Journal of Computer Science 2012, 8 (12), 2062-2067
ISSN 1549-3636

doi:10.3844/jcssp.2012.2062.2067 Published Online 8 (12) 2012 (http://www.thescipub.com/jcs.toc)

The Framework of Intelligent Telepresence Robot Based on Stereo Vision

Widodo Budiharto, Muchsin Shodiq, Bayu Kanigoro and Hutomo

Department of Computer Science, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia

Received 2012-10-18, Revised 2012-10-23; Accepted 2012-10-23

ABSTRACT

This research reports and presents a BeeBot, BINUS multi-client intelligent telepresence robot, a custom-build robot system specifically designed for teleconference with multiple people. The robot is controlled using computer networks, so the manager/supervisor can direct the robot to the intended person to start a discussion/inspection. People tracking and autonomous navigation are intelligent features of this robot using stereo vision. We build a web application for controlling the multi-client telepresence robot and open-source teleconference system used. Experimental result presented and we evaluated its performance.

Keywords: Telepresence Robot, Robot Vision, Stereo Vision, Intelligent Robot

1. INTRODUCTION

Telepresence robots can be deployed in a wide range of application domains, e.g., in workplaces, the public sector or for home use. The idea of a mobile telepresence robot stems from the inherent limitations imposed by traditional video conferencing systems, in which interaction is restricted to the meeting room only. Telepresence robots are already being used in hospitals to allow doctors and specialists to give consultations from afar. In Telerobotics, Automation and Human Supervisory Control, Thomas B. Sheridan shows that progress in robotics depends not only on change in technology, but also on advances in human's relationships to machines. He said tthat "human supervisory control" has the potential to bring robotics out of the laboratory and into the difficult and messy world (Sheridan, 1992).

Traditionally, telepresence robots can also be used to give people with restricted mobility a new way to outreach and interact beyond their usual living quarters (Alers *et al*., 2011; Lazewatsky and Smart, 2011; Quigley *et al*., 2009). In office or factory, sometimes manager/supervisor wants to discuss/supervise staffs remotely. So, based on that situation, telepresence robot with capability of multi-client can be used on that scenario. An assisted teleoperation feature for a Mobile Remote Presence (MRP) system focusing on both the system-oriented dimensions (e.g., autonomous assistance vs. no assistance) and human-oriented dimensions (e.g., gaming experience, locus of control and spatial cognitive abilities) (N = 24) have been developed but there is no capability to control multi client (Takayama *et al*., 2011).

State of the art of this research is to propose the framework of a multi-client intelligent telepresence robot called BeeBot. This robot can be used in office/factory with the features such as video teleconferencing using JITSI framework, people tracking, obstacles avoidance and fast movement using omniwheel mechanism. We also develop an easy web application for controlling the robots on the web.

2. MATERIALS AND METHODS

2.1. Kinematics Model of Omni Wheels Robot

A holonomic or omni-directional robot is capable of driving in any direction. If we want to prescribe the robot's movements in the environment, we need to know how these variables relate to the primary variables we

Corresponding Author: Widodo Budiharto, Department of Computer Science, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia

can control: The angular positions and velocities of the wheel shafts. Therefore, a kinematical model of the robot has to be developed. A global frame [x, y] represents the environment of the robot and the robot's location can be represented as (x, y, θ). The global velocity of the robot can be written as $\dot{x}, \dot{y}, \dot{\theta}$. We can also define a local frame [xL, yL] that is attached to the robot itself. The center of this local frame coincides with the center of gravity of the robot. The three omni-wheel are located at an angle a_i (i = 1, 2, 3) relative to the local frame as shown in **Fig. 1**. If we take the local axis xL_{as} starting point and count degrees in the clockwise direction as positive, we have $\alpha_1 = 0$, $\alpha_2 = 120$ and $\alpha_3 = 240$°C.

The translational velocities of the wheels v_i on the floor determine the global velocity of the robot in the environment $\dot{x}, \dot{y}, \dot{\theta}$ and vice versa. The translational velocity of wheel hub v_i can be divided into a part due to pure translation of the robot and a part due to pure rotation of the robot Equation 1:

$$v_i = v_{trans,i} + v_{rot} \tag{1}$$

When the platform executes a pure rotation, the hub speed v_i needs to satisfy the following Equation 2:

$$v_{rot} = R\dot{\theta} \tag{2}$$

Here R is the distance from the center of gravity of the robot to the wheels along a radial path. The angular velocity of each wheel are Equation 3-5:

$$\dot{\phi}_1 = \left(-\sin(\theta)\, c\dot{o}s(\theta)\dot{x}_L + \cos^2(\theta)y_L + R\dot{\theta}\right)/r \tag{3}$$

$$\dot{\phi}_2 = \left(-\sin(\theta + \dot{a}_2)\cos(\theta)\dot{x}_L + \cos(\theta + a_2)\cos(\theta)\dot{y}_L + R\dot{\theta}\right)/r \tag{4}$$

$$\dot{\phi}_3 = \left(-\sin(\theta + \dot{a}_3)\cos(\theta)\dot{x}_L + \cos(\theta + a_3)\cos(\theta)\dot{y}_L + R\dot{\theta}\right)/r \tag{5}$$

For control, we have the angular positions ϕ_1, ϕ_2, ϕ_3 and the velocities of the wheel shafts $\dot{\phi}_1, \dot{\phi}_2, \dot{\phi}_3$ at our disposal by changing the value to the PWM controller.

2.2. System Specification

We use Stereo vision becase its is very useful for telepresence robot for depth estimation. The basic of visual-perception model for a face tracking system is shown in **Fig. 2**. After image acquisition, image pre-processing used to filter the noise. In feature extraction phase, the system will try to recognizes someone and in interpretation phase, commands to controller issued to follow a person (Budiharto *et al*., 2011a).

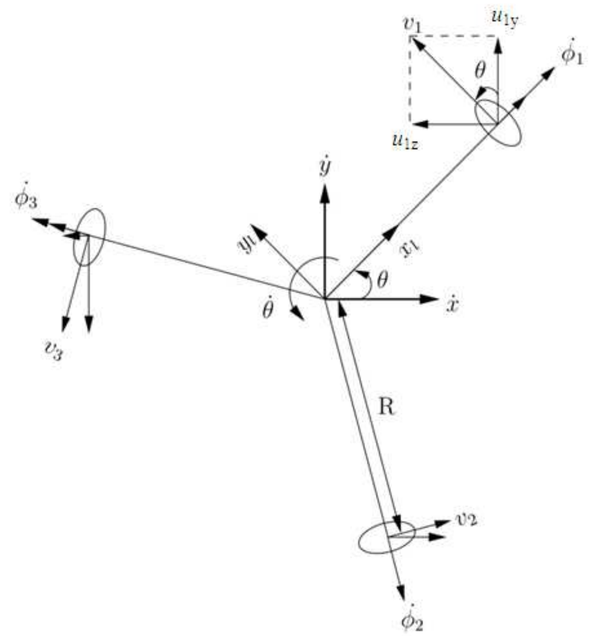

Fig. 1. Kinematics model of omnidirectional wheels

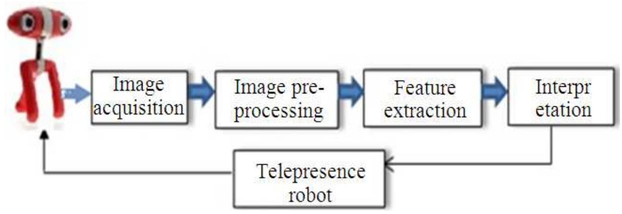

Fig. 2. Visual perception model for telepresence robot

The BeeBot platform is based on the propeller processor with 3 omni-directional wheels. A pole is fitted on the base plate and serves as the elevated attachment point for the laptop, speakers, 1 camera for pointing forward for conversations, 1 stereo camera in used for people tracking and obstacles avoidance and can be used to extract additional features from the environment. The robot has an overall height of 130 cm, the size of a small person, allowing for natural conversation while standing or being seated. Additionally, obstacle avoidance is implemented using 4 ultrasonic sensors, which provides assistance during manual operation or full autonomous navigation if desired. People and face tracking can be used for natural interactions using open CV. We implement the multiple obstacles avoidance based on stero camera Minoru 3D with Kalman filtering from previous work (Budiharto *et al*., 2011b).

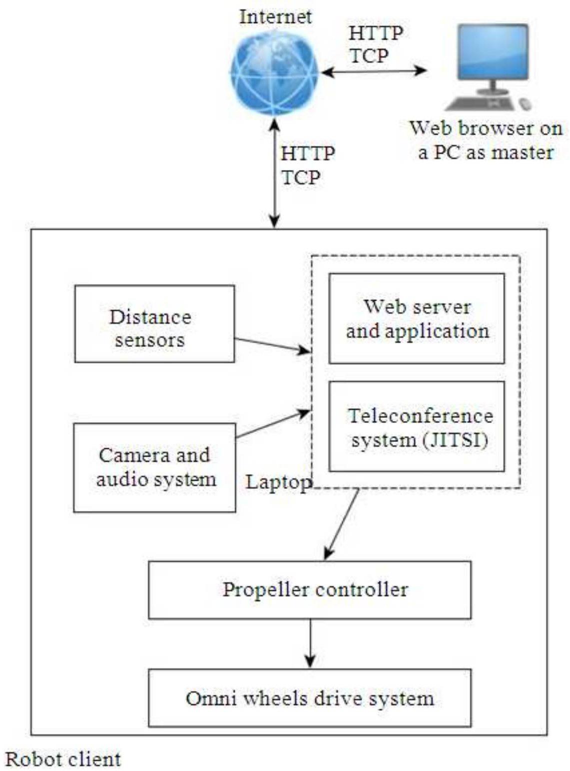

Fig. 3. Interconnecting robot client to master through internet

Fig. 4. Manager/supervisor have the ability to teleconference /inspection with many person

Interconnecting master PC to client robot through Internet shown in **Fig. 3**. To increase mobility of the robot, Ethernet WiFi will be employed and HTTP over TCP is the connection protocol for the robot. The video and audio information will be sent through HTTP based delivery system (Kutinsky, 2007). We develop web application using ASP. Net for controlling the robot under Web.

2.3. Softwares

For this research we are using openfire, Jingle Nodes and JITSI for video conference. Openfire is a Real Time Collaboration (RTC) that use XMPP (also called Jabber) as a protocol for instant messaging (http://www.igniterealtime.org/projects/openfire/). The XMPP network uses a client-server architecture (clients do not talk directly to one another). Jingle Nodes is an XMPP Extension that enables users to share and discover P2P Media Relays that can be used to enable voice and video chat. JITSI is an audio/video and chat communicator that supports XMPP/Jabber protocols (http://code.google.com/p/jinglenodes/). Openfire and Jingle Nodes are used in server side and JITSI as client in this research. Jingle nodes is using the Real Time Transport Protocol (RTP). RPT provides end-to-end network transport functions suitable for applications transmitting real-time data, such as audio, video or simulation data, over multicast or unicast network services [The Internet Engineering Task Force (IETF), http://www.ietf.org/]. The model of manager/supervisor to staff shown in **Fig. 4**.

2.4. Bayesian Filtering for Reducing the Noise of Camera

In this research we improve the ability of camera to be able to recognized face. Because of noise in camera such as failure to detect a face, we use Bayesian filtering to make the system more robust. Bayes filter probabilistically estimate a dynamic system state from noise observations. It represent the state at time t by random variable x_t. Within the Bayesian framework, a probability density of an object's state $x \in S$ is maintained conditioned on the observations z (sensor measurements) obtained. Examples of measurement data include camera images and range scan. If x is a quantity that we would like to infer from y, the probability p(x) will be referred to as prior probability distribution. The Bayesian update formula is applied to determine the new posterior p(x, y) whenever a new observation is obtained Equation 6:

$$p(x, y) = \frac{p(y|x, z)p(x|z)}{p(y|z)} \quad (6)$$

At each point in time, a probability distribution over x_t called belief and denoted bel(x_t), it represents the uncertainty. A belief reflects the robot's internal knowledge about the state of the environment. Probabilistic robility to each possible hypothesis with regards to the true otics represents belief through conditional probability distributions. A belief distribution assigns a probabstate. The sensor data consist of a sequence of time-indexed sensor observations $z_1, z_2,\ldots,z_t$. The belief $bel(x_t)$ then defined by the posterior density over the random variable x_t conditioned on all sensor data available at time t Equation 7 and 8:

$$bel(x_t) = p(x_t \mid z_{1:t}, u_{1:t}) \quad (7)$$

$$= \eta p(z_t \mid x_t)\overline{bel}x_t) \quad (8)$$

This posterior is the probability distribution over the state x_t at time t, conditioned on all past measurements Z_t and all past controller $\mu 1$: t. In this experiment, control data is the status of the recognized face and η is a normalizer in bayes rule variable. To calculate a posterior before incorporating Z_t just after executing the control μ_t Equation 9-11:

$$\overline{bel}(x_t) = p(x_t \mid z_{1:t-1}, u_{1:t}) \quad (9)$$

$$= \int p(x_t \mid u_t, x_{t-1}) bel(x_{t-1}) dx_{t-1} \quad (10)$$

$$= \sum_{x_{t-1}} p(x_t \mid u_t, x_{t-1}) bel(x_{t-1}) \quad (11)$$

This probability distribution is often referred to as prediction in the context of probabilistic filtering (Braunl, 2008). In this context, calculating $bel(x_t)$ from $\overline{bel}(x_t)$ is measurement update that can be used to update the absence of the face or obstacle. The most general algorithm for calculating beliefs is given by the bayes filter algorithm (Thrun *et al*., 2005). This algorithm calculates the belief distribution bel from measurement and control data. The Bayes filter is recursive, the belief $bel(x_t)$ at time t is calculated from the belief $bel(x_{t-1})$ at time t-1. Its input is the belief bel at time t-1 along with the most recent control u_t and the most recent measurement z_t. The proposed method for face detection and tracking is shown in algorithm 1. This method is combination of face recognition and tracking using Bayesian Filtering.

Algorithm 1. Robust Face Detection and Tracking for Telepresence Robot:

```
do
if (face_detected) then ‘ if face detected
call Bayes_filter () ‘update the probability of the face/obstacle
track face ‘track the face
loop
endif
loop
```

functionBayes_filter $(bel(x_{t-1}), u_t, z_t)$
for all x_t do
$\overline{bel}(x_t) = \int p(x_t \mid u_t, x_{t-1}) bel(x_{t-1}) dx_{t-1}$
$bel(x_t) = \eta p(z_t \mid x_t)\overline{bel}(x_t)$
endfor
return $bel(x_t)$

And for autonomous navigation, we propose an algorithm as shown below.

Algorithm 2. Autonomous Navigation:

```
do
        read distance sensors
        start running the robot
        distance estimation using stereo vision and
        Kalman filtering
        if (obstacle = detected) then
          maneuver with opposite direction from obs.
        endif
loop
```

3. RESULTS

Experiments of intelligent telepresence robot have been tested for navigating the robot to the staff in our office as shown in **Fig. 5**. Face tracking and recognition based on eigenspaces with 3 images per person used and databases for the images have been developed (Budiharto *et al*., 2011a). Based on the experiment, the system run very well with the capability to avoid obstacles and people and face tracking.

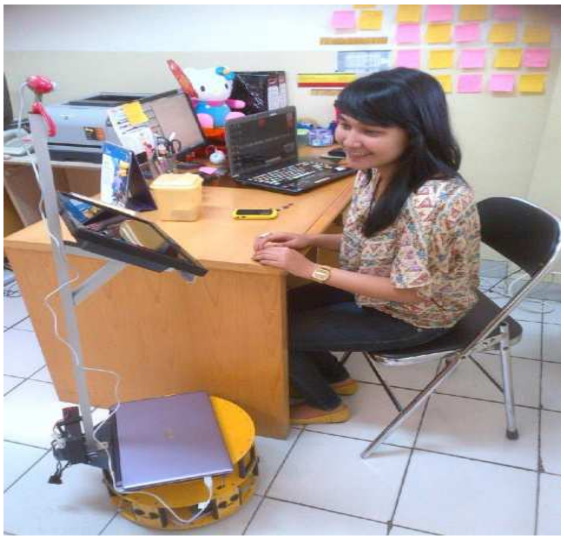

Fig. 5. Teleconference with a client using robot

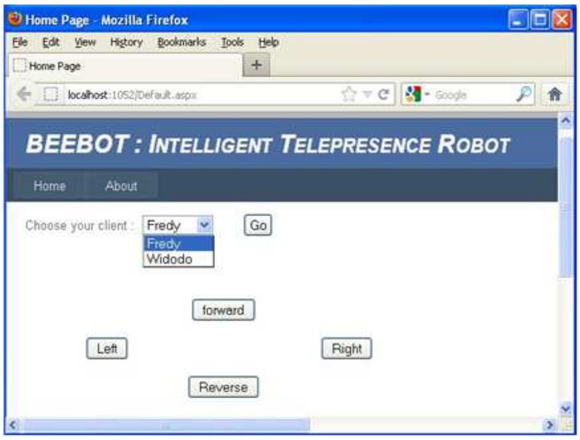

Fig. 6. Web application for choosing a client and direct a robot to a specified person

Fig. 7. JITSI softwares used for teleconference system

Robot controlled using integrated web application that owned by each robot as shown in **Fig. 6** below.

As shown in **Fig. 6** master computer connect to the robot using port 1052 and HTTP protocol. With the high speed internet connection 1 Mbps, the result of video conferencing enough smooth, as shown in **Fig. 7**.

4. DISCUSSION

Camera and audio system take an image and sound from a person before the robot. Therefore, the image is processed through image processing as depicted in **Fig. 2** and is sent through Internet network using HTTP TCP protocol to the master computer. The sound is sent thorough Intenet uses same protocel as the image. The master computer receive the image and sound from the robot then an operator can hear and watch the image. Distance between robot and the person is determined by distance sensor located in the base of the robot.

The navigation of the robot is controlled by the operator using web application that represent the control of the robot. **Figure 6** depicted the control by representing the direction of the robot using four button which labelled as left, right, forward and reverse. The drop-down menu in the controller is purposed for choosing the robot. For example, if operator choose Fredy in the drop-down menu then the controller will connect to Fredy's robot by accessing its IP Address.

5. CONCLUSION

The intelligent telepresence robot with multi-client capability running well with fast movement using omniwheel drive using our method. People tracking system and autonomous navigation successfully developed using stereo vision. This model potentially used by manager/supervisor at office/factory environment. Furthermore, we will develop autonomous navigation to a chosen location (e.g., production area) using stereo vision.

6. ACKNOWLEDGMENT

This researcher funded by grant from Binus University-Jakarta and NUNI for the year 2012.

7. REFERENCES

Alers, S., D. Bloembergen, D. Hennes and K. Tuyls, 2011. Augmented mobile telepresence with assisted control. Maastricht University.

Braunl, T., 2008. Embedded Robotics: Mobile Robot Design and Applications with Embedded Systems. 3rd Edn., Springer, ISBN-10: 3540705333, pp: 542.

Budiharto, W., D. Purwanto and A. Jazidie, 2011b. A navigation system for service robot using stereo vision and kalman filtering. Proceedings of the 11th IEEE International Conference on Control, Automation and Systems, Oct. 26-29, IEEE Xplore Press, Gyeonggi-do, pp: 1771-1776.

Budiharto, W., D. Santoso, A. Jazidie, 2011a. A robust obstacle avoidance for service robot using Bayesian approach. Int. J. Adv. Rob. Syst., 8: 37-44.

Kutinsky, V., 2007. Client-server HTTP communications made easy. SIGAPL APL Quote Quad, 35: 15-20. DOI: 10.1145/1354923.1354925

Lazewatsky, D.A. and W.D. Smart, 2011. An inexpensive robot platform for teleoperation and experimentation. Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), May 9-13, IEEE Xplore Press, Shanghai, pp: 1211-1216. DOI: 10.1109/ICRA.2011.5980230

Quigley, M., B. Gerkey, K. Conley, J. Faust and T. Foote *et al*., 2009. ROS: An open-source robot operating system. Proceedings of the Open-Source Software workshop at the International Conference on Robotics and Automation, (RA' 09), pp: 6-6.

Sheridan, T.B., 1992. Telerobotics, Automation and Human Supervisory Control. 1st Edn., MIT Press, ISBN-10: 0262515474, pp: 415.

Takayama, L., E. Marder-Eppstein, H. Harris and J.M. Beer, 2011. Assisted driving of a mobile remote presence system: System design and controlled user evaluation. Proceedings of the IEEE International Conference on Robotics and Automation, May. 9-13, IEEE Xplore Press, Shanghai, pp: 1883-1889. DOI: 10.1109/ICRA.2011.5979637

Thrun, S., W. Burgard and D. Fox, 2005. Probabilistic Robotics, 1st Edn., The MI Press, ISBN-10: 0262201623, pp: 647.