Particle Filtering for Position based 6DOF Visual Servoing in Industrial Environments

Aitor Ibarguren, José María Martínez-Otzeta and Iñaki Maurtua
Fundación Tekniker, Apdo. 44 Otaola 20, 20600 Eibar, Gipuzkoa, Spain

Keywords: Robotics, Visual Servoing, Particle Filter.

Abstract: Visual servoing allows the introduction of robotic manipulation in dynamic and uncontrolled environments. This paper presents a position-based visual servoing algorithm using particle filtering. The objective is the grasping of objects using the 6 degrees of freedom of the robot manipulator (position and orientation) in non-automated industrial environments using monocular vision. A particle filter has been added to the position-based visual servoing algorithm to deal with the different noise sources of those industrial environments (metallic nature of the objects, dirt or illumination problems). This addition allows dealing with those uncertainties and being able to recover from errors in the grasping process. Experiments performed in the real industrial scenario of ROBOFOOT project showed accurate grasping and high level of stability in the visual servoing process.

1 INTRODUCTION

Traditional industrial robotic applications, like part placement or spot welding, require precise information about the position of the objects to perform their task. Visual servoing (Weiss et al., 1987; Hutchinson et al., 1996) can enhance those industrial applications allowing corrections on the robot trajectories.

Even so, industrial environments raise their own challenges in the inclusion of visual servoing techniques, especially when the production line is not completely automated. Dirt, imprecision in the workpiece placement or changing lighting conditions are some of the problems that must be tackled in this kind of environments, introducing uncertainties in the trajectory correction process.

This paper presents a position-based visual servoing algorithm using particle filtering. Based on the real industrial scenario of ROBOFOOT project, the paper proposes an algorithm to grasp a workpiece (a shoe last specifically) from a not constrained workshop, correcting the 6 degrees of freedom of the robot during the visual servoing process.

The paper is organized as follows. In Section 2 the related work is presented. Section 3 exposes briefly particle filters. Task specification and configuration is shown in Section 4. Section 5 is devoted to the proposed approach, while in Section 6 the experimental results are shown. Finally, Section 7 presents the conclusions as well as the future work to be done.

2 RELATED WORK

Several approaches tackle the use of visual servoing in industrial environments, posing different industrial scenarios and approaches.

Sung-Hyun et al. (Han et al., 1999) propose an image-based visual servoing based on stereo vision. The use of stereo vision allows guiding the robot manipulator to the desired location without giving such prior knowledge about the relative distance to the desired location or the model of the object.

Nomura et al. (Nomura and Naito, 2000) describe a visual servoing system able to track and grasp industrial parts moving on a conveyor using a 6DOF robot arm. A hybrid Kalman Filter is also incorporated to track a moving object stably against visual data noise. Experiments are also presented, performing both 3DOF and 6DOF visual servoing.

Finally, Lippiello et al. (Lippiello et al., 2007) presented visual servoing applications on Industrial Robotic cells. On their setup, composed of two industrial robot manipulators equipped with pneumatic grippers, vision systems and a belt conveyor, a position-based visual servoing is proposed. The system also uses Extended Kalman Filters (EKF) (Julier and Uhlmann, 2004) to manage the occlusions during the multi-arm manipulation.

Ibarguren A., Martínez-Otzeta J. and Maurtua I..
Particle Filtering for Position based 6DOF Visual Servoing in Industrial Environments.
DOI: 10.5220/0003965501610166
In *Proceedings of the 9th International Conference on Informatics in Control, Automation and Robotics* (ICINCO-2012), pages 161-166
ISBN: 978-989-8565-22-8

3 PARTICLE FILTER

Particle filters (A. Doucet and Gordon, 2001; Kotecha and Djuric, 2003), also known as sequential Monte Carlo methods (SMC), are sequential estimation techniques that allow estimating unknown states x_t from a collection of observations $z_{1:t} = \{z_1, ..., z_t\}$. The state-space model is usually described by state transition and measurement equations

$$x_t = f_t(x_{t-1}, v_{t-1}) \quad (1)$$

$$z_t = g_t(x_t, u_t) \quad (2)$$

where f and g are the state evolution and observation model functions respectively and v_t and u_t denote the process and observation noise respectively.

Based on the previous equations, particle filters allow approximating the posterior density (PDF) by means of a set of particles $\{x_t^{(i)}\}_{i=1,...,n}$ using equation

$$p(x_t|z_{1:t}) = \sum_{i=1}^{N} \omega_t^{(i)} \delta(x_t - x_t^{(i)}) \quad (3)$$

where each particle $x_t^{(i)}$ has an importance weight $\omega_t^{(i)}$ associated and δ is the Kronecker delta. These weights are computed following equation

$$\omega_t^{(i)} = \omega_{t-1}^{(i)} \frac{p(z_t|x_t^{(i)}) p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{0:t-1}^{(i)}, z_{0:t})} \quad (4)$$

where $p(z_t|x_t^{(i)})$ is the likehood function of the measurements z_t and $q(x_t^{(i)}|x_{0:t-1}^{(i)}, z_{0:t})$ is the proposal density function.

Based on the previously presented equations the particle set evolves along time, changing the weights of the particles and resampling them in terms of the observations.

4 TASK SPECIFICATION AND CONFIGURATION

Based on the needs of ROBOFOOT project, an object grasping task has been designed using real specifications of footwear workshops. The grasping scenario has been specified as:

- Lasts, material with the shape of a foot used to build shoes, are the object to be grasped. An iron piece (*grasping device*) has been added to lasts to allow a precise and stiff grasping, see Fig. 1, as well as to protect the leather during the grasping process. Those *grasping devices* will be the objects to be identified during the visual servoing process.

Figure 1: Lasts with the *grasping device* on the trolley.

- Lasts are carried in specific trolleys mounted on a manovia. The trolleys are designed to allow the placement of lasts of different shapes and sizes. Lasts are placed in the trolley by human operators. Due to those previous facts it is not possible to know the pose of the last in the trolley, as seen in Fig. 1.
- A 6DOF robot arm with a gripper and a camera and lighting system mounted on the end-effector with an eye-in-hand configuration.
- Based on the design of the gripper and the *grasping device*, the grasping process requires a precision of around a millimeter and 1-2 degrees on each axis to grasp the last smoothly. In the same way the maneuver should take no more than 5-6 seconds.

Based on this scenario, the initial set-up of the system has raised some problems related with the pose estimation of the *grasping device*:

- Illumination is a key aspect in a vision system. In this industrial scenario is complicated to place a suitable external illumination, that is why it was decided to put a specific lighting system on the gripper. Even so, the metallic nature of the *grasping device* makes it difficult to get a good image due to the brightness, reflection and the impossibility of lighting all the image properly.
- Some of the tasks to be performed by both the human operators and robots involve the use of ink, wax or generate dust (roughing process). This dirt can be adhered to the *grasping device*, complicating the visual servoing process.

Those previous points will make it difficult to acquire clear images of the *grasping device*, adding uncertainties to the 6DOF pose estimation that will be the base of the visual servoing process.

5 PROPOSED APPROACH

Taking into account the needed precision, which implies a high resolution camera, and the demanding image processing due to the unstable conditions of the image, a dynamic look-and-move approach has been adopted. A particle filter has also been added to the system to manage the uncertainties of the vision system.

Next lines will describe the general structure of the system, as well as the vision module, pose estimation, the particle filter and the grasping algorithm.

5.1 System Modeling and Architecture

In the described scenario, the space can be represented by $P \in \Re^6$, a set of three positions and three orientations, where $P = [x, y, z, \alpha, \beta, \gamma]^T$. In the same way, this scenario will be composed of two different frames, the robot frame r and the camera frame c. Given those two frames, the homogeneous transformation matrix, denoted by ${}^rT_c$, transforms poses from frame c to frame r as:

$$P^r = {}^rT_c P^c \tag{5}$$

The error of the positioning task involved in the grasping process is represented by vector $E \in \Re^6$ which represents the difference between the pose of the object P^r_o in the robot frame and the pose of the end-effector P^r_e in the robot frame (6). The grasping process can be seen as a minimization of this error that will be fulfilled when $|E| = 0$.

$$E = P^r_e - P^r_o = \begin{bmatrix} x^r_e - x^r_o \\ y^r_e - y^r_o \\ z^r_e - z^r_o \\ \alpha^r_e - \alpha^r_o \\ \beta^r_e - \beta^r_o \\ \gamma^r_e - \gamma^r_o \end{bmatrix} \tag{6}$$

For pose estimation, position-based visual servoing systems extract features from the acquired images and estimate the pose of the object P^r_o and perform the corrections. Even so, the described scenario introduces uncertainties in the feature extraction step (illumination, metallic workpiece...), introducing errors in the pose estimation. To deal with this problem, the use of a particle filter is proposed. From each image, a set of n feature vectors $F_i = \{f_1, f_2, ..., f_m\}_{i=1...n}$ will be extracted for the pose estimation, each of them related with a specific image analysis procedure. Each of those n vectors will be a hypothesis of the values of the m features used for the pose estimation, as it will not be possible to have a unique feature vector extracted from each image due to the uncertainties in the image.

From each feature vector F_i, $P^c_{o_i}$ and $P^r_{o_i}$ will be calculated,

$$P^c_{o_i} = \mathrm{PE}(F_i) \tag{7}$$

$$P^r_{o_i} = {}^rT_c P^c_{o_i} \tag{8}$$

where $P^c_{o_i}$ is the i-th hypothesis of the pose of the object in the camera frame, $P^r_{o_i}$ is the i-th hypothesis of the pose of the object in the robot frame and pose estimation function PE is the function that relates a set of features with a pose of the object in the camera frame.

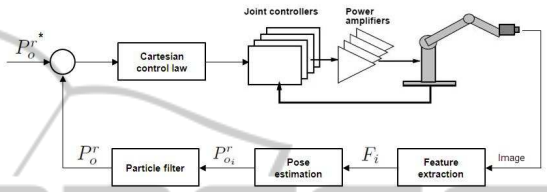

Figure 2: Dynamic position based look-and-move structure with particle filter.

Those n poses, $P^r_{o_{i=1..n}}$ will be the observations of the particle filter, which will output the final pose estimation of the object in the robot frame P^r_o. This final pose will be used to calculate the error E between the object and the end-effector, used to calculate the next robot movement. Fig. 2 shows the structure of the proposed Visual Servoing system.

Next lines will describe the feature extraction, pose estimation, particle filtering and grasping algorithm of the grasping process.

5.2 Feature Extraction

As stated before, one of the challenges of the presented scenario is the feature extraction for pose estimation. The metallic nature of the *grasping device* and the illumination problems make it difficult to detect the different features (edges, corners, holes) precisely. Taking also into account the perspective of the camera through the grasping process, the image features used for pose estimation, shown in Fig. 3, are:

- The center of the three holes (1, 2, 3) of the *grasping device*. Only the pixels of the center of the holes are included, excluding the size and dimensions of the holes, due to the difficulties of extracting their contour precisely.
- The inclination of the left edge (4) of the *grasping device*.

To detect those image features different thresholds, edge detection algorithms and filters are used. Even so, in some images it is not possible to determine the exact position of the three holes' centres as

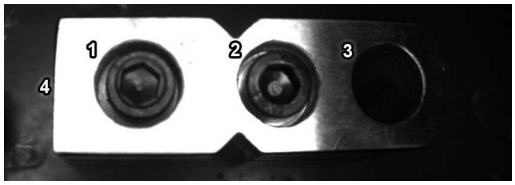

Figure 3: Visual features for pose estimation.

there are various possible circular shapes in each position (ex. the inner screw, outer circle and dirt around it). In those cases it is not possible to define a universal rule to determine which the real contour of the holes is. To overcome this problem, this approach proposes to use all those possible centers of the three holes (left, central and right), creating a set of hypothesis that will be used for pose estimation.

Once the centers of the holes and the left edge are detected, a feature vector will be calculated for each hypothesis as:

$$F_i = \{c_2, d_{12}, d_{13}, d_{23}, \phi_{12}, \phi_{13}, \phi_{23}, \phi_{edge}, \lambda\} \quad (9)$$

where c_2 is the coordinate in pixels of the central hole, d_{ij} is the distance in pixels between the holes i and j, ϕ_{ij} is the angle between the holes i and j, ϕ_{edge} is the angle of the left side of the *grasping device* and λ is a coefficient that measures the noise (quality) of the hypothesis based on the similitude of the circular shapes and their alignment and calculated as

$$\lambda = \frac{C_v(p_1, p_2, p_3) + \frac{|\phi_{12} - \phi_{13}| + 1}{|\phi_{13}| + 1}}{\mathrm{Min}(r_1^{xy}, r_2^{xy}, r_3^{xy})} \quad (10)$$

where p_i is the perimeter of the ith hole, C_v is the *coefficient of variation* of the perimeters and r_i^{xy} is the xy axis ratio of the bounding box of the ith hole.

Those are the features that will be used to estimate the pose of the workpiece.

5.3 Pose Estimation

Once the image is analyzed and the features are extracted, the pose of the object in the camera frame for each of the possible hypothesis are calculated as

$$P_{o_i}^c = [x_i, y_i, z_i, \alpha_i, \beta_i, \gamma_i]^T = \mathrm{PE}(F_i) \quad (11)$$

where each position and orientation is a quadratic function based on some of the features. The coefficients of the quadratic functions are omitted from the paper as they are related to the size of the *grasping device* and the aberration of the lens.

$$\mathrm{PE}(F_i) = \begin{cases} x_i = f(c_2, d_{12}, d_{13}, d_{23}) \\ y_i = g(c_2, d_{12}, d_{13}, d_{23}) \\ z_i = h(d_{12}, d_{13}, d_{23}) \\ \alpha_i = j(\phi_{12}, \phi_{13}, \phi_{23}) \\ \beta_i = k(d_{12}, d_{23}) \\ \gamma_i = l(\phi_{edge}) \end{cases} \quad (12)$$

Once the hypothetical poses of the object in the camera frame $P_{o_i}^c$ are estimated, the poses in the robot frame $P_{o_i}^r$ are calculated using the homogeneous transformation matrix ${}^rT_c$. Those hypothesis will be the observations of the particle filter.

5.4 Particle Filter

Once the possible hypothesis are calculated it is necessary to merge and fuse this information to perform the grasping process. To this end a particle filter is proposed, as it fits in this kind of non-gaussian problem.

Focusing on the posed problem, the state in time t will be defined as a pose of the object in the robot frame

$$X_t = [x_t, y_t, z_t, \alpha_t, \beta_t, \gamma_t]^T \quad (13)$$

As it is not possible to model the pose estimation error a priori, the state transition is defined as

$$X_t = X_{t-1} + V_{t-1} \quad (14)$$

where X_{t-1} is the previous state vector and V_{t-1} is the process noise.

The observation, on the other hand, is defined by a set of hypothetical poses of the object in the robot frame

$$Z_t = P_{o_{i=1..n}}^r \quad (15)$$

So based on this information source each particle will be defined by a probability $P(X_t|Z_t)$.

5.4.1 Probability

To calculate the probability of a state given an observation, initially the distance between the poses is calculated as

$$distPos_i = \sqrt{(x_t - x_i)^2 + (y_t - y_i)^2 + (z_t - z_i)^2} \quad (16)$$

$$distAng_i = \sqrt{(\alpha_t - \alpha_i)^2 + (\beta_t - \beta_i)^2 + (\gamma_t - \gamma_i)^2} \quad (17)$$

where $distPos_i$ is the Euclidean distance between x, y and z positions of the state t and the ith hypothesis and $distAng_i$ is the Euclidean distance between the α, β and γ orientations of the state t and the ith hypothesis.

$$P(X_t|Z_t) = \prod_{i=1}^{n} e^{-distPos_i \cdot distAng_i \cdot (1+\lambda_i)} \quad (18)$$

Based on this distance, the probability of the state is calculated as the product of the exponential of the distances of all the hypothesis ponderated by the λ_i noise coefficient.

5.4.2 Particle Filtering Procedure

Finally the procedure of the particle filter is given as:

- 1: Find the *grasping device* in the initial image and initialise N particles $X_0^{(i)}$ with the different hypothesis randomly, where $w_0^{(i)} = 1/N$
- 2: if $ESS < threshold$ (*Effective Sample Size*), draw N samples with *selection with replacement*
- 3: Predict $x_t^{(i)} = x_{t-1}^{(i)} + v_{t-1}$
- 4: Replace the particles with the lowest weight by the new hypothesis found in the image
- 5: Update importance weights $w_t^{(i)} = w_{t-1}^{(i)} P(X_t|Z_t)$
- 6: Normalize weights $w_t^{(i')} = w_t^{(i)} / \sum_{j=1}^{N} w_t^{(j)}$
- 7: Set $t = t+1$, goto Step 2

In this procedure EES (Liu et al., 2000) (Effective Sample Size) is calculated as

$$cv_t^2 = \frac{var(w_t^{(i)})}{E^2(w_t^{(i)})} = \frac{1}{N}\sum_{i=1}^{N}(Nw_t^{(i)} - 1)^2 \quad (19)$$

$$ESS_t = \frac{N}{1+cv_t^2} \quad (20)$$

where N is the number of particles and $w_t^{(i)}$ is the weight of particle i in time t. The fourth step has been added to allow a fast convergence.

Based on this discrete approximation of the posterior probability, the object is tracked along the grasping procedure.

5.5 Grasping Algorithm

The feature extraction step has shown that the best images are acquired when camera is perpendicular to the *grasping device* and the end-effector to a 30mm distance, $E = [0,0,30,0,0,0]^T$, as it solves in part the illumination problems. Taking it into account the grasping algorithm will try to minimize the error until this value is reached, adding an small tolerance of ± 1 mm in position and $\pm 1.5^\circ$ in orientation to avoid an infinite loop. Once this error is reached the robot will make a final approach in just one axis and perform the grasping.

6 EXPERIMENTAL RESULTS

To test the performance of the proposed approach an experiment has been designed in order to measure its suitability. Those are the specifications of the experiment:

- Six different particle filter configurations have been set-up, mixing different state estimation methods and number of particles. Specifically the estate estimation methods are:
  - Best particle (the one with maximum weight)
  - Robust mean with the 3 particles with maximum weight (denoted as R.M. 3)
  - Robust mean with the 5 particles with maximum weight (denoted as R.M. 5)
- For each configuration, 250 repetitions have been performed using different shoes and *grasping devices*. Most of the *grasping devices* have been dirtied up to include variety and simulate real conditions. Due to the structure of the manovia, the *grasping device* will be placed in a space of 200x100x200mm (depending on the shoe and its placement) and with a rotation of $\pm 15^\circ$ in each axis.
- The grasping process will fail if does not achieve to pick up the shoe. There are two reasons for this fail, the *grasping device* has not been found in the initial image (ex. not well illuminated due to its orientation) or a wrong pose estimation which leads to movement that leaves the *grasping devices* out of the scope of the camera.

TABLE 1 shows the results of the experiment. The first column describes the number of particles and the estimation method, the second one the success rate, third and fourth columns show the mean (μ) and the standard deviation (σ) of the grasping time in seconds, fifth and sixth columns the mean (μ) and the standard deviation (σ) of the number of movements required to grasp the shoe and finally the last column shows the time required to process each cycle of the particle filter in milliseconds.

Table 1: Results of the experiment.

		Time (s)		Mov.		
	%	μ	σ	μ	σ	ms/image
50 - Best	96.4	4.84	1.79	10.39	4.45	127.53
100 - Best	95.6	4.94	2.03	10.41	5.34	133.50
50 - R.M. 3	97.2	4.96	1.85	10.36	4.99	133.79
100 - R.M. 3	98	4.98	1.71	10.27	4.68	136.71
50 - R.M. 5	**99.6**	4.78	1.94	10.25	5.05	**123.03**
100 - R.M. 5	99.2	**4.69**	**1.65**	**9.89**	**4.14**	127.94

Results show a better performance of the system using the *robust mean* estimation method with 5 particles, both in success rate and in grasping time. In the case of the success rate, in all the configurations a part of the fails were related with the search of the *grasping device* in the initial image (more or less the same quantity for each configuration, around a 1-2%).

In the same way, it seems that the addition of more particles does not help to improve the success rates although it does not increase significantly the processing time of each visual servoing iteration.

7 CONCLUSIONS AND FUTURE WORK

This paper presents a dynamic position-based look-and-move architecture to perform visual servoing with 6DOF in industrial environments. This kind of environments usually suffers from unstable conditions like changing lighting condition or dirt, introducing uncertainties in the visual servoing process. To overcome the above mentioned problem, this paper proposes the use of a particle filter to manage multiple hypothesis of the poses of the workpiece to grasp.

The results show a high success rate of the grasping system, reaching around a 99% of success in the different experiments performed. The use of particle filtering allows the use and fuse various hypothesis, overcoming the noise problems of the presented scenario. The system also performs the grasping process in a suitable time, increasing few processing time with the addition of the particle filter.

As further work, there are two interesting paths to follow. On one hand, test this approach in similar scenarios (different workpiece, environment, noise source...) to test its suitability. On the other hand, one or more sensors could be attached to the end-effector as new data sources, using the particle filter to fuse the information received from the different sources as done in different robotic applications.

ACKNOWLEDGEMENTS

This work has been performed within the scope of the project "ROBOFOOT: Smart robotics for high added value footwear industry ". ROBOFOOT is a Small or Medium-scale focused research project supported by the European Commission in the 7th Framework Programme (260159). For further information see http://www.robofoot.eu

REFERENCES

A. Doucet, N. De Freitas, N. and Gordon, N. (2001). *Sequential Monte Carlo methods in practice*. Springer-Verlag.

Han, S.-H., Seo, W., Yoon, K., and Lee, M.-H. (1999). Real-time control of an industrial robot using image-based visual servoing. In *Intelligent Robots and Systems, 1999. IROS '99. Proceedings. 1999 IEEE/RSJ International Conference on*, volume 3, pages 1762 –1767 vol.3.

Hutchinson, S., Hager, G., and Corke, P. (1996). A tutorial on visual servo control. *Robotics and Automation, IEEE Transactions on*, 12(5):651 –670.

Julier, S. and Uhlmann, J. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401 – 422.

Kotecha, J. and Djuric, P. (2003). Gaussian particle filtering. *Signal Processing, IEEE Transactions on*, 51(10):2592 – 2601.

Lippiello, V., Siciliano, B., and Villani, L. (2007). Position-based visual servoing in industrial multirobot cells using a hybrid camera configuration. *Robotics, IEEE Transactions on*, 23(1):73 –86.

Liu, J., Chen, R., and Logvinenko, T. (2000). A theoretical framework for sequential importance sampling and resampling. *Sequential Monte Carlo Methods in Practice*, pages 1–24.

Nomura, H. and Naito, T. (2000). Integrated visual servoing system to grasp industrial parts moving on conveyer by controlling 6dof arm. In *Systems, Man, and Cybernetics, 2000 IEEE International Conference on*, volume 3, pages 1768 –1775 vol.3.

Weiss, L., Sanderson, A. C., and Neuman, C. P. (1987). Dynamic sensor-based control of robots with visual feedback. *IEEE Journal on Robotics and Automation*, RA-3(5).