

基于 Legendre 神经网络的单目 视觉相对位姿解算

郝 帅,程咏梅,马 旭,赵建涛,崔蓝月

(西北工业大学 自动化学院,陕西 西安 710072)

摘 要: 将 BP 神经网络、RBF 神经网络以及 Legendre 正交基神经网络应用于单目视觉相对位姿强非线性解算,并且在梯度下降法的基础上推导了 Legendre 正交基神经网络的最优权值,最后进行了相应的仿真及物理实验.实验结果表明,基于 Legendre 正交基神经网络的单目视觉相对位姿解算方法不仅精度高、鲁棒性强,而且实时性较好.

关键词: Legendre 正交基;神经网络;相对位姿解算;单目;视觉

中图分类号: TP391

文献标识码: A

文章编号: 1000-7180(2015)03-0041-05

Monocular Vision Relative Position and Attitude Solution Based on Legendre Neural Network

HAO Shuai, CHENG Yong-mei, MA Xu, ZHAO Jian-tao, CUI Lan-yue

(School of Automation, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: A method of monocular vision relative pose based on Legendre orthogonal basis neural network is proposed. The BP neural network, RBF neural network and Legendre orthogonal basis neural network are combined to solve the strong nonlinear solution in relative position and attitude solution algorithm for monocular vision. The optimal weights in Legendre orthogonal basis neural network are derived from gradient descent method. The developed method can deal with the problem of strong nonlinearity in relative position and attitude solution algorithm for monocular vision. The computer simulations and semi-physical simulations show that the proposed method can obtain satisfactory accuracy, robustness and real-time.

Key words: legendre orthogonal basis; neural network; relative position and attitude solution; monocular; vision

1 引言

计算机视觉相对导航由于其自主性强、近距离高等优点,在飞机自主着陆/着舰、空中加油以及航天器交会对接等领域有着广泛的应用^[1]. 相对位姿解算是视觉相对导航中一个重要环节,其解算的准确性及实时性往往决定着视觉相对导航的精度及可靠性. 基于视觉的相对位姿解算主要是通过图像像素坐标系、图像物理坐标系、摄像机坐标系以及世界坐标系之间的转换关系,推导出相机在世界坐标

系中的位置和姿态,从而进一步可以得到相机载体在世界坐标系中的位置和姿态.

文献[2-3]中利用 BP(Back Propagation)神经网络对摄像机进行标定,其中对摄像机外部参数的标定也就是相对位姿解算. 文献[4]中采用 RBF(Radical Basis Function)神经网络求解相对位姿. 相比于 BP 神经网络, RBF 神经网络具有全局逼近能力. 但是与 BP 神经网络一样, RBF 神经网络神经元之间的连接权值同样需要不断迭代训练才能获取,导致训练时间较长,而且逼近精度有限. 此外,目

收稿日期: 2014-12-20; 修回日期: 2014-05-22

基金项目: 国家自然科学基金项目(60702066); 西安市科技计划项目(CXY1350(2)); 西安科技大学培养基金项目(2014015)

前利用神经网络进行相对位姿解算时大都采用的双目摄像机,而对单目摄像机如何进行相对位姿的求解研究较少.

针对单目摄像机相对位姿非线性解算问题,本文采用 Legendre 正交基神经网络进行求解.首先在摄像机成像模型的基础上推导了单目摄像机相对位姿解算方程中的独立参数,进而确定神经网络的输入和输出;然后基于梯度下降法推导出神经网络的最优权值;最后将 BP 神经网络、RBF 神经网络以及 Legendre 神经网络应用于单目视觉相对位姿非线性解算问题中.仿真实验及物理实验验证了本文所提出的算法在求解相对位姿参数时的鲁棒性与实时性.

2 摄像机模型

相对位姿求解主要是建立图像像素坐标 (u, v) 与世界坐标系 (x_w, y_w, z_w) 之间的映射关系.为了推导出像素点坐标和世界坐标之间的关系,还需要用到图像物理坐标系以及摄像机坐标系.

2.1 图像像素坐标系与图像物理坐标系

图像像素坐标系 uO_1v 与图像物理坐标系 XO_1Y 的关系如图 1 所示.

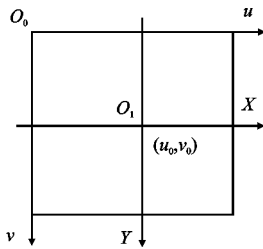


图 1 图像像素坐标系与图像物理坐标关系

图像像素坐标 (u, v) 表示该像点在图像中的行数和列数,它与图像物理坐标 (X, Y) 的关系如式(1)所示.

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & u_0 \\ 0 & \frac{1}{dy} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (1)$$

式中, (u_0, v_0) 为图像物理坐标原点在图像像素坐标系中的坐标, dx, dy 分别为每个像素在 X 轴与 Y 轴方向上的物理尺寸.

2.2 摄像机坐标系与世界坐标系

摄像机坐标系用 $o-xyz$ 来表示,世界坐标系用 $O_W - X_W Y_W Z_W$ 来表示,二者之间的关系如图 2 所示.其中 x 轴和 y 轴分别与图像物理坐标系中的 X

轴和 Y 轴平行, z 轴与光轴重合,并且与图像平面垂直, oO_1 为摄像机的焦距.

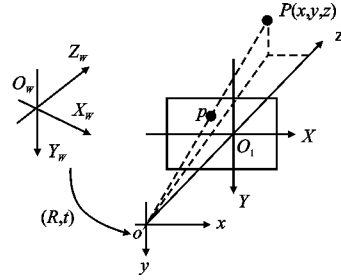


图 2 摄像机坐标系与世界坐标关系

设空间中某一点 P 在世界坐标系与摄像机坐标系下的坐标分别为 $(X_W, Y_W, Z_W, 1)^T$ 和 $(x, y, z, 1)^T$, 则有如下关系式:

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \quad (2)$$

式中, R 是 3×3 的旋转矩阵,它包含 3 个独立变量记为方位角 α 、俯仰角 γ 以及滚转角 β ,三个角度的定义可参考文献[5], $0^T = (0, 0, 0)$, $t = (t_x, t_y, t_z)^T$ 是三维平移向量. R 的表达式如式(3)所示.

$$R = \begin{bmatrix} \cos\alpha\cos\gamma & \sin\alpha\sin\beta\cos\gamma - \cos\beta\sin\gamma & \sin\alpha\cos\beta\cos\gamma + \sin\beta\sin\gamma \\ \cos\alpha\sin\gamma & \sin\alpha\sin\beta\sin\gamma + \cos\beta\cos\gamma & \sin\alpha\cos\beta\sin\gamma - \sin\beta\cos\gamma \\ -\sin\alpha & \cos\alpha\sin\beta & \cos\alpha\cos\beta \end{bmatrix} \quad (3)$$

2.3 摄像机成像模型的建立

当不考虑非线性因素时,通过针孔成像模型描述其成像几何关系为

$$\frac{x}{X} = \frac{y}{Y} = \frac{z}{f} \quad (4)$$

式中, f 为摄像机的焦距.将式(4)写成矩阵的形式可以得到:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{z} \begin{bmatrix} f & 0 \\ 0 & f \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (5)$$

引入比例因子 s , 并另 $s = z$, 则可进一步得到:

$$s \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (6)$$

将式(1)和式(2)带入式(6)可得摄像机成像模型的表达式如式(7)所示.

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} a_x & 0 & u_0 & 0 \\ 0 & a_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \quad (7)$$

式中, $a_x = 1/dx$, $a_y = 1/dy$ 分别定义为 X 方向和 Y 方向上的等效焦距, 从式(3)和式(7)可知, 在已知内参 a_x 、 a_y 、 u_0 、 v_0 时, 求相对位姿参数 R 和 t 是一个非线性方程。

3 Legendre 正交基神经网络的相对位姿解算

3.1 Legendre 神经网络模型

本文构造的 Legendre 神经网络模型是由输入层、隐层和输出层组成, 其中隐层神经元用 Legendre 正交多项式作为激励函数, 输入层和输出层采用线性函数。以 2 输入 1 输出为例, 其模型结构如图 3 所示。

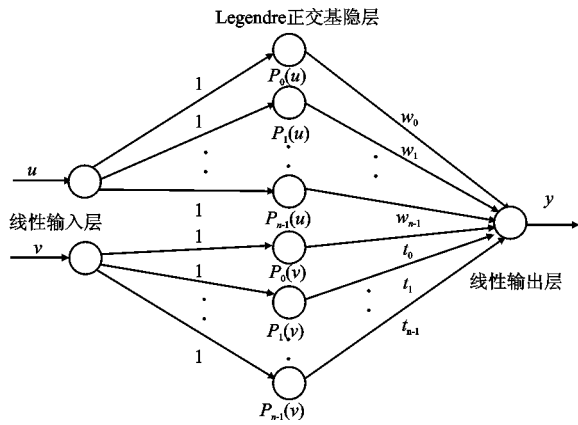


图 3 Legendre 神经网络模型

文中构建的神经网络中所有神经元阈值都设置为 0, 输入层和隐层神经元之间的连接权值设为 1, 而隐层和输出层神经元之间的权值设为 w_i ($i = 0, 1, \dots, n-1$) 和 t_i ($i = 0, 1, \dots, n-1$)。隐层神经元 $P_i(x)$ ($i = 0, 1, \dots, n-1$) 的表达式如式(8)所示^[5-6]。

$$\begin{cases} P_0(x) = 1 \\ P_1(x) = x \\ P_2(x) = (3x^2 - 1)/2 \\ P_3(x) = (5x^3 - 3x)/2 \\ \dots \\ P_n(x) = \frac{n!}{2^n n!} \times \frac{d^n}{dx^n} [(x^2 - 1)^n] \end{cases} \quad (8)$$

式中, $x \in [-1, 1]$ 。因此, 可以得到神经网络输入输出表达式如式(9)所示。

$$y = w_0 P_0(u) + w_1 P_1(u) + \dots + w_{n-1} P_{n-1}(u) + t_0 P_0(v) + t_1 P_1(v) + \dots + t_{n-1} P_{n-1}(v) \quad (9)$$

设 $w = [w_0, w_1, \dots, w_{n-1}]^T$, $t = [t_0, t_1, \dots, t_{n-1}]^T$, $Y = [y_0, y_1, \dots, y_{m-1}]^T$,

$$X_1 = \begin{bmatrix} P_0(u_0) & P_1(u_0) & \dots & P_{n-1}(u_0) \\ P_0(u_1) & P_1(u_1) & \dots & P_{n-1}(u_1) \\ \dots & \dots & \dots & \dots \\ P_0(u_{m-1}) & P_1(u_{m-1}) & \dots & P_{n-1}(u_{m-1}) \end{bmatrix}$$

$$X_2 = \begin{bmatrix} P_0(v_0) & P_1(v_0) & \dots & P_{n-1}(v_0) \\ P_0(v_1) & P_1(v_1) & \dots & P_{n-1}(v_1) \\ \dots & \dots & \dots & \dots \\ P_0(v_{m-1}) & P_1(v_{m-1}) & \dots & P_{n-1}(v_{m-1}) \end{bmatrix}$$

按照式(9)可以得到矩阵形式, 如式(10)所示。

$$Y = [X_1 \ X_2] \begin{bmatrix} w \\ t \end{bmatrix} = X\psi \quad (10)$$

式中, $X = [X_1 \ X_2]$, $\psi = [w \ t]^T$ 。

3.2 权值的确定

神经网络权值的选取直接影响着神经网络的逼近精度、以及训练时间。本文在基于梯度下降法的基础上推导了 Legendre 神经网络最优权值的确定方法。根据图 3, 假设神经网络的样本集 $A = \{X, Y_1\}$, X 为样本集的输入, Y_1 为样本集的输出, Y 为神经网络的输出其表达式如式(10)所示。定义该神经网络的误差函数 E 如式(11)。

$$E = \frac{1}{2} (Y_1 - Y)^2 = \frac{1}{2} (Y_1 - X\psi)^2 \quad (11)$$

权值迭代修正式可写为

$$\psi(k+1) = \psi(k) - \Delta\psi = \psi(k) - \eta(\partial E / \partial \psi) |_{\psi=\psi(k)} \quad (12)$$

式中, η 为学习率并且大于 0, 迭代次数 $k = 0, 1, 2, \dots$ 。将式(11)代入式(12)可得

$$\psi(k+1) = \psi(k) - \eta X^T (X\psi - Y_1) \quad (13)$$

假设当神经网络训练得到了最优权值 ω , 则 ω 应该满足如下式子:

$$\lim_{k \rightarrow \infty} \psi(k) = \lim_{k \rightarrow \infty} \psi(k+1) = \omega \quad (14)$$

式中, ω 为权值矩阵常数, 将式(14)代入式(13), 最终可得最优权值为

$$\omega = (X^T X)^{-1} X^T Y_1 \quad (15)$$

同理, 进一步可以推导出 n 个输入 1 个输出的神经网络最优权值为

$$\omega_n = (X_n^T X_n)^{-1} X_n^T Y_1 \quad (16)$$

式中, $X_n = [X_1 \ X_2 \ \dots \ X_n]$ 。

3.3 Legendre 正交基神经网络的搭建

由于摄像机内部参数可以认为是固定不变的, 而且从 2.3 节的分析可知, 相对位姿参数的求解包含 6 个独立变量(俯仰角、滚转角、方位角以及 X 、 Y 和 Z 方向的偏移量)。所以在图像中利用 3 个像素点(3 对横、纵坐标)即可确定出唯一的相对位姿参

数. 因此, 本文构建了一个 6 输入 6 输出的神经网络, 输入参数为图像坐标系中 3 个点的像素纵横坐标, 输出为相机相对于世界坐标系的 X 轴方向、Y 轴方向、Z 轴方向的偏移量以及俯仰角、滚转角和方位角. 此外, 为了提高神经网络的训练速度以及精度, 进一步将这个神经网络分解成 6 个 6 输入 1 输出的神经网络, 如图 4 所示.

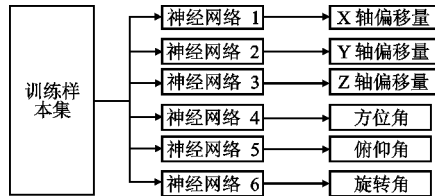


图 4 并行神经网络解算

基于 Legendre 神经网络的位姿解算步骤如下:

- (1) 以棋盘格作为实验对象, 标定好棋盘格中三个点世界坐标系下的坐标, 记为 A、B 和 C 点;
- (2) 通过改变相机与棋盘格的距离以及调整相机与棋盘格的相对姿态, 记录 A、B 和 C 点三点的像素坐标及其对应的 3 个姿态角和 3 个偏移量, 从而建立训练样本集;
- (3) 对样本集进行训练;
- (4) 用测试样本对神经网络进行测试, 验证本文算法的性能.

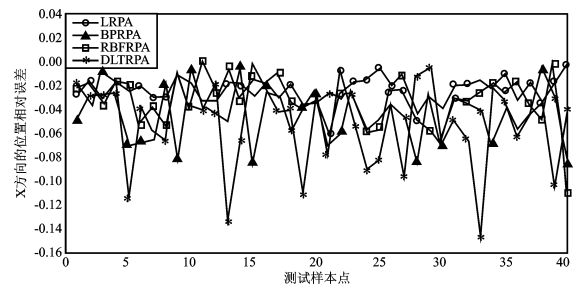
4 实验及分析

4.1 仿真实验

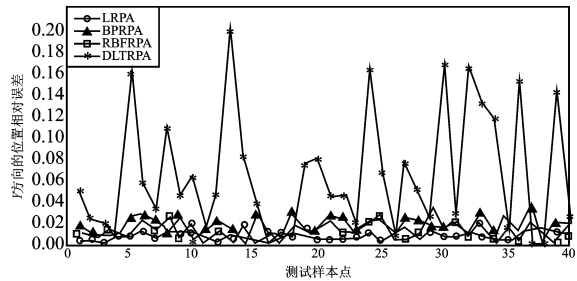
仿真实验时在 T410 的联想笔记本电脑上进行. 仿真实验中, 首先给定摄像机的内参, 如表 1 所示. 然后通过调整摄像机与被测目标的位姿来获取 160 组样本数据 (图像中 3 个像素点的坐标及其对应的 3 个姿态角和 3 个偏移量), 并用其中 40 组数据作为测试样本, 剩下的 120 组数据作为训练样本. 考虑到真实情况中摄像机模型中存在的噪声以及图像像素点检测的误差, 在每个测试样本中加入均值为 0, 方差为 1 的高斯白噪声, 并分别采用 DLTRPA (DLT 相对位姿解算)、BPRPA (BP 相对位姿解算) 以及 RBFRPA (RBF 相对位姿解算) 与 LRPA (Legendre 相对位姿解算) 作对比, 其结果如图 5 所示.

表 1 给定摄像机内部参数

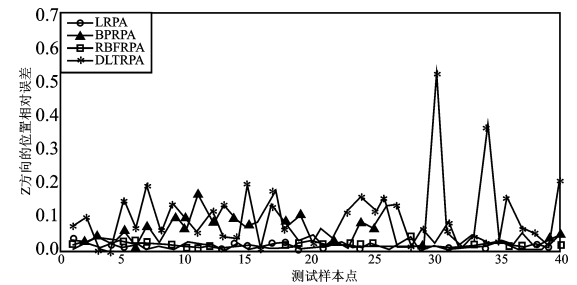
内参	X 方向	Y 方向
光轴与图像平面交点	785	642
等效焦距	100	100



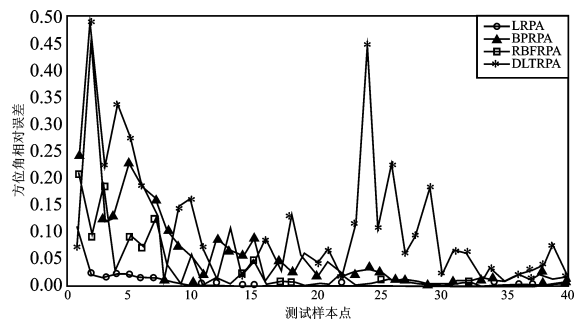
(a) X 方向位置相对误差



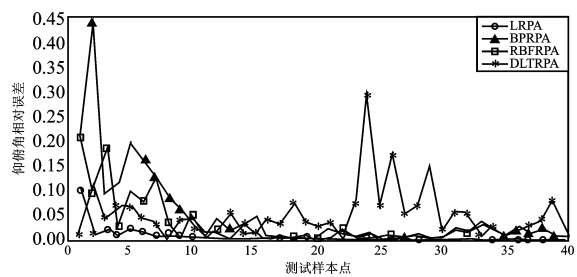
(b) Y 方向位置相对误差



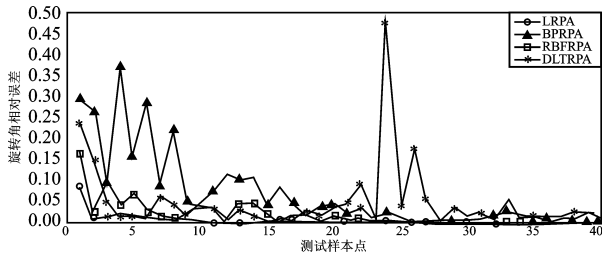
(c) Z 方向位置相对误差



(d) 方位角相对误差



(e) 俯仰角相对误差



(f) 滚转角相对误差

图5 相对位姿误差曲线

从图5可看出, BPRPA、RBFRPA 以及 LRPA 相比于 DLTRPA 算法相对误差更小, 具有更好的抗噪能力, 鲁棒性更强. 其中 LRPA 算法的相对误差最小. 各种方法的平均相对误差和运行时间 t 如表2所示.

表2 平均相对误差和运行时间的比较

算法	平均相对误差						t/s
	X 方向	Y 方向	Z 方向	俯仰角	滚转角	偏航角	
DLT	0.075	0.105	0.187	0.203	0.085	0.147	0.014
BP	0.054	0.020	0.085	0.148	0.077	0.158	0.244
RBF	0.042	0.017	0.032	0.021	0.023	0.040	0.120
Leg	0.017	0.008	0.009	0.003	0.004	0.003	0.002

通过表2的分析可知, BPRPA 算法以及 RBFRPA 算法的平均相对误差虽然小于 DLTRPA 算法, 但实时性都不如 DLTRPA 算法, 而 LRPA 算法不仅平均相对最小, 且实时性最好.

4.2 物理实验

采用的实验设备主要有: 北京博创兴盛科技有限公司生产的 UP-Voyager IIA 自主移动机器人、罗技 C310 摄像头(分辨率为 640×480)、棋盘格标定板以及型号为 T410 的联想笔记本电脑等. 摄像头的内参如表3所示.

表3 摄像头内部参数

内参	X 方向	Y 方向
光轴与图像平面交点	300	233
等效焦距	821	821

由于该移动机器人具有 Z 方向、俯仰角以及方位角 3 个方向的自由度, 所以实验主要验证这 3 个自由度方向的解算精度. 实验中分别改变机器人的 3 个自由度来获取 100 组实验数据, 其中 20 组数据作为测试样本, 80 组用来训练. 对所获取的图像利用 Harris 算法检测棋盘格的角点如图6所示, 并选取了其中 3 个点(A、B 和 C)的纵横坐标作为神经网络的输入. 各种算法的对比实验结果如表4所示.

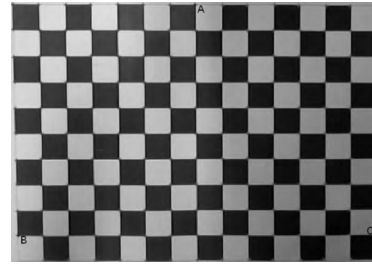


图6 Harris 角点检测

表4 平均相对误差的比较

方法	平均相对误差/%		
	Z 方向	俯仰角	偏航角
DLTRPA	0.380 1	0.419 1	0.342 2
BPRPA	0.268 3	0.211 4	0.172 4
RBFRPA	0.116 4	0.094 2	0.103 3
LRPA	0.034 4	0.026 3	0.027 0

从表4的结果可以看出, LRPA 方法精度较高, 优于其他三种方法.

5 结束语

本文采用一种基于 Legendre 正交基神经网络相对位姿解算方法来解决单目视觉相对位姿解算中的强非线性问题. 实验结果表明, Legendre 神经网络解算算法不仅解算精度高, 而且实时性好.

参考文献:

- [1] 王鹏, 孙长库, 张子森. 单目视觉位姿测量的线性求解[J]. 仪器仪表学报, 2011, 32(5):1126-1131.
- [2] Li G J, Li G R. Camera calibration for monocular vision system based on harris corner extraction and neural network [C]// 2011 International Conference on Consumer Electronics, Communications and Networks, China: Wuhan, 2011:1-4.
- [3] George N R A, Fadi F. Stereo-vision calibration by multi-layer perceptions of an artificial neural network [C]// 2011 7th International Symposium on Image and Signal Processing and Analysis, Shanghai, 2011: 621-626.
- [4] Hazem H Ali, Mohammed T Haweel. Legendre neural networks with multi input multi output system equation [C]// 2012 Seventh International Conference on Computer Engineering & System, China: Tianjin, 2012:92-97.

(下转第 49 页)

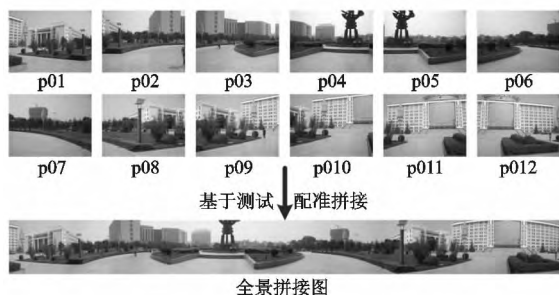


图6 全景拼接仿真效果

6 结束语

针对图像拼接特征点配准计算量大、配准正确率存在误差情况,本文研究了基于标定配准的视频拼接方法.此方法要求视频采集摄像头位置相对固定,通过试验和仿真证明了此方法的有效性、实时性和实用性.通过此方法设计的标定配准图像拼接系统可以用于装甲车辆周边环境监控及无人平台环境监控,并为地面复杂背景视频图像运动目标识别和跟踪提供技术基础.

参考文献:

- [1] Brown L G. A survey of image registration techniques [J]. ACM Computing Surveys, 1992, 24 (4):

325-376.

- [2] 欧阳宁,翟紫伶,首照宇,等.基于图切割的多频带融合图像拼接[J].微电子学与计算机,2013,30(7):107-110.
- [3] 黄琼丹,邱跃洪,田小平.基于特征点及优化理论的图像自动拼接方法[J].光子学报,2009,38(8):2139-2143.
- [4] 尚明妹.基于改进SIFT特征匹配的快速图像拼接算法[J].微电子学与计算机,2014,31(1):64-67.
- [5] 李阳,林意.一种改进的Harris角点图像拼接方法[J].计算机科学,2012,39(11):264-266.
- [6] 陈洪猛,李明,贾璐,等.一种高精度的DBS图像拼接算法[J].西安电子科技大学学报:自然科学版,2014,41(2):46-53.
- [7] 魏川,张功国,吕晓萌.图像拼接技术在双摄像头监控中的应用[J].激光与红外,2014,44(4):447-451.

作者简介:

赵凯 男,(1970-),硕士,研究员.研究方向为信息与控制总体、图像处理.

耿修堂(通讯作者) 男,(1975-),博士,高级工程师.研究方向为信息与控制、智能计算、图像处理. E-mail: gxt1028@163.com.

何浩 男,(1969-),高级工程师.研究方向为无线电通信.

(上接第45页)

- [5] Nithin V George, Ganapati Panda. A reduced complexity adaptive legendre neural network for nonlinear active noise control [C] // 2012 19th International Conference on System, Signals and Image Processing. Vienna, 2012:560-563.
- [6] Zhang Yunong, Li Wei, Yi Chenfu, et al. A Weights-directly-determined simple neural network for nonlin-

ear system identification [C] // 2008 IEEE International Conference on Fuzzy System and Knowledge Discovery. China:Jinan, 2008:455-460.

作者简介:

郝帅 男,(1986-),博士研究生.研究方向为视觉导航、模式识别及图像处理. E-mail: haoxust@163.com.