Supervisory Control for Automatically 3D Pose Estimation

Xiaoli Liu, Xiaoyan Wang

School of Science, Xi'an Shiyou University, Xi'an, China

E-mail: liuxl166@163.com

Keywords: Supervisory Control; virtual vision servo; augmented reality; Automatically Registration

Abstract: Automatically obtaining the pose of 3D object is one of the difficulties in computer vision. In this paper, feedback control theory is used to solve these problems. It is mainly a dynamic target cognitive mechanism with a supervision controller and virtual vision servo controller. Supervisory controller finds the potential matching feature sets according to the feature performance evaluation index, and virtual vision controller gets the possible pose and the integral error under matching hypothesis. The supervisory controller switches in the candidate matches until the integral error is less then specific threshold, then the corrected matchers are found. The augmented reality experiment shows the effectiveness and robustness of the algorithm at the end.

1. Introduction

Visual servo control[1] refers to the use of computer vision data to control the motion of a robot. Visual servo control relies on techniques from image processing, computer vision, and control theory. There are two very different approaches because of the desired features, generally, one is the position-based visual servo control (PBVS)[2], in which desired features consist of a set of 3-D parameters, which must be estimated from image measurements in deed; the other is the image-based visual servo control (IBVS), in which desired features consist of a set of features that are immediately available in the image data.

The virtual visual servo(VVS)[3], an eye-in-hand method based on IBVS, deals with the registration problem in augmented reality, in this algorithm the controlled object is not a physical one but a virtual camera. The image errors are eliminated by change the position of virtual camera. While in [4], firstly assumed the world coordinate is concentric with the camera coordinate , the controlled object is a virtual model (checkerboard as an example) which consists of 3D data at a proper initial pose; secondly the algorithm makes the pose of the virtual model move constantly to pose 1,pose2…, reduces the error between the projection image and the real image; eventually when the image error is minimum, the virtual model and real object get overlapped in three-dimensional space, namely realizes pose estimation of real object. One of the differences between traditional VVS[3] and ours is that the control object is the virtual model not the virtual camera; it is easier and more accurate to get the 3D position estimation by our VVS. These methods can only guarantee for the local asymptotic or local convergence[5].

The VVS methods need to compare the detected image features information and the ones projecting from virtual model in 3D frame to get the 3D pose estimation. It is in essence to the problem of PNP[7]. PNP problem can be divided into two categories: the iteration method and the no-iteration one. The iteration method[8] can get more accuracy solution by the minimum iteration strategy based on nonlinear cost equation. Its drawbacks are too depending on the initial assumption and high complex computation[9] put forward no-iteration algorithm with O(n), and has high efficiency. However it is not stable in noise environment, especially when n is less than or equal to 5.What's more, dealing the stability and accuracy of the camera pose estimation problem under quasi-singular case are lack of effective theoretical analysis and technical method. P3P issue[10] is the minimum subset of PNP. Estimating the object's real pose with three feature points is a typical

  DOI: 10.25236/icemit.2019.035

multiple solution problem[11] and it can get four space poses at most[12]. But the current solutions about P3P or PNP are under the hypothesis that the feature points are matched correctly. However, feature automatic and accurate matching is a difficulty in computer vision.

The right match is that there is a one-to-one correspondence between image features and the corresponding object in 3D space. When mismatching occurs, PNP analytical algorithm will construct a wrong and fictive 3D object which does not exist at all. This 3D object satisfies the PNP equation forever, so the PNP algorithm cannot find such a mismatching phenomenon.

In real scene, the detected features from the image are more than the expected ones on the model, so how to pick the right matching feature sets automatically? How to get the only one right pose? How to get out of the local convergence in the process of VVS[4,12,13]? To solve all these problems, supervisory control theory based on VVS is introduced in this paper.

The algorithm uses the image square error as the performance evaluation index of supervisory control to get automatic and accurate feature matching and obtains the accurate 3D pose of object. The AR experiment shows the feasibility of our algorithm. In the remainder of the paper the principle of the approach is presented, and the algorithm flow details are shown. Brief introduction, results and comparison with the traditional VVS method are given in the experiment part. Finally, the conclusions are made in this paper.

2. Supervisory Control based VVS

Suppose 3D features P_i^M , i =1…n, on a rigid model M (t, R), where t and R are transition and rotation matrices with respect to the world coordinates. The corresponding image features are f_i^M at P_i^M, i=1…n.

1) Capture an image I and extract feature pi and their feature vectors fi.

2) Match f_i^M ,i=1…n, with fi and sort in descending order by similarity.

3) Pick up the top three points pi , i=1…3, which correspond to P_i^M in the model. Calculate k initial poses M (t1, R1), M (t2, R2)… M(tk,Rk)with the three points pi for virtual visual servo.

4) Repeat capturing image for the pi (t), i=1..3, and tracking f_i^M; in a fixed sampling period, the following a), b) and c) are continuously performed for several times.

- Visual servo for the right matching points

Define tracking error to be $e(k)=[e_1(k)\ e_2(k)\ e_3(k)]^T$, where $e_i(k)=p_i-p_i^{M(k)}$, $i=1,2,3$ and $k=1,2,3,4$ .We introduce a modified error with deadzone

$$e_\Delta(k)=e(k)-\varphi(k) \tag{1}$$

where $\varphi(k)=[\epsilon sat(\frac{e_1(k)}{\varepsilon})\ \epsilon sat(\frac{e_1(k)}{\varepsilon})\ \epsilon sat(\frac{e_1(k)}{\varepsilon})]^T$ with dead zone width ϵ. According to formula (1), we know that the image mean error can be controlled in dead zone when the feature points are matched correctly.

- Supervisory control for initial pose

Define a monotonically non-decreasing cost function $I(k,t)$ as

$$I(k,t)=\int_0^t \sum_{i=0}^n \left\| e_{\Delta i}(k,t) \right\| dt . \tag{2}$$

Let $\hat{k}_j \in K=\{k:1,2,3,4\}$ and j=0 is selected as the initial pose or visual servo based AR. At time t during the visual servo, if $I(\hat{k}_j,t)>\min_{k\in K} I(k,t)+\varepsilon$, then j=j+1 and

$$\hat{k}_j = \arg\min_{k\in K} I(k,t) ,$$

$\hat{k}_j$ is selected for visual servo based AR after time t.

- Falsified poses

Definition: A pose $k \in K$ is said to be falsified by measurement information if this information is

sufficient to deduce that the performance specification (p,e(k),u(k))∈Tspec would be violated if the object was controlled from the kth pose. Otherwise the pose k is said to be unfalsified.

Consider a γ dependent performance specification

$$T_{spec} = \{p,e,u \,|\, I(k,t) \le \gamma\}$$

Where γ is a positive threshold and design parameter; if for a given pose, the performance index exceeds this threshold at any time, it is not suitable for the actual unknown object and hence is falsified and taken out of the candidate pose set. Switching is done among as yet unfalsified candidate controllers only.

If all of the current 4 poses are falsified, go to step 1 for detecting new candidate poses.

3. Experiments

3.1 Experimental Conditions

In practice, it is hardly to get the actual 3D pose of object. Augmented reality (AR) is a live direct or indirect view of a physical, real-world environment whose elements are augmented (or supplemented) by computer-generated sensory input such as sound, video, graphics or GPS data. The 3D registration, which means that the coordinate system of virtual model and that of the real object is overlapped, is a key step in vision based AR. Our algorithm in which the 3D pose is estimated realizes the 3D registration in deed. The experiment shows AR effects based on a color cube by augmenting a teapot in the specific position.

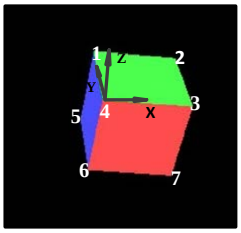

Figure 1. The virtual cube model.

Build a 3D color cube in which the length is 100mm. Every face of the model is filled with different colors Teapot model.

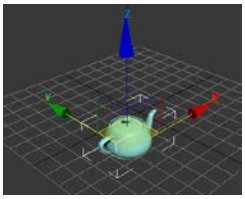

Figure 2. The teapot model.

In our paper, the coordinate system is attached on 4th vertex of the cube where the teapot is augmented. That is to say, whatever the pose of 3D cube is, the teapot will be augmented on the coordinate system of 4th vertex.

After the 3D pose of cube is estimated, the camera intrinsic parameters are used to get the image of AR objects.

3.2 Experiment Steps

Step 1 Do VVS

Choose the vertexes of the most possible relationship f_1 as the initial matcher; we will get 2 possible poses used 3 points by P3P. Take them as the initial pose for virtual vision servo respectively, after VVS the corresponding image mean square errors is shown as Fig 3. And t presents time presents the value of MSE. Fig 4 shows the actual teapot pose after virtual vision servo, clearly it is wrong.

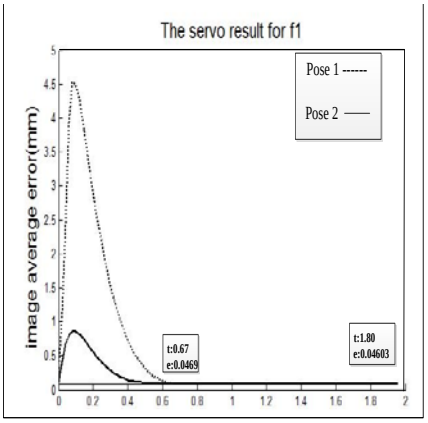

Figure 3. The integral square errors.

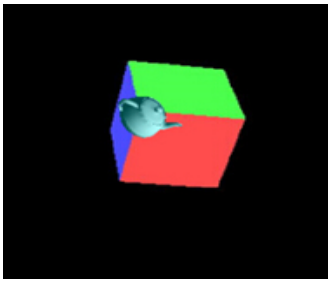

Figure 4. The end pose after VVS.

Step 2 Make switch decision based on supervisory control

Under the intervention of supervisory control, it finally switches to the correct matching feature set, namely realizing automatic object recognition. After that, the image MSE changes with the changing frames, and the change value depends on the camera's velocity. At end, it converges to global minimum, and the correct AR results of 1st input image are show in figure 5.

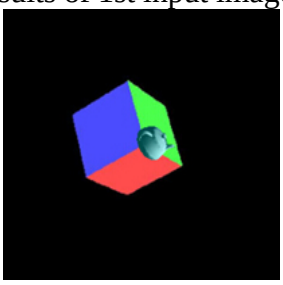

Figure 5. AR Results of 1st input image.

4. Conclusion

Our algorithm based on the existing performance evaluation function, logic switching evaluates n possible feature sets, and selects one possible feature set which matches the target object, and makes this feature set as the input of virtual visual servo. In addition, logic switching should switch among all the possible space poses. By this way the algorithm can reject the uncorrected matchers and can get the pose of 3D object automatically. The proposed algorithm can get global convergence value with any initial pose. It's more robust, accurate and can be widely used in AR. There are several advantages in this paper as following: a) Use feedback control method to solve vision cognition problem; b) Switch in the alternative candidate feature sets with supervisory controller; c) Make full use of the matching relationships on the 3D model, and achieve the accurate feature matching with the internal exercise of VVS; d) Solve the problem of 3D pose estimation and tracking as well as object recognition.

References

[1] F. Chalmette, S. Hutchinson. “Visual servo control. I. Basic approaches,” IEEE Robot. Autom. Mag., Vol.13, Issue 4, pp. 82-90, 2006.

[2] S. Hutchinson, G. D. Hager, and P. I. Corke. “A tutorial on visual servo control,” IEEE Trans. Robot. Autom., Vol. 12, Issue 5, pp. 651-670, 1996.

[3] A. I. Comport, E. Marchand, M. Pressigout, et al. “Real-time markerless tracking for augmented reality: the virtual visual servo framework,” IEEE Trans. Vis. Comput. Graphics, Vol. 12, Issue 4, pp. 615-628, 2006.

[4] M. Li, X. N. Wang, and J. Zhu. “A New Virtual Visual Servo Algorithm ,” Advanced Materials Research, pp. 1615-1620, 2014.

[5] C. P. Lu, G. D. Hager, and E. Mjolsness. “Fast and globally convergent pose estimation from video images,” IEEE Trans. Pattern Anal. Mach. Intell., Vol. 22, Issue 6, pp. 610-622, 2000.

[6] D. Kragic, V. Kyrki. “Initialization and System Modeling in 3-D Pose Tracking,” 18th International Conference on Pattern Recognition(ICPR’06), IEEE Press, Aug. 2006, pp. 643-646.

[7] M. A. Fischler, R. C. Bolles. “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” Communications of the ACM, Vol. 24, Issue 6, pp. 381-395, 1981.

[8] G. Schweighofer, A. Pinz. “Robust pose estimation from a planar target,” IEEE Trans. Pattern Anal. Mach. Intell., Vol. 28, Issue 12, pp. 2024-2030, 2006.

[9] V. Lepetit, F. Moreno-Noguer, and P Fua. “Epnp: An accurate o (n) solution to the pnp problem,” Int. J. Comput. Vision, Vol. 81, Issue 2, ,pp. 155-166, 2009.

[10] X. S. Gao, X. R. Hou, J. Tang, et al. “Complete solution classification for the perspective-three-point problem,” IEEE Trans. Pattern Anal. Mach. Intell., Vol. 25, Issue 8, pp. 930-943, 2003.

[11] I. Misra, S. M. Moorthi, D. Dhar, and R. Ramakrishnan. “An automatic satellite image registration technique based on Harris corner detection and Random Sample Consensus (RANSAC) outlier rejection model,” 2012 1st International Conference on Recent Advances in Information Technology (RAIT), IEEE Press, March 2012, pp. 68-73.

[12] Emmanuel Dean-León, Gordon Cheng. “A new method for solving 6D Image-Based Visual Servoing with Virtual Composite camera model,” 2014 14th IEEE-RAS International Conference on Humanoid Robots, IEEE Press, Nov. 2014, pp. 519-525

[13] A. Assa, F. Janabi-Sharifi. “Virtual Visual Servoing for Multicamera Pose Estimation,” IEEE-ASME T. Mech., Vol. 20, Issue 2, pp.789-798, 2015.

[14] F. Wang, Z. Feng, S. Liu, and P. Jiang. “Robust Supervisory Control of Fuzzy Discrete Event Systems,” IET Control Theory & Applications, Vol. 2, Issue 5, pp. 384-391, 2008.

[15] H. Logemann, B. Mårtensson. “Adaptive stabilization of infinite-dimensional systems,” IEEE Trans. Autom. Control, Vol. 37, Issue 12, pp. 1869-1883, 1992.

[16] L. Vu, D. Liberzon. “Supervisory Control of Uncertain Linear Time-Varying Systems,” IEEE Trans. Autom. Control, Vol. 56, Issue 1, pp. 27-42,2011.