# Human-Face-Tracking Using Visual Servoing by Patient Robot

Shota Oonishi<sup>1</sup>, Akira Yanou<sup>1</sup> and Mamoru Minami<sup>1</sup>

<sup>1</sup>Graduate School of Natural Science and Technology, Okayama University

(Tel,Fax: +81-86-251-8924;Email: en20815@s.okayama-u.ac.jp, {yanou-a, minami-m}@cc.okayama-u.ac.jp)

**Abstract:** Human-body models imitating a part of human body have been used for medical training. But those models seem not to suit training because they cannot be felt to be similar to actual humans. Therefore, the simulator being felt like a real human is required for the medical trainees being able to assess patient's physical and emotional conditions, since medical care procedures are vital to watch out patient's pains given by the care treatment itself to avoid failures leading the patients to dangerous conditions. Based on the needs for medical training by using human-like robot, we have developed a simulator for injection training called "patient robot." Patient robot has to behave like humans to make the medical training effective, so we introduced a visual servoing system into the patient robot so that the patient robot can gaze at a trainee's face like humans or turn away its face form the medical procedures. We executed some experiments of gazing at a rotating object using visual servoing, in which actual object's positions, patient robot's gazing point and pose-tracking data have been measured. We confirmed patient robot can track real human faces.

Keywords: Patient robot, Visual servoing, Object tracking

## **1. INTRODUCTION**

Nowadays, some human-body models called "phantom" imitating parts of human body have been developed, most of them are used for technical training for particular training objective. Those phantoms do not suit for medical trainings such as injection-the medical procedure that should be conducted with a sense of carefully monitoring the patients' body condition to avoid medical accidents—since the phantoms are just parts of body and then cannot be felt to be humans alive. What is important for nurse to prevent medical accidents is the constant awareness to monitor patients' physical conditions. Medical workers-especially beginners-may fall themselves in a state of concentration too much on medical procedures without paying attention to the patients' condition, which may change quickly and dangerously. Therefore constant awareness of the patients' state is important, which is the reason human-like patient robot is required instead of phantoms. On-line monitoring training through patient robot helps the trainees notice sudden change of patients' conditions, preventing medical malpractices before dropping in irrecoverable situation.

Therefore, we have developed a new simulator called "patient robot" as shown in Fig.1. To offer safe and effective nursing training, the patient robot must present its mental activity through expressions and body behaviors since nurses are required to monitor the patients' conditions during nursing procedures. On the other hand the robot can monitor the nurse students' action, e.g., injection, to evaluate their ability from the view point of patients. Moreover, the patient robot should behave autonomously and naturally like humans to make the injection training effective. In order to implement human-like behavior, we added a new function of visual servoing to the robot's behavior.

In the field of robot vision, a control method called visual servoing has attracted attention [1]-[4]. Visual servoing is a method of controlling robots' motion through



Fig. 1 Patient robot

visual information in feedback loop. Thus this method is expected to help the robot to adapt changing or unknown environments.

A fixed-hand-eye system—a robotic system for visual servoing where eye cameras are set at the robot's hand with fixed orientation—has some disadvantages, making the observing ability deteriorated depending on the relative geometry between the camera and the seroving target. Such as: the robot cannot observe the object well when it is near the cameras (Fig.2 (a)), small intersection of the possible sight space of the two cameras (Fig.2 (b)), and the image of the object cannot appear in the center of both cameras, thus we could not get clear image information of target, reducing the pose measurement accuracy (Fig.2 (c)). To solve the problems above, Eye-Vergence system that gives the cameras an ability to rotate themselves to focus target at center of the images has been thought to be effective that enables eye-camera's orien-



Fig. 2 Disadvantage of fix camera system



Fig. 3 Advantage of Eye-vergence system

tation change in case of Eye-Vergence system being installed into patient robot.

There has not seemed that such rotatable hand-eye system for training robot in medical field has been developed. Thus it is possible to change the pose of the cameras in order to observe the object better, as it is shown in Fig.3, enhancing the measurement accuracy in trigonometric calculation and avoiding peripheral distortion of camera lens by observing target at the center of lens. Moreover, recent researches on visual servoing are limited generally in a swath of tracking an object while keeping a certain constant distance [5], [6], [7].

In this paper, we propose visual servoing system that enables patient robot recognize and track human face. In the following sections, a Model-based Matching method and human recognition technique using GA search are specifically introduced, and human's face is detected by using orientation expression of quaternion, making the patient robot's motion resemble real humans motions, i.g. looking into the opponent's face to estimate his/her emotions.

## **2. PATIENT ROBOT**

The patient robot we developed is shown in Fig. 1. We mounted the robot's head with two CCD cameras as eyes to observe the nurse being trained and installed some servomotors inside the head for generating face expression, as shown in Fig.4. The moving parts of patient robot's body are shown in Fig.5. Left arm is made by arm model for blood drawing training, and the artificial vein letting imitated blood flow inside is buried in the arm. Since checking the state of patient periodically is necessary to avoid danger during nursing, the robot detects student's face with eye-cameras to evaluate whether the nurse is paying attentions to the state of patient while injecting[8].



Fig. 4 Structure of robot's head



Fig. 5 Structure of robot's body



Fig. 6 Coordinate system

## **3. ON-LINE TRACKING**

## 3.1 Kinematics of Stereo Vision

A perspective projection is used as projection transformation. The coordinate systems of left and right cameras and object (here we take a solid column model as an example) in Fig.6 represent world coordinate system  $\Sigma_W$ , model coordinate system  $\Sigma_M$ , camera coordinate systems  $\Sigma_{CR}$  and  $\Sigma_{CL}$ , image coordinate systems  $\Sigma_{IR}$  and  $\Sigma_{IL}$ . A point *i* on a solid model of the target object can be described using these coordinates and homogeneous transformation matrices. At first, a homogeneous transformation matrix from  $\Sigma_{CR}$  to  $\Sigma_M$  is defined as  ${}^{CR}T_M$ . And an arbitrary point *i* on the target object in  $\Sigma_{CR}$  and  $\Sigma_M$  is defined as  ${}^{CR}r_i$  and  ${}^{M}r_i$ . Then  ${}^{CR}r_i$  is,

$$^{CR}\boldsymbol{r}_{i} = ^{CR} \boldsymbol{T}_{M}{}^{M}\boldsymbol{r}_{i}. \tag{1}$$

The position vector of i point in right image coordinates,  ${}^{IR}r_i$  is described by using projection matrix P of camera as,

$$^{IR}\boldsymbol{r}_{i}=\boldsymbol{P}^{CR}\boldsymbol{r}_{i}. \tag{2}$$

Using a homogeneous transformation matrix of fixed values defining the kinematical relation from  $\Sigma_{CL}$  to  $\Sigma_{CR}$ ,  ${}^{CL}\boldsymbol{T}_{CR}$ ,  ${}^{CL}\boldsymbol{T}_{CR}$ ,  ${}^{CL}\boldsymbol{r}_{i}$  is,

$$^{CL}\boldsymbol{r}_{i} = ^{CL} \boldsymbol{T}_{CR} ^{CR} \boldsymbol{r}_{i}. \tag{3}$$

By the same way as we have obtained  ${}^{IR}r_i$ ,  ${}^{IL}r_i$  is described by the following Eq.(4) through projection matrix P.

$$^{IL}\boldsymbol{r}_{i} = \boldsymbol{P}^{CL}\boldsymbol{r}_{i} \tag{4}$$

Then position vectors projected in the  $\Sigma_{IR}$  and  $\Sigma_{IL}$  of arbitrary point *i* on target object can be described as  ${}^{IR}r_i$ and  ${}^{IL}r_i$ . Here, position and orientation, i.e. pose of the origin of  $\Sigma_M$  based on  $\Sigma_{CR}$ , are represented as  $\phi = [t_x, t_y, t_z, \phi, \theta, \psi]^T$ , in which  $\phi, \theta$ , and  $\psi$  are roll, pitch and yaw angles respectively, and then Eq. (2) and Eq. (4) are rewritten as,

$$\begin{cases} {}^{IR}\boldsymbol{r}_i = f_R(\boldsymbol{\phi}, {}^{M}\boldsymbol{r}_i) \\ {}^{IL}\boldsymbol{r}_i = f_L(\boldsymbol{\phi}, {}^{M}\boldsymbol{r}_i). \end{cases}$$
(5)

This relation connects the arbitrary points on the object and projected points on the left and right images with the variables  $\phi$  representing the human face's pose, which is considered to be unknown in this paper. When evaluating each left and right point *i* above mentioned, the matching problem of corresponding point in left and right images is arisen, and it is sometimes difficult to be solved. Therefore, to avoid this problem, the 3-D model-based matching that treats the points of the object model as a set, is chosen instead of point-based corresponding.

### 3.2 3-D Object Pose Tracking

In this paper, patient robot recognizes human face to decide the transition of each process automatically. This method is given by using Model-Based Matching(MBM) method and genetic algorithm(GA)[9]. In action patterns of patient robot of the previous system, the border of each process had been decided by operator of the robot. Since each process must relate to the recognition of human face, we propose 3-D object pose recognition method[10] of patient robot by using two CCD cameras as robot's eyes.

## **4. PATIENT ROBOT SYSTEM**

We have built a model of patient robot to design a dynamics of it. The parameters are shown in Table 1. In this report we built a simplified model since it is difficult to solve a solution of inverse kinematics. Specifically we cut the injection of link 3 and we built the model in a case of  $q_i = 0$  (i = 1, 2, 3, 4) so that an end-effector of link 4 becomes vertical to the ground. Each links are defined as *i*. The model is shown in Fig.7. A body is defined as  $q_{1,2,3,4}$  and head is defined as  $q_{5,6,7}$ .



Fig. 7 Coordinate of model

## **5. POSITION-BASED CONTROLLER**

#### 5.1 Desired-trajectory Generation

In Fig.8, the world coordinate frame is denoted by  $\Sigma_W$ , the target coordinate frame is denoted by  $\Sigma_M$ , and the desired and actual end-effector coordinate frame is denoted by  $\Sigma_{Hd}$ ,  $\Sigma_H$  respectively. The desired relation between the target and the end-effector is given by Homogeneous Transformation as  ${}^{Hd}T_M$ , the relation between the target and the actual end-effector is given by  ${}^{H}T_M$ , then the difference between the desired end-effector pose  $\Sigma_{Hd}$  and the actual end-effector pose  $\Sigma_H$  is denoted as  ${}^{H}T_{Hd}$ , which can be described by:

$${}^{H}\boldsymbol{T}_{Hd}(t) = {}^{H}\boldsymbol{T}_{M}(t){}^{Hd}\boldsymbol{T}_{M}^{-1}(t)$$
(6)

(6) is a general representation of pose tracking error that satisfies arbitrary object motion  ${}^{W}\boldsymbol{T}_{M}(t)$  and arbitrary visual servoing objective  ${}^{Hd}\boldsymbol{T}_{M}(t)$ . The relation  ${}^{H}\boldsymbol{T}_{M}(t)$  can be estimated by 1-step GA [5], having been presented as an on-line model-based pose estimation method that will be introduced in next subsection. Let  $\Sigma_{\hat{M}}$  denote the detected object, it is natural there should always exist an error between the actual object  $\Sigma_{M}$  and the detected one  $\Sigma_{\hat{M}}$ . So in visual servoing, (6) will be rewritten based on  $\Sigma_{\hat{M}}$  that includes the error  ${}^{M}\boldsymbol{T}_{\hat{M}}$ , as

$${}^{H}\boldsymbol{T}_{Hd}(t) = {}^{H}\boldsymbol{T}_{\hat{M}}(t){}^{Hd}\boldsymbol{T}_{\hat{M}}^{-1}(t).$$
 (7)

Differentiating (7) with respect to time yields

$${}^{H}\dot{T}_{Hd}(t) = {}^{H}\dot{T}_{\hat{M}}(t)^{\hat{M}}T_{Hd}(t) + {}^{H}T_{\hat{M}}(t)^{\hat{M}}\dot{T}_{Hd}(t).$$
(8)

Differentiating Eq. (8) with respect to time again

ŀ

$${}^{H}\ddot{T}_{Hd}(t) = {}^{H}\ddot{T}_{\hat{M}}(t)^{\hat{M}}T_{Hd}(t) + 2^{H}\dot{T}_{\hat{M}}(t)^{\hat{M}}\dot{T}_{Hd}(t) + {}^{H}T_{\hat{M}}(t)^{\hat{M}}\ddot{T}_{Hd}(t),$$
(9)

Table 1 Physical parameters of the patient robot

Joint	Base	link1	link2	link3	link4	link5	link6	link7
Length(m)	0.58	0.08	0.205	0.270	$0.3_x, 0.3_z$	0.070	0.035	0.035
Center of mass $(m)$	0.0	-0.04	0.1025	0.135	$0.15_x, 0.15_z$	0.0	0.0175	0.0175
mass~(Kg)	N/A	3.0	3.0	3.0	1.0	1.0	1.0	1.0



Fig. 8 Motion of the end-effector and object

where  ${}^{\hat{M}}\boldsymbol{T}_{Hd}$ ,  ${}^{\hat{M}}\dot{\boldsymbol{T}}_{Hd}$ ,  ${}^{\hat{M}}\ddot{\boldsymbol{T}}_{Hd}$  are given as the desired visual servoing objective.  ${}^{H}\boldsymbol{T}_{\hat{M}}$ ,  ${}^{H}\dot{\boldsymbol{T}}_{\hat{M}}$ ,  ${}^{H}\ddot{\boldsymbol{T}}_{\hat{M}}$  can be observed by cameras. As shown in Fig. 8, there are two errors that we have to decrease in the visual servoing process. First one is the error between the actual object and the detected one,  ${}^{M}\boldsymbol{T}_{\hat{M}}$ , and the other one is the error between the desired end-effector and the actual one,  ${}^{H}\boldsymbol{T}_{Hd}$ .

#### 5.2 Hand & Eye Visual Servoing Controller

5.2.1 Hand Visual Servoing Controller

Based on the above analysis of the desired-trajectory generation, the desired hand velocity  ${}^{W}\dot{r}_{d}$  is calculated as,

$${}^{W}\dot{\boldsymbol{r}}_{d} = \boldsymbol{K}_{P_{p}}{}^{W}\boldsymbol{r}_{H,Hd} + \boldsymbol{K}_{V_{p}}{}^{W}\dot{\boldsymbol{r}}_{H,Hd}, \qquad (10)$$

where  ${}^{W}\boldsymbol{r}_{H,Hd}, {}^{W}\dot{\boldsymbol{r}}_{H,Hd}$  can be calculated from  ${}^{H}\boldsymbol{T}_{Hd}$ and  ${}^{H}\dot{\boldsymbol{T}}_{Hd}$ .  $\boldsymbol{K}_{P_{p}}$  and  $\boldsymbol{K}_{V_{p}}$  are positive definite matrix to determine PD gain.

The desired hand angular velocity  ${}^{W}\boldsymbol{\omega}_{d}$  is calculated as,

$${}^{W}\boldsymbol{\omega}_{d} = \boldsymbol{K}_{P_{o}}{}^{W}\boldsymbol{R}_{H}{}^{H}\Delta\boldsymbol{\epsilon} + \boldsymbol{K}_{V_{o}}{}^{W}\boldsymbol{\omega}_{H,Hd}, \quad (11)$$

where  ${}^{H}\Delta\epsilon$  is a quaternion error [11] calculated from the pose tracking result, and  ${}^{W}\omega_{H,Hd}$  can be computed by transforming the base coordinates of  ${}^{H}T_{Hd}$  and  ${}^{H}\dot{T}_{Hd}$ from  $\Sigma_{H}$  to  $\Sigma_{W}$ . Also,  $K_{P_{o}}$  and  $K_{V_{o}}$  are suitable feedback matrix gains. The desired hand pose is defined as  ${}^{W}\psi_{d}^{T} = [{}^{W}r_{d}^{T}, {}^{W}\epsilon_{d}^{T}]^{T}$ 

The desired joint variable  $oldsymbol{q}_d$  and  $\dot{oldsymbol{q}}_d$  is obtained by

$$\boldsymbol{q}_d = \boldsymbol{f}^{-1}(^W \boldsymbol{\psi}_d^T) \tag{12}$$

$$\dot{\boldsymbol{q}}_d = \boldsymbol{J}^+(\boldsymbol{q}) \begin{bmatrix} & \boldsymbol{r}_d \\ & \boldsymbol{W}_{\boldsymbol{\omega}_d} \end{bmatrix}$$
 (13)

where  $f^{-1}({}^{W}\psi_{d}^{T})$  is the inverse kinematic function and  $J^{+}(q)$  is the pseudoinverse matrix of J(q), and  $J^{+}(q) = J^{T}(JJ^{T})^{-1}$ .

The hardware control system of the velocity-based servo system of PA10 is expressed as

$$\boldsymbol{\tau} = \boldsymbol{K}_{SP}(\boldsymbol{q}_d - \boldsymbol{q}) + \boldsymbol{K}_{SD}(\dot{\boldsymbol{q}}_d - \dot{\boldsymbol{q}})$$
(14)

where  $K_{SP}$  and  $K_{SD}$  are symmetric positive definite matrices to determine PD gain.

#### 5.2.2 Eye-vergence Visual Servoing Controller

In this paper, we use two pan-tilt cameras for eyevergence visual servoing. For camera system,  $q_5$  is tilt angle for both right and left eyes,  $q_6$  and  $q_7$  are pan angles.

 ${}^{H}x_{\hat{M}}, {}^{H}y_{\hat{M}}, {}^{H}z_{\hat{M}}$  express position of the detected object in the hand coordinate. The desired angle of the camera joints are calculated by:

$$q_{5d} = atan2({}^{H}y_{\hat{M}}, {}^{H}z_{\hat{M}})$$
(15)

$$q_{6d} = atan2(-d_{CR} + {}^{H}x_{\hat{M}}, {}^{H}z_{\hat{M}})$$
(16)

$$_{7d} = atan2(d_{CL} + {}^{H}x_{\hat{M}}, {}^{H}z_{\hat{M}})$$
 (17)

where  $d_{CR} = d_{CL} = 35$  [mm] that is the camera location.

Because the motion of camera motor is an open loop, we can only make it rotate a certain degree without getting the actual angle during the rotation, which put us in a situation that we cannot get the accurate camera angle. So the desired camera angles are input in every 33[ms], and the input is limited to a certain value.

## **6. EXPERIMENT**

Two experiments to confirm patient robot's tracking ability have been done. Each experiment is shown below.

### 6.1 Rotating Object Tracking

q

Having had the patient robot track human face model rotating at a constant cycle T = 30, 40[s] in an experimental environment as shown in Fig.9, and then the actual object's position, patient robot's gazing point and fitness values have been measured. Patient robot's gazing point is defined by the center of image as shown in Fig.10. Fitness function is shown in the following formula.

$$F_{site}(\boldsymbol{\phi}) = \left\{ \left( \sum_{IR} \boldsymbol{r}_{i \in S_{R,in}(\boldsymbol{\phi})} p(^{IR}\boldsymbol{r}_{i}) - \sum_{IR} p(^{IR}\boldsymbol{r}_{i}) \right) + \left( \sum_{IL} \boldsymbol{r}_{i \in S_{L,in}(\boldsymbol{\phi})} p(^{IL}\boldsymbol{r}_{i}) - \sum_{IL} p(^{IL}\boldsymbol{r}_{i}) - \sum_{IL} p(^{IL}\boldsymbol{r}_{i}) \right) \right\} / 2$$



Fig. 9 Situation(rotate object)



Fig. 10 Camera image



Fig. 11 Searching model

$$= \left\{ F_{R,site}(\boldsymbol{\phi}) + F_{L,site}(\boldsymbol{\phi}) \right\} / 2 \tag{18}$$

 $p({}^{IR}\boldsymbol{r}_i)$  is the brightness value in the right image area  ${}^{IR}\boldsymbol{r}_i$ , and  $p({}^{IL}\boldsymbol{r}_i)$  is the brightness value in the left image area  ${}^{IL}\boldsymbol{r}_i$ . In order to evaluate facial ratings and the surrounding luminance value change, as shown Fig.11, search model is composed of surface areas comprised of  $S_{R,in}$  and  $S_{L,in}$  and stripe areas comprised of  $S_{R,out}$  and  $S_{L,out}$ .

The results are shown in Fig.12  $\sim$  Fig.17. Fig.12, Fig.15 show y-position of actual object and gazing point. Fig.13, Fig.16 show z-position of actual object and gazing point. Fig.14, Fig.17 show time profile of fitness value.

As shown in Fig.12, 13, 15 and 16, gazing point is attenuated sometime, however, Fig.14 and Fig.17 show high fitness value. Therefore, patient robot can track the rotating model.

#### 6.2 Tracking To Real Human

In this experiment, having had the patient robot track a living human face, and then control error as shown Fig.10 between object position and gazing point and fitness value have been plotted. Human facing the front moved about 0.5[m] side to side and returned started position.

The results are shown in Fig.18  $\sim$  Fig.20. Fig.18 and Fig.19 show control error. Fig.20 shows fitness value.

As the results of Fig.18 and Fig.19, control error is only up to about 60[mm], so that patient robot can follow real human face. There are two points fitness value increase greatly in Fig.20. It is considered that the face was slowed down because these points are changing human move direction.

According to these results, we confirmed patient robot can track the object.

## 7. CONCLUSION

We have proposed a eye-vergence visual servoing system for patient robot that can track the nurse trainee's face. The obtained experimental data show that the eyevergence system can improve pose tracking performance to follow human's face.

## REFERENCES

- S.Hutchinson, G.Hager, and P.Corke, "A Tutorial on Visual Servo Control", IEEE Trans. on Robotics and Automation, vol. 12, no. 5, pp. 651-670, 1996.
- [2] P.Y.Oh, and P.K.Allen, "Visual Servoing by Partitioning Degrees of Freedom", IEEE Trans. on Robotics and Automation, vol. 17, no. 1, pp. 1-17, 2001.
- [3] E.Malis, F.Chaumentte and S.Boudet, "2-1/2-D Visual Servoing", IEEE Trans. on Robotics and Automation, vol. 15, no. 2, pp. 238-250, 1999.
- [4] P.K.Allen, A.Timchenko, B.Yoshimi, and P.Michelman, "Automated Tracking and Grasping of a Moving object with a Robotic Hand-Eye System", IEEE Trans. on Robotics and Automation, vol. 9, no. 2, pp. 152-165, 1993.
- [5] W. Song, M. Minami, Y. Mae and S. Aoyagi, "Online Evolutionary Head Pose Measurement by Feedforward Stereo Model Matching", IEEE Int. Conf. on Robotics and Automation (ICRA), pp.4394-4400, 2007.
- [6] Omar Tahri and Francois Chaumette, "Point-Based and Region-Based Image Moments for Visual Servoing of Planar Objects", IEEE Tran. on Robotics, vol. 21, no. 6, Dec 2005.
- [7] Tarek Hamel and Robert Mahony, "Visual Servoing of an Under-Actuated Dynamic Rigid-Body System: An Image-Based Approach", IEEE Trans. on Robotics and Automation, VOL. 18, NO. 2, APRIL 2002.
- [8] Yasushi Mae, Wei song, Tsuyoshi Ueki and Mamoru Minami: "Patient Robot for Training and Evaluation of Nursing Skill-Front Face Detection from View of

Patient in Injection," *Journal of RSJ*, vol.23, 3J35, 2005.

- [9] H. Suzuki and M. Minami: "Visual Servoing to catch fish Using Global/local GA Search," *Proc. of IEEE/ASME Transactions on Mechatronics*, Vol.10, Issue 3, pp.352–357, 2005.
- [10] Wei Song, Mamoru Minami and Yasushi Mae: "Evolutionary Head Pose Measurement by Improved Stereo Model Matching," *Proc. of SICE-ICASE International Joint Conference*, pp.4234– 4239, 2006.
- [11] W. Song, M. Minami, S. Aoyagi, "On-line Stable Evolutionary Recognition Based on Unit Quaternion Representation by Motion-Feedforward Compensation", International Journal of Intelligent Computing in Medical Sciences and Image Processing (IC-MED) Vol. 2, No. 2, Page 127-139 (2007).



Fig. 15 Y-position(T = 30[s])



Fig. 20 Fitness value