

An assembly system based on industrial robot with binocular stereo vision

Hong Tang^a and Nanfeng Xiao

School of Computer Science & Engineering, South China University of Technology, 510006, Guangzhou, China

Abstract. This paper proposes an electronic part and component assembly system based on an industrial robot with binocular stereo vision. Firstly, binocular stereo vision with a visual attention mechanism model is used to get quickly the image regions which contain the electronic parts and components. Secondly, a deep neural network is adopted to recognize the features of the electronic parts and components. Thirdly, in order to control the end-effector of the industrial robot to grasp the electronic parts and components, a genetic algorithm (GA) is proposed to compute the transition matrix and the inverse kinematics of the industrial robot (end-effector), which plays a key role in bridging the binocular stereo vision and the industrial robot. Finally, the proposed assembly system is tested in LED component assembly experiments, and the results denote that it has high efficiency and good applicability.

Keywords: assembly system; industrial robot; binocular stereo vision; genetic algorithm; deep neural network.

1 Introduction

With the development of intelligent manufacturing technology, the electronic component assembly production based on the industrial robots becomes increasingly efficient. For assembly production, industrial robots relieve workers of the monotonous, duplication work. But most robots, if changes made for its electronic components position, sometimes even a slight change, the factory have to re-configure the whole assembly line, which seriously affects the economic benefit of the factory. Therefore, we propose an assembly system based on an industrial robot with binocular stereo vision, and tested in the LED component assembly, the results denote that the assembly system has high efficiency and good applicability.

The 3D robot vision system [1] could pick objects that have 3D shapes and randomly organized in a box. However, calculating all of the correspondent image points will cost too much time. To overcome the problem [2] this paper proposed a method to lock the image area containing the parts and components which take the advantage of the space-based attention mechanism model [3]. This method may work efficiently in a certain condition when the components are changed or are taken away from the visual field of the cameras always gets some nonsense results for space-based attention mechanism model. This paper introduces an assembly system based on an industrial robot with binocular stereo vision. The principle of the stereo vision is shown in Figure 1. With the binocular stereo vision, we can obtain the position and the orientation of the electronic components by matching

^a Corresponding author : tang.hong@mail.scut.edu.cn

the correspondent image points on the left camera and the right camera. In addition, the visual attention mechanism model [4] is introduced to the proposed assembly system, which is very helpful to pinpoint various electronic parts and components on the images of the cameras.

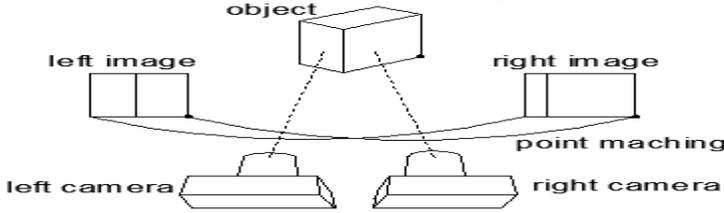


Figure 1. The principle of the binocular stereo vision.

2 Assembly system framework

The binocular stereo vision can use in different assembly tasks to assist the end-effector to grasp the objects. Figure 2 shows the framework of the proposed assembly system. This assembly system includes stereo vision module, coordinate transformation module and industrial robot module. The stereo vision module constructing an ideal binocular stereo vision system and in this module the attention mechanism is applied to obtain a salient area includes the objects. Then, we can extract the object in a complex background, and to compute their positions and orientations. In the coordinate transformation module, the task is to convert the 3D positions and orientations. In this phase, the positions in the camera coordinate system would convert into industrial robot coordinate system. In the industrial robot module, the joint angles of the industrial robot are gotten by the inverse solution with GA, which are used to adjust the grasping positions and orientations of the end-effector to complete the assembly [5, 6].

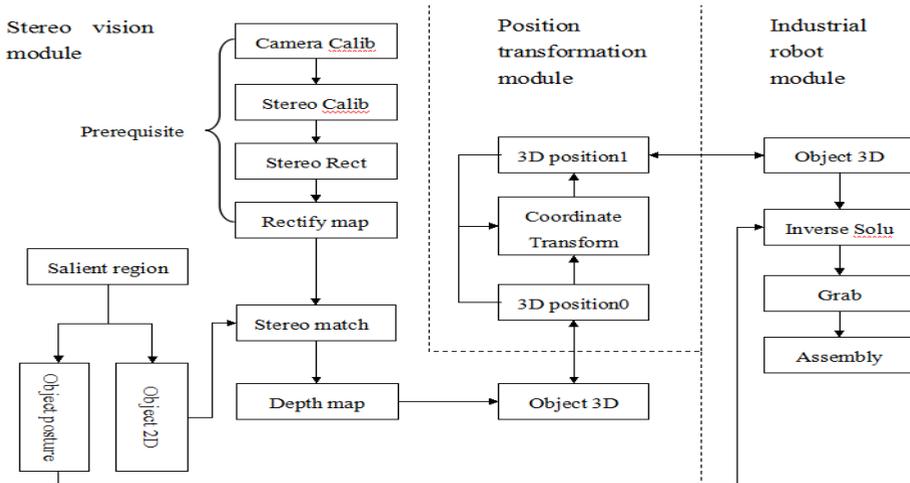


Figure 2. The framework of the assembly system.

3 Binocular stereo vision

The binocular stereo vision includes eliminating distortion, stereo calibration, stereo rectification and stereo matching. The 3D coordinates of the objects will be gotten by overlapping matrix Q and parallax d shown in Figure 3, Z can be calculated as follows by the similar triangle principle that each edge has in the same proportion,

$$\frac{T - (x^l - x^r)}{Z - f} = \frac{T}{Z} \implies Z = \frac{fT}{x^l - x^r} = \frac{fT}{d} \quad (1)$$

where Z is the depth of the point P on the objects, and T is a baseline length between the two cameras, the focal length are all f , x^l and x^r are the coordinates on the left and right images respectively.

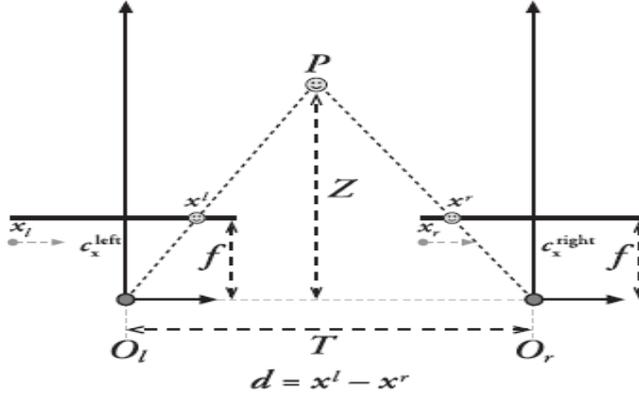


Figure 3. The computation of the 3D coordinates

3.1 Constructing process

In order to implement an ideal binocular stereo vision model shown in Figure 4, the following work needs to be done. Firstly, the chessboard images shown in Figure 5 are utilized to calibrate [7] each camera, in order to obtain their intrinsic matrix \mathbf{M} and distortion vector \mathbf{D} which are defined in following Eq.(2) respectively, and there are nine parameters in removing the internal error and eliminating the image distortions of the cameras.

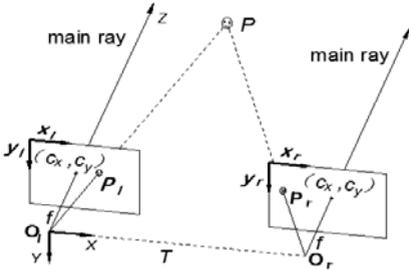


Figure 4. Binocular stereo vision model

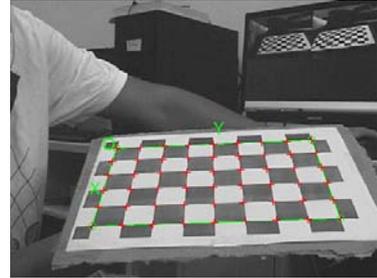


Figure 5. Chessboard grid

$$\mathbf{M} = \begin{bmatrix} f_x & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{D} = [k_1, k_2, p_1, p_2, k_3]^T \quad (2)$$

where, f_x and f_y are respectively the focal length of the cameras. Equally, c_x and c_y are respectively disparities between the main point and the image center, k_1 , k_2 and k_3 are respectively the parameters of the radial distortions, p_1 and p_2 are respectively the parameters of the tangential distortions.

Secondly, the left and right cameras capture the images at the same time which are implemented in the stereo calibration. Here, \mathbf{M} is used to converse the two camera coordinate systems. The parallax d is calculated from the pixel at the same row on the image planes of the left and the right cameras. In the stereo vision model, the block matching (BM) method is introduced for its simple and fast. Therefore, the 3D coordinates of the object can be obtained by

$$Q \begin{bmatrix} x \\ y \\ d \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} \quad \text{here} \quad Q = \begin{bmatrix} 1 & 0 & 0 & -C_x \\ 0 & 1 & 0 & -C_y \\ 0 & 0 & -1/T_x & (C_x - C'_x)T_x \end{bmatrix} \quad (3)$$

where, (x, y) is the 2D coordinates of the point on the object in the image planes, d is a parallax, C_x, C_y are horizontal and vertical pixel coordinates respectively. T_x is a baseline length between the two cameras, and its 3D coordinates of the point are $(X/W, Y/W, Z/W)$.

3.2 Region extraction

From the binocular stereo vision model shown in Figure 4, the 3D coordinates of every matched pixel on the image planes can be gotten but costs too much time, even only parallax to calculate the 3D coordinates of every point on the image planes. Moreover, most of the calculations are not necessary, because only few points on the object are meaningful to pinpoint. Therefore, rapidly extract the image region containing the object before computing parallax is necessary.

As the visual attention mechanism has ability to obtain a salient region and can be used to recognize the objects. So, it was used to pinpoint the objects [8, 9] and the mode is shown in Figure 6. The greatest difference of this model is to bridge the top-down and the bottom-up attention mechanism with 'grouping' such as a point, a feature, an object, a region or a hierarchical structure [10, 11]. In this paper, color, intensity and orientation as well as the object contours are added in grouping. Then, the salient regions are marked out by a minimum external rectangle on the original image. At last, these regions are extracted from the original images whose contour features are input to a deep neural network for recognizing. After the object corresponding for the assembly tasks is found the outermost contour will be extracted for pinpointing the 2D positions and orientations of the object on the image planes. All the process is shown in Figure 7. In Figure 7 the first row is the original images, the second one is a salient map, and the last row is the extraction region.

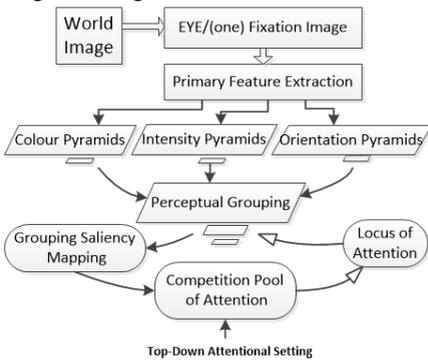


Figure 6. Attention mechanism model

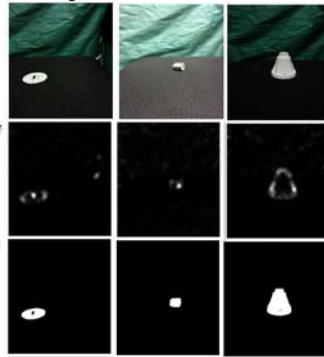


Figure 7. Space segmentation

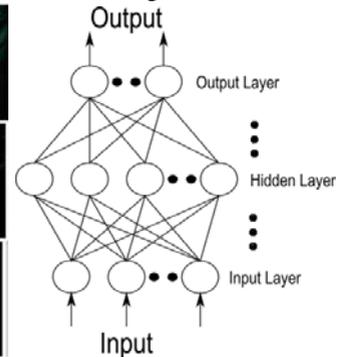


Figure 8. deep neural network

3.3 Object recognition

In this paper, a deep multilayer neural network is used for the object recognition [12-14]. Its learning rule is to apply the steepest descend method to adjust the weights and thresholds of the neural network according to the minimum sum of the square error. The topology of the deep neural network includes an input layer, a hide layer and an output layer shown in Figure 8. Firstly, we obtain the edge features of each object by the canny edge detection algorithm, and then divide the image into a 4×4 grid shown in Figure 9, which gives a 16 dimensional vector. The neuron number of the layer determined by the dimension of the vector, so the input neuron is 16. For there are 4 objects with 2 orientations of each object, therefore, the number of the output neuron is 8.

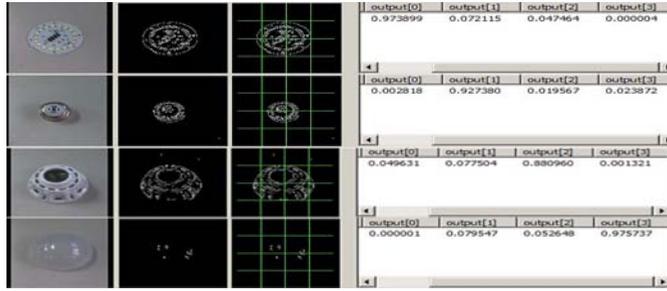


Figure 9. Feature extraction of the objects

Each element of the vector is the ratio of the edge in this region and the feature of grouping for the attention mechanism model. For the assembly task, if only there are enough images for training an available network can be gotten, and the deep neural network can work well in the object recognition. In Figure 9, the column on the left is the recognition results. The deep neural network could work well in confirming whether the object is the assembly task or not if a region is extracted through the attention mechanism.

4 Industrial robots

This paper is based on an industrial robot with six degrees of freedom (6 DOF) shown in Figure 10. It is mainly composed of manipulators, electric control cabinet and s workbench. The D-H parameters of this industrial robot are shown in Table 1.

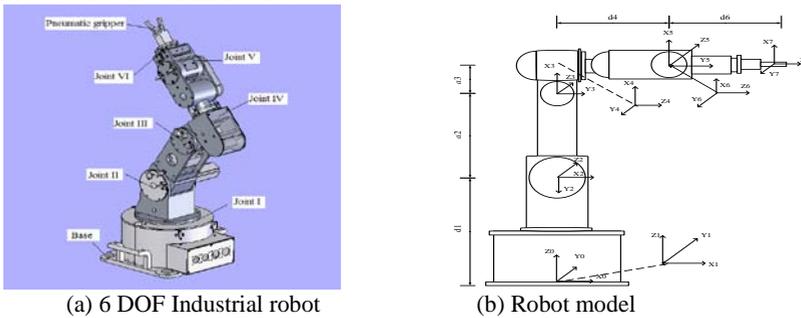


Figure 10. Industrial robot with end-effector

Table 1. D-H parameters of the industrial robot

Joint	α_{i-1}	a_{i-1}	d_i	θ_i
I	0	0	0	θ_1
II	-90°	0	d_1	θ_2
III	0	a_2	0	θ_3
IV	-90°	a_3	0	θ_4
V	90°	0	d_4	θ_5
VI	-90°	0	0	θ_6

According to the D-H parameters of the industrial robot, the kinematics of the industrial robot can be obtained as follows

$$T = {}^0T = {}^0A_1A_2A_3A_4A_5A_6A = \begin{bmatrix} r_{11} & r_{12} & r_{13} & P_x \\ r_{21} & r_{22} & r_{23} & P_y \\ r_{31} & r_{32} & r_{33} & P_z \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} R_{3 \times 3} & P_{3 \times 1} \\ 0_{1 \times 3} & 1 \end{bmatrix} \tag{4}$$

where, the 3×3 matrix $\mathbf{R}_{3 \times 3}$ is the orientation of the end-effector, and its coordinates in the industrial robot system are P_x , P_y respectively. \mathbf{A} is a transformation matrix between the joint coordinate systems. The position of the end-effector can be acquired from the binocular stereo vision.

4.1 Coordinate transformation

In this paper, we adopt an indirect method to calculate the conversion matrix, which can get the matrix inversely through several point coordinates in both coordinate systems [15], the principle is as following. Suppose that there are two coordinate systems \mathbf{A} and \mathbf{B} , and \mathbf{B} comes from \mathbf{A} after rotation \mathbf{R} that around X, Y, Z axis with γ , β and α , respectively, and translation \mathbf{P} that along X, Y, Z axis with P_x , P_y and P_z , respectively. Then the conversion matrix from A to B is

$${}^A_B\mathbf{T} = \begin{bmatrix} cac\beta & cas\beta s\gamma - sac\gamma & cas\beta c\gamma + sas\gamma & P_x \\ sac\beta & sas\beta s\gamma + cac\gamma & sas\beta c\gamma - cas\gamma & P_y \\ -s\beta & c\beta s\gamma & c\beta c\gamma & P_z \\ 0 & 0 & 0 & 1 \end{bmatrix}, \begin{cases} c\alpha = \cos \alpha \\ s\alpha = \sin \alpha \end{cases} \quad (5)$$

Suppose a point P whose coordinate in A and B is $\mathbf{P}_a(x_a, y_a, z_a)$ and $\mathbf{P}_b(x_b, y_b, z_b)$ respectively. Then we have the following equation

$$\begin{bmatrix} P_a \\ 1 \end{bmatrix} = {}^A_B\mathbf{T} \begin{bmatrix} P_b \\ 1 \end{bmatrix} \quad (6)$$

As there are 6 unknown numbers in ${}^A_B\mathbf{T}$ we just need two coordinate pairs to get a conversion matrix. Foresee from Eq.(5), it is too complex to find a solution because of the inverse trigonometric function calculation. To simplify this calculation, we take Eq.(5) into Eq.(4). Though there are still 6 unknown numbers to be calculated the process to find a solution become easier for only linear equation system. The only work needed to be done is to acquire two more coordinate pairs. However, the 12×12 matrices based on Cramer's Rule will be too complicate to be calculated, though it could be divided into 4×4 matrices, there also are fifteen 4×4 determinants to be calculated. This paper applies a novel and convenient method in minimizing calculation. As \mathbf{P}_b is the key to make calculation becoming complicated as it is uncontrollable. Fortunately, \mathbf{P}_a is able to be designed so that Eq.(6) is transformed into Eq.(7)

$$\begin{bmatrix} P_b \\ 1 \end{bmatrix} = ({}^A_B\mathbf{T})^{-1} \begin{bmatrix} P_a \\ 1 \end{bmatrix} \quad (7)$$

Firstly, 4 points in the industrial robot coordinate system are selected for \mathbf{P}_a , which are (0, 0, 0), (1, 0, 0), (0, 1, 0) and (0, 0, 1) respectively. Evidently, after we find the points in the binocular stereo vision, \mathbf{T}^{-1} could be gotten through simple addition and subtraction. As an assembly task is based on a certain criteria, most of its objects have accurate sizes. Therefore, it is not difficult to mark the feature points. An alternative method is to mark the ending of the industrial robot, and it can be moved to the designed points. Besides, the feature points are changeable, if only they are not coplanar and each three points are not collinear, and most importantly, 0 is contained as many as possible in each coordinate.

In general, calculating an inverse matrix is a hard work but \mathbf{T}^{-1} is an exception. For the rotation matrix \mathbf{R} is an orthogonal matrix, whose inverse is equal to its transport matrix. Therefore, the only work is just to calculate three 3×3 determinates, which accelerates 20 times at least contrasting to calculate fifteen 4×4 determinates. What is more, the error level of this method will be close to the error level of the binocular stereo vision theoretically.

4.2 Inverse kinematics solution

Inverse kinematics occupies a significant position in the industrial robotics. Its main work is to converse position and orientation of the end-effector of the industrial robot into the joint angles. In this paper, the genetic algorithm (GA) is applied in solving inverse kinematics. The computation of the inverse kinematics is designed without considering the end-effector orientation and the optimization model is to minimize the distance between the end-effector and the object point which shown as follows:

$$\min \left(\sum_{i=1}^6 \theta_i + \beta D \right) \text{ s.t. } \begin{cases} \theta_1 \in [-150^\circ, 150^\circ], \theta_2 \in [-135^\circ, 45^\circ] \\ \theta_3 \in [-70^\circ, 90^\circ], \theta_4 \in [-90^\circ, 90^\circ] \\ \theta_5 \in [-90^\circ, 90^\circ], \theta_6 \in [-180^\circ, 180^\circ] \end{cases} \quad (8)$$

where, θ_i is the rotation angle of joint i , β is a penalty factor, and D is a hamming distance between the end-effector and the object point. Then we utilize 60 bits to encode each chromosome in the population, namely, every 10 bits represents a gene (a joint angle). The decoding function is defines as

$$\theta_i = \min_i + \left\{ \frac{(B_i - 2^l + 1)}{2^l - 1} + 1 \right\} * (\max_i - \min_i), \quad i=1, 2 \dots 6 \quad (9)$$

where B_i is an angle of joint i , and l is the number of encoding bits, \max_i and \min_i are the joint rotation limitation of the highest and the lowest respectively, as shown in Eq.(8). We set the fitness function as

$$F = \frac{1}{1 + (vD + (1-v)\varphi)}, \text{ here } v = \frac{1}{2} * e^{-D/10} + \frac{1}{2} \quad (10)$$

where φ is the sum of the 6 joint rotation angles, D is a hamming distance, v is the impact factor of D and φ . The roulette wheel selection is applied in choosing breeding chromosomes. Then a two layer mutation strategy is taken in this model. The first layer uses the global search which makes mutation affecting every single one. The second layer uses local search which makes each gene code of the lower five bits to mutate from low to high bit by bit so that we can continue to search for a chromosome with better fitness in their neighborhood.

5 Experiments and analysis

Our experiments are based on the RBT-6T/S01S industrial robots with 6 DOFs, for this study the binocular stereo vision system is the two CCD cameras of TOSHIBA TELI.

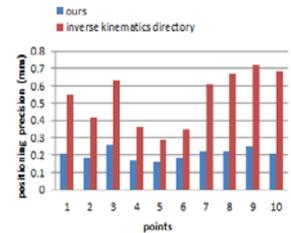
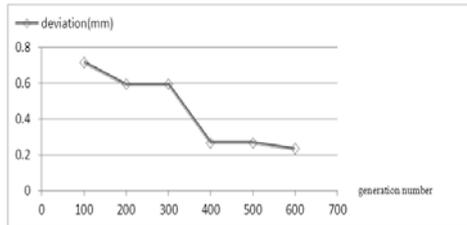


Figure 11. Assembly system. Figure 12. Distance deviation.

Figure 13. positioning precision.

Figure 11 shows the experimental platform for grasping operations. Table 2, 10 records of all grasping test data was listed and each record shows the mapping relationship between the pixel coordinate of the left and right camera images and Cartesian coordinates. Column “left camera” and “right camera” are the pixel coordinates in the left image and right images respectively. The column

“Coordinate Calculation” is the 3D coordinate calculated in space. Since GA is a heuristic method, its genetic operators make the result to ensure randomness. Therefore, choose the best result from several tests is desirable. We have elaborated the derivation of the inverse kinematics above, to get the distances between the object and the end-effector the crossover and mutation rate is 0.1 and 0.09 respectively. With the generation number increasing, the distance between the object and the end-effector is declining illustrated in figure 12, as well as deviation of the joint rotation.

Table 2: experimental data

<i>object NO</i>	<i>Left camera</i>	<i>Right camera</i>	<i>Coordinate calculation</i>	<i>Distance(mm)</i>
1	245,149	155,149	11.06, 49.76, 31.78	167.86
2	296,158	209,158	-53.08, 41.08, 31.15	162.73
3	255,106	156,106	-1.65, 91.19, 39.75	195.14
4	336,101	237,101	-95.02, 95.63, 41.40	188.21
5	199,126	96,126	66.43, 82.31, 39.61	219.37
6	287,104	195,104	-52.51, 83.86, 40.24	227.28
7	248,147	154,147	11.08, 49.75, 10.71	239.50
8	352,154	264,154	-123.21, 42.16, 10.79	224.02
9	202,210	106,210	69.33, -29.24, 20.79	266.60
10	327,207	241,207	-93.86, -32.16, 20.77	253.91
...

Compare the method of inverse kinematics directory the absolute errors of our method are better, result is showed in Figure 13.

6 Conclusions

This paper presented an assembly system based on an industrial robot with binocular stereo vision. And the visual attention mechanism model is applied to accelerate the speed and efficiency of recognition and pinpoint the objects. More, the GA is used to solve the inverse kinematics and compute the joint rotation of the industrial robot. The way to calculate the transition matrix is also a precise and convenient method to bridge the camera coordinates system and the industrial robot coordinate system. The experiment shows the result of our approach is accurate enough to capture the objects on the worktable. And such an assembly system would improve the availability of industrial robots for various intelligent manufacturing.

Acknowledgements

This research is funded by the National Natural Science Foundation (Project No. 61171141, 61573145), the Public Research and Capacity Building of Guangdong Province (Project No. 2014B010104001) and the Basic and Applied Basic Research of Guangdong Province (Project No. 2015A030308018), the authors are greatly thanks to these grants.

References

1. Tudorie, C. Raluca, *Simulation, Modeling, and Programming for Autonomous Robots* **509** (Springer, 2010)
2. L. Yang, NF. Xiao, *AMM* **708** 385(2013)
3. L. Itti, C. Koch, E. Niebur, *TPAMI* **1254**, 11(1998)
4. Y. Sun, R. Fisher, *Artificial Intelligence* **77**, 146(2003)
5. Aguilar, J.C. Huegel, *Advances in Artificial Intelligence* **490** (Springer,2011)
6. Z. Sui, L. Jiang, Y.T. Tian, W. Jiang, *PCIAC2015* **151**(Springer,2015)
7. Z. Zhang, *Pattern Analysis and Machine Intelligence* **1330** 22(2000)

8. L. Lukic, A. Billard, J. Santos, *Autonomous Mental Development* **76** 7(2015)
9. J.J. Craig, *Introduction to robots: Mechanics and Control* 3(PPHUSR, 2005)
10. Q.J. Zhao, D.B. Zhao, W. Hu, *JUESTC* **016** 4(2010)
11. H. Badino, A. Yamamoto, T. Kanade, *International Conference on Computer Vision* 222 (IEEE, 2013)
12. G.E. Hinton, R.R. Salakhutdinov, *Science* **504** 313(2006)
13. P. Swietojanski, S. Renals, *SLT* **171** (2014)
14. J.C. Ban, C.H. Chang, *Neural Networks* **116** 46 (2013)
15. N.F. X, *Intelligent robots* **104** (South China University of Technology, 2008)