Human Pose Refinement for Reliable Robotic Teleoperation

Jae-Min Sa, Byung-Hyun Yun, and Kang-Sun Choi*
Interdisciplinary Program on Creative Engineering, KOREATECH, Cheonan, Korea
E-mail: ks.choi@koreatech.ac.kr

Abstract—Three-dimensional human pose estimation can be effectively applied to teleoperation of humanoids, thanks to its intuitive and easy usability. However, the human pose wrongly inferred can cause critical and dangerous situations for robotic operation. In this paper, a reliable human pose estimation method is presented. Initially, 3-D joint positions are inferred by a conventional pose estimation method using stereo images. The joint positions are refined by Kalman filtering followed by median filtering. The motor control value for each joint is smoothed with an IIR filter to perform stable and flexible operations.

I. INTRODUCTION

Historically, numerous types of mobile robots have been developed and employed for various situations to operate on behalf of humans. Although there are research fields devoted to the automatic control of mobile robots [1], most practical robots are still operated by human operators using video images captured by cameras mounted on the robot.

Due to anthropometrical difference between the operator and a teleoperated humanoid, inverse kinematics-based solutions focusing on only hand-like end effector manipulation can yield a different robotic pose from the human pose. Because this pose inconsistency can cause the humanoid to collide with its surroundings and to be broken. This is why the control accuracy is given top priority even at the expense of manipulating speed. However, both the pose inconsistency and the slow manipulating speed prevent the humanoid teleoperation from being widely utilized. Therefore, natural estimation of the operator’s pose and rapid humanoid manipulation by transferring it has been significantly important in the robotic teleoperation.

In this paper, a reliable and stable human pose estimation is proposed to tele-operate a humanoid in real time. The proposed pose estimation and refinement method is depicted in Fig. 1.

Initially, two versions of the operator’s pose are obtained using a conventional 3-D human pose estimation (HPE) method [2] with a pair of stereo images, respectively. Although a pose can be misdetected, another one extracted from a different point of view can maintain a more appropriate pose. A reliable human pose is determined by Kalman filtering followed by median filtering.

Secondary, the angle of each joint with respect to its parent bone is calculated to control the corresponding motor. This angle based motor control allows human pose imitation to the teleoperated humanoid.

Finally, the control angle of the joint motor is temporally stabilized by IIR filtering, that limits rapid motor angle changes.

II. PROPOSED METHOD

In this Section, each step in the proposed method is explained in detail.

A. Pose refinement using stereo data

Two initial 3-D human poses, denoted by \( P_t^{v} \)'s, \( v \in \{L, R\} \), are obtained using the HPE method [2] with input stereo images acquired at time \( t \), respectively. The 3-D position of the \( j \)th joint within \( P_t^{v} \) is denoted by \( P_t^{v,j} \), where the joint index, \( j \in \{\text{neck}, \text{top}, \text{lwri}, \text{lelb}, \text{lsho}, \text{rsho}, \text{relb}, \text{rsho}\} \), is annotated on the pose configuration in Fig. 2(a).

Fig. 2. Correspondences between the skeletal posture and the humanoid LIMS. (a) The pose configuration of eight joints. (b) The motors located in LIMS and the base position indicated as a green dot.

The initial poses determined by the image-based method are sometimes quite inaccurately estimated due to pose ambiguity.

\[\text{This work was supported by the BK21 plus program through the National Research Foundation (NRF) funded by the Ministry of Education of Korea.}\]
In order to correct such a misdetected pose, Kalman filtering is employed per joint.

\[ \hat{p}_t^{i,j} = \text{Kalman}(p_t^{i,j}). \]  

Then, a reliable joint position is determined effectively combining two corresponding candidate positions as follows:

\[ \hat{p}_t^j = \text{MED} [\hat{p}_t^{j,l}, \hat{p}_t^{j,R}], \]  

where \( \text{MED} [\cdot] \) represents the median filter that selects a median value for each component of the input value. The median filtering can remove an outlier causing glitch artifacts in the robotic movement.

B. Motor angle calculation

In this stage is calculate motors angle value for robot control from filtered 3-D joints \( p_t^{M} \). This method is simple calculate of coordinate and angles expressed as follows

\[
\begin{align*}
Q_t &= J2Q(\hat{p}_t) \\
Q_t &= [q_{\phi_1}^L, q_{\phi_2}^L, q_{\phi_3}^L, q_{\phi_4}^L, q_{\phi_1}^R, q_{\phi_2}^R, q_{\phi_3}^R, q_{\phi_4}^R]
\end{align*}
\]  

where \( J2Q(\cdot) \) represents a humanoid-specific \( Q \)-calculation function. For the humanoid LIMS, four motors \( \phi_i, i \in [1 \ldots 4] \), can be actuated using the pose we have. \( \phi_1, \phi_2, \) and \( \phi_3 \) are located on the shoulder and \( \phi_4 \) actuates the elbow as shown in Fig. 2(b).

C. Motor angle stabilization

Most glitch artifacts have already removed through the pose refinement, but \( Q_t \) still has discontinuous values that can result in critical abrupt movement when controlling the robot. Therefore, the motor angles are stabilized using an IIR filter.

\[ Q_t^{\text{IR}} = \alpha Q_t + (1 - \alpha) Q_{t-1}^{\text{IR}}, \]  

where the constant damping coefficient \( \alpha \) is set to 0.05 in our experiments.

III. EXPERIMENTAL RESULTS

In order to evaluate the proposed method, an experiment was simulated using a stereo video. The pair of initial poses were obtained every 12 ms, while the remaining calculations are very efficient. As a result, the entire pose refinement processing can be performed at 80 Hz. It is very fast compared to the conventional teleoperation method\[4\] running at 7.13 Hz.

As shown in Fig. 3, the initial pose obtained from the HPE method represented as blue graph fluctuates severely and includes large glitches. After performing the proposed refinement, however, such the artifacts are significantly alleviated in the refined pose represented in cyan. The final motor angle values stabilized by the IIR filter, indicated in red, are sufficiently smoothed as well as following the motor angle values for the initial pose with a little delay of at most about 40 ms. In Fig. 3, the performance of the proposed method is visualized comparing the refined poses to the initial poses at the time instances of 1190 and 1575. The operator lifted a large box with open arms, but the operator’s arms were not clearly shown that the initial poses were estimated wrongly. Nonetheless, the final poses were successfully corrected and refined by the proposed method.

IV. CONCLUSIONS

The proposed human pose estimation method is quite suitable for robotic teleoperation, since it does not only run much fast compared to the conventional method, but also yield reliable and stable poses even without a depth sensor.

REFERENCE