置,深度比小于 0.35),WEPnP 算法的估计精度高于 Lu 的非线性算法的估计精度,并且接近于 ML 算法的估计精度.

* 从图 3(a)和图 3(b)可以看出,WEP_nP 算法的估计精度显著地高于 WDLT 算法的估计精度,更接近于 ML 算法.

总之,对于深度比实验,加权线性算法(WDLT,WEP_nP)比原始线性算法(DLT,EP_nP)具有更好的数值稳定性和更高的估计精度.WEP_nP 在更大深度比的范围内优于 Lu 的非线性算法,并且更接近于最优 ML 算法.

• 图像噪声的实验

本实验的目的是比较各种算法随图像噪声变化的估计精度.在实验中,深度比被设置为 0.3,点数对应数 $N=80$,随机选取图像点、深度、旋转矩阵和光心坐标.图像点添加均值为 0、标准差为 σ 的高斯噪声, σ 在[0.3,3]内变化(步长为 0.3 个像素).对每个噪声水平进行 10 000 次独立随机实验,图 4 是各种算法的实验结果.

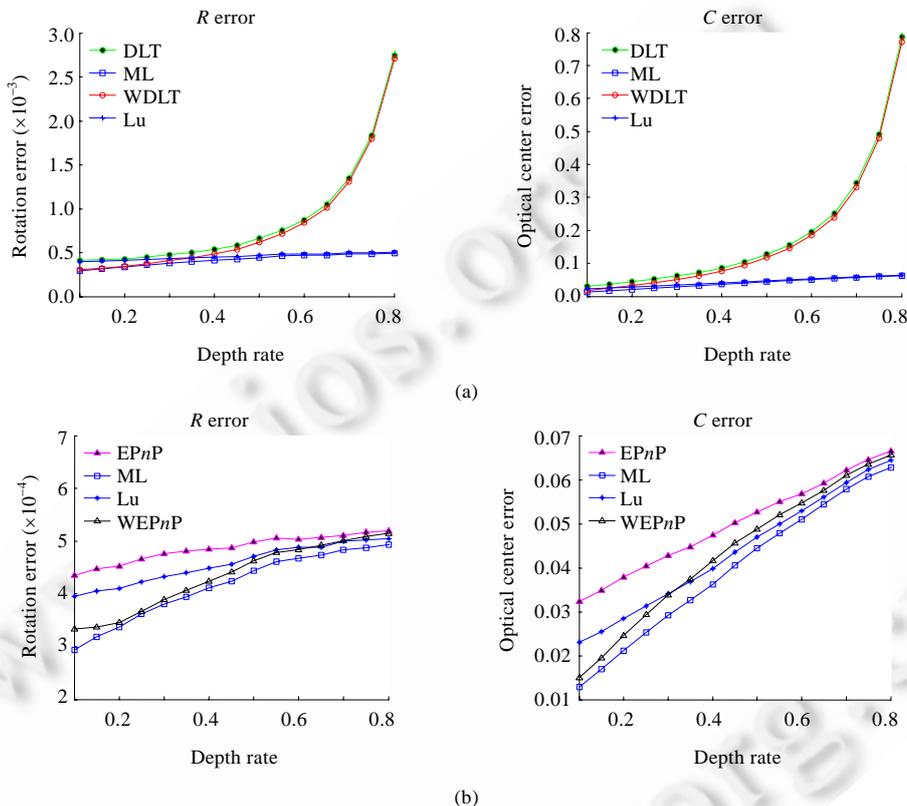


Fig.3 Results of depth ratio experiments

图 3 深度比的实验结果

由图 4 我们可以看出:

- * 对于旋转矩阵,DLT 和 EP_nP 两种算法具有相似的估计精度.随着噪声的增加,EP_nP 的精度略高于 DLT,它们在各种算法中是最差的;WDLT 和 WEP_nP 两种加权线性算法都显著地提高了相应线性算法的估计精度,并且都优于 Lu 的非线性算法;WEP_nP 的精度比 WDLT 的精度要高,它非常接近于最优 ML 算法.
- * 对于摄像机位置,DLT 的估计精度最差.虽然 WDLT 显著地提高了 DLT 的估计精度,但它仍劣于其他算法;WEP_nP 显著地提高了 EP_nP 的估计精度,与 Lu 的非线性算法有相同的估计精度,与最优 ML 算法的估计精度也非常接近.

总之,对于图像噪声实验,两种加权线性算法(WDLT,WEP_nP)比相应的原始线性算法(DLT,EP_nP)具有更高的估计精度.WEP_nP 算法比 WDLT 算法具有更好的抗噪能力,其估计精度优于 Lu 的非线性算法,并且十分接近于最优 ML 算法.

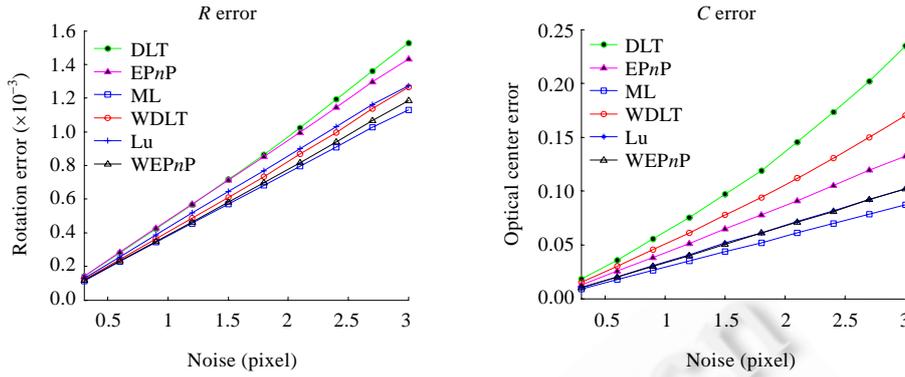


Fig.4 Results of noise experiments

图 4 图像噪声的实验结果

• 对应点数目的实验

本实验的目的是比较各种算法随对应点数目变化的估计精度.在实验中,深度比被设置为 0.3,随机选取图像点、深度、旋转矩阵和光心坐标,图像点添加均值为 0、标准差为 1 的高斯噪声.点对应数目的变化范围为 [20,100],变化步长为 10.对每个对应点数目进行 10 000 次独立随机实验,图 5 是各种算法的实验结果.

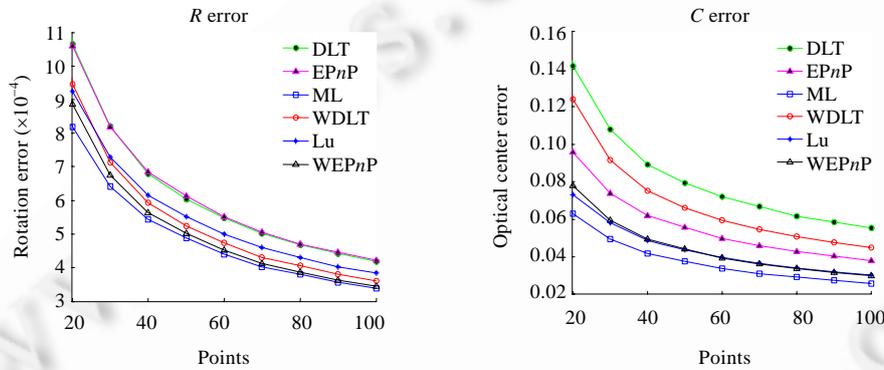


Fig.5 Results of points experiments

图 5 对应点数目的实验结果

从图 5 我们可以看出:所有算法的估计精度都随点对应数目的增加而有所提高.对于所有的对应点数目,两种加权线性算法 WDLT 和 WEPnP 都提高了相应原始线性算法的估计精度,WEPnP 算法优于 Lu 的非线性算法,并且十分接近于最优 ML 算法.

• 实验总结

加权线性算法 WDLT 和 WEPnP 比相应的原始线性算法 DLT 和 EPnP 具有更好的估计精度.WEPnP 算法优于 Lu 的非线性算法,其估计精度十分接近于 ML 算法.因此,WEPnP 算法是一种优秀的线性算法,是一种在实践中值得推荐的线性算法.

4.2 真实图像

在真实图像实验中,选取 Nikon D40 拍摄的 8 幅不同角度的图像,如图 6 所示.每幅图像的大小为 3008×2000,相机的焦距为 2 320,每对图像手工选取 108 个点对应(深度比为 0.6 左右).选取标定块垂直面的交点为世界坐标系的原心,3 条垂直边分别为世界坐标系的 x 轴、y 轴和 z 轴.根据标定块的结构关系,我们可以得到空间点在世界坐标系下的真实坐标.标定算法选取 Heikkilä 的标定算法^[15],使用 Heikkilä 提供的 Camera Calibration Toolbox for Matlab^[16]对真实图像进行标定,再根据估计的畸变因子对图像进行矫正,最后用各种算法计算摄像机的外参数.在实验中,图像点添加均值为 0、标准差为 σ 的高斯噪声, σ 在 [0,1] 内变化(步长为 0.5 个像素).对每个噪声水平

进行 500 次独立随机实验,进而模拟更多图像.因为无法得到真实的摄像机外参数,所以使用重投影误差和重构误差的均方根来间接衡量各种算法的精度.图 7 是各种算法的实验结果.



Fig.6 Real images

图 6 真实图像

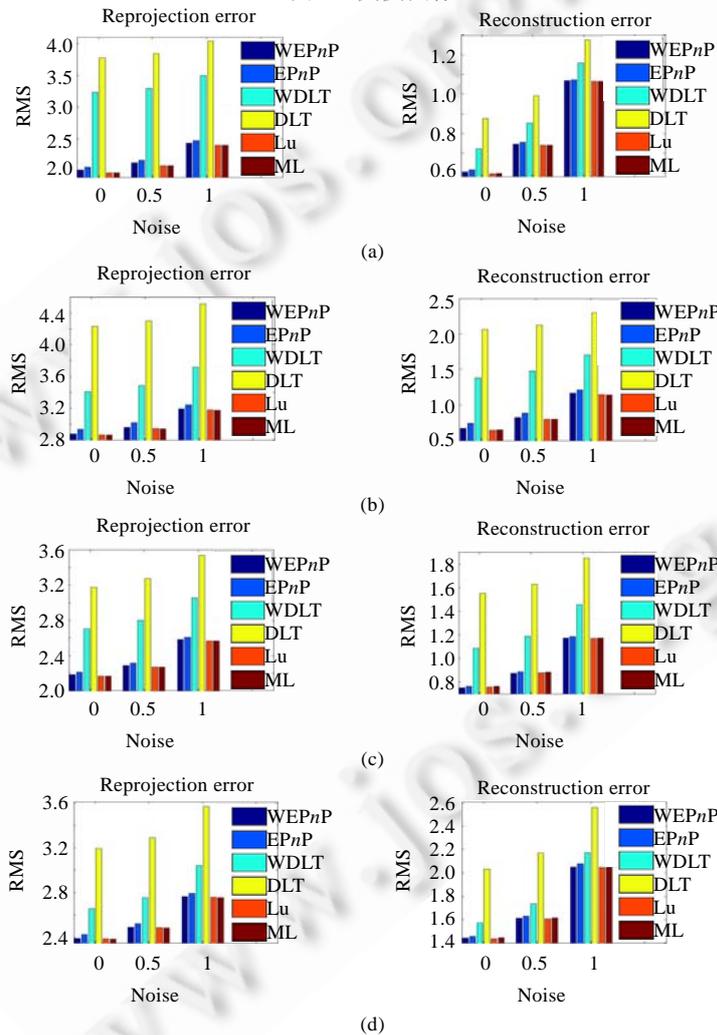


Fig.7 Results of experiments on real data

图 7 真实实验结果

由图 7 我们可以看出:

- 对于重投影误差均方根,WDLT 和 WEPnP 两种加权线性算法的重投影误差均方根明显小于相应线性算法的重投影误差均方根;WEPnP 的重投影误差均方根明显比 WDLT 的重投影误差均方根要小,它最为接近于最优 ML 算法.
- 对于重构误差均方根,DLT 的重构误差均方根最大.虽然 WDLT 重构误差均方根明显小于 DLT 的重构误差均方根,但它仍大于其他算法;WEPnP 重构误差均方根明显小于 EPnP 的重构误差均方根,与 Lu 的非线性算法和最优 ML 算法非常接近.

总之,对于真实图像实验,两种加权线性算法(WDLT,WEPnP)比相应原始线性算法(DLT,EPnP)具有更高的估计精度.WEPnP 算法比 WDLT 算法具有更好的估计精度,且十分接近于 Lu 的非线性算法和最优 ML 算法.

5 结束语

本文利用加权求精的思想,使得加权后的代数误差近似于几何误差,给出了快速且精度较高的 WDLT 算法和 WEPnP 算法.两种加权线性算法 WDLT 和 WEPnP 都明显提高了相应原始线性算法的估计精度,并且在深度变化剧烈的情况下,WDLT 和 WEPnP 精度都可以接近最优 ML 算法,甚至比 Lu 的算法更为精确,并且计算量大为减小.大量实验结果表明,WDLT 算法和 WEPnP 算法与其他线性算法相比,能够更有效地提高精度.

References:

- [1] Abdel-Aziz YI, Karara HM. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. In: Proc. of the ASP/UI Symp. on Close-Range Photogrammetry. 1971. 1–18.
- [2] Quan L, Lan ZD. Linear N -point camera pose determination. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1999,1(8): 774–780. [doi: 10.1109/34.784291]
- [3] Fiore PD. Efficient linear solution of exterior orientation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001,23(2): 140–148. [doi: 10.1109/34.908965]
- [4] Lepetit V, Moreno-Noguer F, Fua P. EPnP: An accurate $O(n)$ solution to the PnP problem. Int'l Journal of Computer Vision, 2009, 81(2):155–166. [doi: 10.1007/s11263-008-0152-6]
- [5] Ansar A, Daniilidis K. Linear pose estimation from points or lines. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2003,25(5):578–589. [doi: 10.1109/TPAMI.2003.1195992]
- [6] Lu CP, Hager GD, Mjolsness E. Fast and globally convergent pose estimation from video images. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2000,22(6):610–622. [doi: 10.1109/34.862199]
- [7] Hartley RI, Kahl F. Global optimization through rotation space search. Int'l Journal of Computer Vision, 2009,82(1):64–79. [doi: 10.1007/s11263-008-0186-9]
- [8] Schweighofer G, Pinz A. Globally optimal $O(n)$ solution to the PnP problem for general camera models. In: Proc. of the British Machine Vision Conf. London: BMVA Press, 2008. 1–10.
- [9] Hartley R, Zisserman A. Multiple View Geometry in Computer Vision. Cambridge: Cambridge University Press, 2000.
- [10] Fischler MA, Bolles RC. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 1981,24(6):381–395. [doi: 10.1145/358669.358692]
- [11] Haralick RM, Lee CN, Ottenberg K, Nölle M. Review and analysis of solutions of the three point perspective pose estimation problem. Int'l Journal of Computer Vision, 1994,13(3):331–356. [doi: 10.1007/BF02028352]
- [12] Gao XS, Hou XR, Tang JL, Cheng HF. Complete solution classification for the perspective-three-point problem. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2003,25(8):930–943. [doi: 10.1109/TPAMI.2003.1217599]
- [13] Wu WJ. Basic Principles of Mechanical Theorem Proving in Geometries, vol. I: Part of Elementary Geometries. Beijing: Science Press, 1984 (in Chinese).
- [14] Horn BKP, Hilden HM, Negahdaripour S. Closed-Form solution of absolute orientation using orthonormal matrices. Journal of the Optical Society of America, 1988,5(7):1127–1135. [doi:10.1364/JOSAA.5.001127]

- [15] Heikkilä J. Geometric camera calibration using circular control points. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2000,22(10):1066–1077. [doi: 10.1109/34.879788]
- [16] Camera Calibration Toolbox for Matlab. <http://www.ee.oulu.fi/~jth/calibr/>

附中文参考文献:

- [13] 吴文俊.几何定理机器证明的基本原理(初等几何部分).北京:科学出版社,1984.



杨森(1985—),男,河南禹州人,硕士,主要研究领域为计算机视觉.



吴福朝(1957—),男,研究员,博士生导师,主要研究领域为计算机视觉,模式识别.



中国科学院软件研究所诚聘英才

中国科学院软件研究所成立于1985年3月1日,是一所致力于计算机科学理论和软件高新技术的研究与发展的综合性基地型研究所。秉承“创新求实,永竞一流”的精神,软件研究所不断迈上新的台阶,目前已形成了以多个国家重点实验室和国家工程研究中心为龙头的基础前沿研究、战略高新技术研究和高技术应用三大科研体系,建立了计算机科学、计算机软件、计算机应用技术等重点学科领域和相关重点学科方向的学科布局。为进一步加强软件所人才队伍建设,现面向海内外诚聘英才。

一、招聘方向

计算机科学与软件理论,基础软件技术与系统,互联网信息处理的理论、方法与技术,综合信息系统技术四个重点学科。

二、申请条件

1. 重点学科(方向)领军人物申请条件

(1) 申请人应为业界公认的世界级专家(相当于国际著名大学担任正教授职位),取得过国际同行公认的重要研究成果,如:发表过多篇有重要影响的学术论文,或掌握关键技术、拥有重大发明专利等;

(2) 申请人应能够把握学科发展方向,具备创新性构想,并能够组织和带动研究团队做出具有国际领先水平的创新成果;

(3) 申请人年龄原则上不超过55岁(申请时),并能在协议期内全时专职投入。

2. 杰出青年人才申请条件

(1) 申请人年龄在35周岁以内,具有博士学位或副高级及以上的专业技术职称;

(2) 申报的研究方向创新性突出,有明显的理论意义或应用前景,个人具备扎实的研究基础和突出的研究成果;

(3) 对于国家杰出青年科学基金、中国青年科技奖、中国青年科学家奖、中国科学院杰出成就奖(个人)和国家三大科技奖二等奖以上(排名前3)获得者,可直接入选,且不受本期支持指标数控制。

三、申请时间

重点学科(方向)领军人物申请截止时间:2011年12月25日

杰出青年人才申请截止时间:2011年10月8日

详细情况请登陆软件研究所主页(<http://www.iscas.ac.cn>)中的中长期发展规划专栏查阅。

四、联系方式

地址:北京中关村南四街4号,中国科学院软件研究所

邮编:100190

联系人:人力资源处 张会涵

电话:010-62661016 E-mail: huihan@iscas.ac.cn