Proceedings of the 2003 IEEE
International Conference on Robotics & Automation
Taipei, Taiwan, September 14-19, 2003

Enabling Real-time Full-Body Imitation: A Natural Way of Transferring Human Movement to Humanoids

Marcia Riley *† Ales Ude ‡* Keegan Wade §* Christopher G. Atkeson ¶*

Abstract

We seek intuitive, efficient ways to create and direct human-like behaviors for humanoid robots. Here we present a method to enable humanoid robots to acquire movements by imitation. The robot uses 3D vision to perceive the movements of a human teacher, and then estimates the teacher's body postures using a fast full-body inverse kinematics method that incorporates a kinematic model of the teacher. This solution is then mapped to the robot and reproduced in real-time. The robustness of the method is tested on a 30-degree-of-freedom Sarcos humanoid robot located at ATR using 3D vision data from external cameras and from head-mounted cameras.

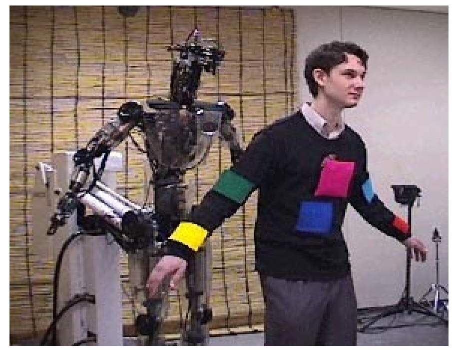

Figure 1: The robot imitating its teacher. The color patches used for vision can be seen on the teacher's shirt.

1 Introduction

We need ways to easily teach humanoid robots new behaviors. Currently, a large amount of work and specialized knowledge are often required to create or modify behaviors. To change this we are building more natural and efficient interfaces between humans and humanoids to communicate new behaviors. For inspiration we look to successful strategies used by people, such as imitation [7], where people observe the behavior of another, and reproduce it.

Here we present our work on enabling a humanoid robot to acquire new movements by observing and reproducing its human teacher's movements in real time. To solve for the teacher's full-body postures quickly, we divide the inverse kinematics(IK) problem into simpler sub-problems. First we use an analytic solution to find the position and orientation of the teacher relative to a world coordinate system. Then we employ a simplified hierarchical version of the iterative numerical solution used in our earlier work [9] to find the joint angles describing the full-body configuration of the teacher. A kinematic model is used to describe the interdependencies of the human skeleton, allowing us to phrase the IK problem in terms of a series of rigid body rotations organized along 6 hierarchically dependent chains. After finding the IK solution, the robot reproduces the perceived configuration by matching the recovered joint angles.

We test two types of stereo vision to track the teacher's movements: an off-the-shelf color vision system, QuickMag (OKK Inc., Japan), using external cameras, and vision provided by the robot's head-mounted cameras via a probablistic tracking framework similar to that described in [13]. Both vision systems provide information at 60 Hz.

2 Related Work

Work on humanoid movement spans both the increasingly active community of humanoid robots, and the well-established 3D animation community. In robotics, several researchers have proposed frameworks where imitation plays a role. Mataric [5] cites

*ATR Human Information Science Laboratories, Japan
†College of Computing, GVU, Georgia Institute of Technology, USA, mriley@cc.gatech.edu
‡Jožef Stefan Institute, Ljubljana, Slovenia, aude@atr.co.jp
§University of Victoria, British Columbia, Canada, kwade@atr.co.jp
¶Robotics Institute, Carnegie Mellon University, USA, cga@cmu.edu

Enabling Real-time Full-Body Imitation: A Natural Way of Transferring Human Movement to Humanoids

Marcia Riley *† Ales Ude ‡* Keegan Wade §* Christopher G. Atkeson ¶*

Abstract

We seek intuitive, efficient ways to create and direct human-like behaviors for humanoid robots. Here we present a method to enable humanoid robots to acquire movements by imitation. The robot uses 3D vision to perceive the movements of a human teacher, and then estimates the teacher's body postures using a fast full-body inverse kinematics method that incorporates a kinematic model of the teacher. This solution is then mapped to the robot and reproduced in real-time. The robustness of the method is tested on a 30-degree-of-freedom Sarcos humanoid robot located at ATR using 3D vision data from external cameras and from head-mounted cameras.

Figure 1: The robot imitating its teacher. The color patches used for vision can be seen on the teacher's shirt.

1 Introduction

We need ways to easily teach humanoid robots new behaviors. Currently, a large amount of work and specialized knowledge are often required to create or modify behaviors. To change this we are building more natural and efficient interfaces between humans and humanoids to communicate new behaviors. For inspiration we look to successful strategies used by people, such as imitation [7], where people observe the behavior of another, and reproduce it.

Here we present our work on enabling a humanoid robot to acquire new movements by observing and reproducing its human teacher's movements in real time. To solve for the teacher's full-body postures quickly, we divide the inverse kinematics(IK) problem into simpler sub-problems. First we use an analytic solution to find the position and orientation of the teacher relative to a world coordinate system. Then we employ a simplified hierarchical version of the iterative numerical solution used in our earlier work [9] to find the joint angles describing the full-body configuration of the teacher. A kinematic model is used to describe the interdependencies of the human skeleton, allowing us to phrase the IK problem in terms of a series of rigid body rotations organized along 6 hierarchically dependent chains. After finding the IK solution, the robot reproduces the perceived configuration by matching the recovered joint angles.

We test two types of stereo vision to track the teacher's movements: an off-the-shelf color vision system, QuickMag (OKK Inc., Japan), using external cameras, and vision provided by the robot's head-mounted cameras via a probablistic tracking framework similar to that described in [13]. Both vision systems provide information at 60 Hz.

*ATR Human Information Science Laboratories, Japan

†College of Computing, GVU, Georgia Institute of Technology, USA, mriley@cc.gatech.edu

‡Jožef Stefan Institute, Ljubljana, Slovenia, aude@atr.co.jp

§University of Victoria, British Columbia, Canada, kwade@atr.co.jp

¶Robotics Institute, Carnegie Mellon University, USA, cga@cmu.edu

2 Related Work

Work on humanoid movement spans both the increasingly active community of humanoid robots, and the well-established 3D animation community. In robotics, several researchers have proposed frameworks where imitation plays a role. Mataric [5] cites human movement as a rich source of non-verbal information, and a powerful tool which we can exploit to learn new skills. She uses primitives to describe movement, and tests her method in a humanoid simulation.

In [2] Breazeal and Scassellati discuss social interaction and skill transfer, including imitation as an effective means for humanoids to acquire new skills. Difficult challenges such as having the humanoid know what to imitate and how to evaluate success are also identified. They show an active vision system for an upper-torso robot, Cog, and an expressive head and face robot, Kismet.

Bakker and Kuniyoshi [1] navigate the terminology and offer a clear definition of imitation which we adopt here: imitation is an agent learning a behavior from observing a teacher executing that behavior. They also describe a 3-part framework to enable imitation: observe the action, represent the action, and reproduce it. They point out that a non-trivial mapping is necessary to make a correspondence between observed action and the execution of a similar action by the robot. Inverse kinematics plays a key role in this mapping. Our implementation is compatible with the framework of Bakker and Kuniyoshi. We observe with 3D vision. To represent what we perceive, we rely on a model of human kinematics to help us solve the inverse kinematics problem for the observed postures. We use this joint angles representation to achieve a similar posture on the robot.

In much previous work, off-line methods are used to reproduce motion for a humanoid robot. In [12] Ude estimates human motion from video using a kinematic model of the human skeleton. The work is interesting, although to-date not achievable in real-time. In our earlier work [9], we reproduce Okinawan folk dance for a humanoid robot with off-line methods starting with motion capture data. We were concerned with scaling the recovered motion for the robot's joint ranges to preserve the style of the motion. In [8] Pollard et al. further explore methods to scale human motion to a humanoid robot while maintaining the individual style of the performer in the final motion.

Cheng and Kuniyoshi [3] use 2D vision with a simple body model to implement real-time humanoid robot imitation where perceived head and arm motion is mapped directly to specific motor outputs in the robot head, neck, torso, and arms. Their decision to use a simple model and 2D vision means that more assumptions must be made in the mapping function from person to robot. Also, representing motion with motor velocities is not as intuitive as joint angles. However, their solution is elegant in that it requires no special markers and is part of an integrated system that can detect the person entering the room and then begin imitation.

Much early work in robotics centered on solutions for robotic arms, thus making end-effector-driven inverse kinematics an important problem. Closed-form solutions were particularly emphasized for industrial robots. More recent end-effector work also includes numerical solutions. Tevatia and Schaal [11] discuss and compare the relationship between the pseudo inverse with explicit optimization and the extended Jacobian methods for the 30-degree-of-freedom (DOF) robot at ATR.

Ude and Atkeson [13] recently describe a system which tracks either a person's face or hand. The motion of the face or hand is then translated to desired end-effector coordinates of the robot's right hand, and the resulting movement is produced by using end-effector inverse kinematics as described in Tevatia and Schaal [11]. This work requires no special color markers and presents a fast, probablistic framework for vision with filtering for noise reduction, but the imitative behavior is limited. Only a simple translation of the perceived movement of hand or face is mapped to the end-effector position.

In the graphics community, Zhao and Badler [14] formulated inverse kinematics as a constrained non-linear optimization problem taking into account a sequential definition of constraints to reconcile the joint chains. End-effector position is still important, but they established that numerical solutions are robust enough to handle non-trivial inverse kinematics problems for complex humanoids, and can achieve sufficient performance for interactive systems.

In Hodgins et al. [4] off-line physical simulation is used to find physically-plausible movement solutions for humanoid athletes. This work is novel in that dynamics as well as kinematics are taken into account to achieve more realistic solutions for graphical characters.

Shin et al. [10] animate graphical characters for a television show by tracking a person's movements in real-time. They begin with magnetic marker positions and orientations, and solve the inverse kinematics problem in three stages while varying the importance of the end-effector position in the solution. Our approach is similar in that we also sub-divide the IK problem to maximize performance. However, vision data restricts us to position data only to begin the motion recovery for a physical system.

3 Real-time Full-Body Imitation

Our approach to imitation begins with 3D vision to perceive the movements of the person acting as a teacher. We use vision because it is a natural interface for human-human interaction, and thus for human-humanoid interaction. We simplify the vision problem by tracking color patches placed on the teacher's clothing (Figs.1 and 2). To make sense of what it sees, the robot has a model of human kinematics. This model consists of 14 rigid parts divided into 6 interconnected kinematic chains: the head; the upper arms, lower arms, and hands; the torso; the upper legs, lower legs and feet. There are 32 DOFs in the model where 6 describe the body's position and orientation in the world, and 26 DOFs describe the body configuration as shown in [9].

The association between the color patches and their initial positions on the model's body at zero configuration is set as follows. We collect vision data of the teacher standing still for 2 seconds. The average perceived position of each patch is then transformed to the model via a straightforward analytic mapping between a triangle on the teacher's body and a triangle on the model's body (Fig.2).

At this time we also determine which DOFs will be active for a particular session. The number of DOFs we can estimate depends upon the number and placement of color patches on the body. For our robot experiments with real vision data, we currently track 6 or 7 patches, and solve for 7 DOFs: 3 torso DOFs (flexion/extension, abduction/adduction, and rotation), and 2 shoulder DOFs (flexion/extension and abduction/adduction) for each shoulder. However, the method described here is intended to estimate all 32 DOFs, provided we have enough vision data. To test this, we use simulated vision data as described below. In the following sections we also describe details of our 3D vision and inverse kinematics solutions, and discuss our results.

3.1 Vision

In our initial experiments we used external cameras and a commercial color tracking system Quickmag to track up to 6 colored patches on the body at 60 Hz. However, the use of a commercial system is cumbersome because it does not allow us to change the way visual processing is done. This makes it difficult to program operations such as vision initialization or error recovery, which can often be done more efficiently if the nature of the task can be taken into considera-

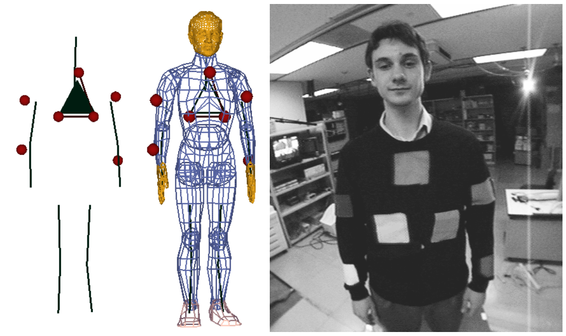

Figure 2: Patch positions are mapped by aligning the solid triangle on the model with one made by 3 patches on the teacher. Patches are shown after mapping as dots on a graphical model represented first with simple lines connecting centroids of body parts, and then with its polygonal body. If patches do not fall on the correct model part, scaling of the model may be necessary. On the right is the teacher as seen through the robot's left eye using head-mounted cameras.

tion.

To make the solution of these problems easier and also to increase the number of color patches that can be tracked, we decided to implement our own PC-based vision system. PCs with modern chipsets enable real-time transfer of full frame color images from two synchronized NTSC cameras to computer memory at 30 Hz (or 60 Hz if the digitizer allows field transfer). In the core of our vision system is a probabilistic tracker that uses color and shape information to find relevant areas in the image. Markers on the body, here color patches, are represented by a number of mutually independent random processes (blobs). We also introduce an additional outlier process that models everything not captured by other processes. Tracking and motion estimation is implemented in a Bayesian framework using an EM (expectation-maximization) algorithm. The system converts the incoming RGB color images into the HSV color space via a lookup table. Only hue and saturation are used for processing. The probabilistic approach together with the conversion to the HSV color space makes our tracker fairly resistant to changes in lighting conditions. Such robustness would be impossible to achieve with more conventional techniques based on straightforward thresholding. Specialized vision initialization and error recovery schemes were also implemented.

A 2 GHz dual processor PC and multithreading techniques were used to achieve stereo tracking of 7 or more color patches on the body at 60 Hz (see Fig.2). The system starts a number of threads dealing with frame/field grabbing, probability calculation, and stereo processing. Techniques such as masking, windowing, and warping were used to speed up processing. The use of wide angles lenses for the robot eyes ensured that the human instructor stays within the view field of the robot even while the robot moves. However, wide angle lenses also introduce significant distortion into our images that need to be dealt with to make the results of stereo triangulation usable. The noise was further reduced using a Kalman filter based on a random jerk kinematic model [13].

3.2 Inverse Kinematics

We solve for the teacher's full-body postures in real-time by breaking the IK problem into simpler subproblems. First we solve the world position and orientation of the teacher relative to the model's coordinate system using an analytic mapping between triangles on the model and the teacher. We then employ an iterative numerical solution to hierarchically solve for the joint angles describing the current body configuration.

Our numerical solution uses twist coordinates to model the kinematics [6]. For a revolute joint, the twist has the form

$$\boldsymbol{\xi}_i = \begin{bmatrix} -\boldsymbol{n}_i \times \boldsymbol{q}_i \\ \boldsymbol{n}_i \end{bmatrix}, \tag{1}$$

where $\boldsymbol{n}_i$ is the unit vector in the direction of the joint axis and $\boldsymbol{q}_i$ is any point on the axis, both given in a global body coordinate system. See [6] for mathematical details.

A patch position in a global body coordinate system is given by $\boldsymbol{Y}_j$. Its 3-D position at body configuration $(\boldsymbol{R}, \boldsymbol{d}, \theta_1, \ldots, \theta_n)$ can be calculated as follows

$$\begin{aligned} \tilde{\boldsymbol{Y}}_j &= \boldsymbol{g}(\boldsymbol{R}, \boldsymbol{d}) \cdot \exp(\theta_{i_1} \boldsymbol{\xi}_{i_1}) \cdot \ldots \cdot \exp(\theta_{i_{n_j}} \boldsymbol{\xi}_{i_{n_j}}) \cdot \boldsymbol{Y}_j \\ &= \boldsymbol{H}_j(\boldsymbol{R}, \boldsymbol{d}, \theta_1, \ldots, \theta_n). \end{aligned} \tag{2}$$

Here exp is the function mapping twists $\boldsymbol{\xi}_i$ representing the body kinematics into rigid body transformations, $\boldsymbol{R}$ and $\boldsymbol{d}$ are the orientation and position of a body coordinate system with respect to the world coordinate system, and $\boldsymbol{g}(\boldsymbol{R}, \boldsymbol{d})$ denotes the homogeneous matrix corresponding to $\boldsymbol{R}$ and $\boldsymbol{d}$. Note that the set of twists affecting the motion of a patch varies according to the identity of the body part to which the patch is attached. Here our world coordinate system is the model coordinate system.

For our purposes, $\boldsymbol{g}(\boldsymbol{R}, \boldsymbol{d})$ are derived analytically and immediately applied to the patch positions. Thus our system simplifies to:

$$\tilde{\boldsymbol{y}}_j = \boldsymbol{h}_j(\theta_1, \ldots, \theta_n). \tag{3}$$

Where $\tilde{\boldsymbol{y}}_j$ (and $\boldsymbol{y}_j$ below) are expressed in world coordinates.

To attain a close match between the observed and estimated joint angles, we wish to minimize the difference between the measured patch positions and the positions estimated by the recovered joint angles for each frame of motion:

$$\begin{aligned} &\tfrac{1}{2} \textstyle\sum_{j=1}^{N} \|\boldsymbol{h}_j(\theta_1(t_k), \ldots, \theta_n(t_k)) - \boldsymbol{y}_j(t_k)\|^2 \\ &= \tfrac{1}{2} \|\boldsymbol{h}(\theta_1(t_k), \ldots, \theta_n(t_k)) - \boldsymbol{y}(t_k)\|^2, \end{aligned} \tag{4}$$

where $\boldsymbol{y}_j(t_k)$ denotes the j-th measured patch at time t_k expressed in world coordinates, $\boldsymbol{h} = [\boldsymbol{h}_1^T, \ldots, \boldsymbol{h}_N^T]^T$ and $\boldsymbol{y}(t_k) = [\boldsymbol{y}_1(t_k)^T, \ldots, \boldsymbol{y}_N(t_k)^T]^T$.

Here $\theta_i(t_k)$ denotes the current estimate for the teacher's joint angles. This resulted in the following overconstrained system of linear equations that must be solved at each iteration step

$$\begin{aligned} &\boldsymbol{J} \cdot [\Delta\theta_1, \ldots, \Delta\theta_n]^T = \\ &\boldsymbol{y}(t_k) - \boldsymbol{h}(\theta_1(t_k), \ldots, \theta_n(t_k)), \end{aligned} \tag{5}$$

where $\boldsymbol{J}$ is the Jacobian of $\boldsymbol{h}$ at $(\theta_1(t_k), \ldots, \theta_n(t_k))$.

The complexity of the Jacobian is reduced by using an analytic solution for the world position and orientation, thus removing it from the optimization problem. Also, the numerical solution is solved hierarchically, exploiting the dependencies of the body geometry to reduce the amount of work needed to find a good solution. For example, the torso DOFs are found first, and this partial solution is then used in solving for the arm and head DOFs. Both of these factors significantly improve the speed of the IK solution, as compared with our earlier work [9], in which all parameters including joint angles and world position and orientation were found simultaneously by solving one optimization problem.

3.3 Results and Robustness

We tested the above methods on a 30-DOF humanoid robot tracking 6 color patches with external cameras

Test	Condition	L Arm FE	L Arm AA	R Arm FE	R Arm AA	R Elbow FE	Torso Rot	Torso AA	Torso FE
Model	Noise 1	0.470°	0.211°	1.25°	0.792°	1.93°	0.070°	0.026°	0.099°
Noise	Noise 2	0.495°	0.361°	3.29°	0.158°	2.88°	0.024°	0.011°	0.011°
Vision	Noise 1	0.718°	0.423°	1.05°	0.464°	2.90°	0.205°	0.161°	0.110°
Noise	Noise 2	0.735°	0.780°	4.32°	1.10°	8.97°	0.231°	0.118°	0.144°
Both	Noise 1	0.878°	1.39°	2.34°	1.45°	5.59°	0.760°	0.517°	0.274°
	Noise 2	0.906°	1.36°	4.84°	1.06°	11.68°	0.653°	0.451°	0.416°

Table 1: Mean absolute errors of joint angles recovered from simulation (AA = abduction/aduction, FE = flexion/extension, Rot = Rotation).

(Fig.3, first row), and 7 color patches with the head-mounted cameras (Fig.3, second row). In both cases we ran several sessions which lasted about 3 minutes each, and solved for 7 DOFs with vision data supplied at 60 Hz. Sessions were ended by the teacher, and can be extended for longer periods. The numerical solution used to find the joint angles converged in an average of 3.7 iterations per frame. This plus the fast analytic step to find the teacher-model correspondence accounts for the total solution per frame.

Our method makes the assumption that the last solution is a good place to start to solve the next frame. This assumption remains true only if the motion velocities are not too large with respect to our tracking frequency. To ensure this is true, we use a movement threshold $\boldsymbol{m}$ as follows:

$$\boldsymbol{m} = \sqrt{\sum_{i=1}^{N} d\theta_i^2} \tag{6}$$

where $\boldsymbol{d\theta}_i$ is the change in joint angle for a degree of freedom since the previous vision frame, and is set to 10 degrees for all N degrees of freedom used in the current solution. Movement exceeding this is considered too fast for the robot to track. In this case the robot signals the teacher it is lost in one of 2 ways: the robot stays in the last known position waiting for the human to re-direct behavior from that point, or the robot returns to its initial zero configuration, standing with arms and legs straight, and again waits for the human to start from this point. As the human goes to that point, the robot begins tracking again. In our tests this worked well. We believe this to be a natural way to signal a teacher that she is going too fast, as people also stop when they get lost tracking a teacher in settings such as exercise and dance classes.

If the movement threshold is not exceeded for a particular iteration, the robot immediately moves to imitate the joint angles, setting desired joint velocities and accelerations by finite differencing. This configuration is then used as the initial solution for the next iteration.

Although our method has limitations inherent to numerical methods, that is, we are not guaranteed convergence if the step size between subsequent frames is too large, even in analytic solutions the velocity may exceed the robot's capabilities. In either case, we must check for excessive movement velocity, and signal the teacher to slow down.

Presently, we are working on ways to extend our 3D vision tracking beyond 7 patches. In the meantime, we tested our method for higher numbers of DOFs by using simulated vision data. In one experiment, we used 31 patches to recover all 32 DOFs in our model. 26 DOFs representing body joint angles were found by the numerical method described in Section 3.2, and converged with an average of 3.9 iterations per frame with an error threshold of 0.01 radians. Our previous method where one optimization problem was used to solve for all DOFs took an average of 25 iterations to solve for 25 DOFs (19 joint angles and 6 DOFs describing the world position and orientation) for the same error threshold. The simulated vision data had no noise, and was compared with data from an Optotrak, a more accurate optical motion capture system.

The accuracy of the recovered joint angles depends primarily on model geometry and vision accuracy. To test how these errors affect our method, we use simulated data as follows. We create a trajectory of joint angles for 8 DOFs and use it to generate seven 3D points (or patches) placed on a 3D humanoid body. The trajectory has 147 frames of data.

In the first test, we change the rigid body representing the teacher by adding random noise to the initial patch positions. This same noise is then added to each frame of data (Test 1), creating an inexact match between the teacher and model geometry. In the next test, we

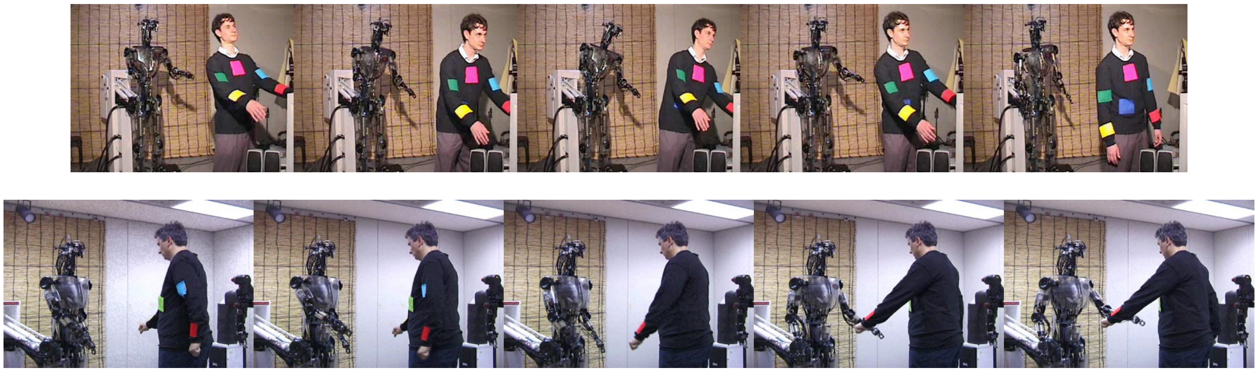

Figure 3: The top row shows the robot using external cameras to imitate from 6 color patches. Frames are 1.33 seconds apart. In the bottom row the robot uses head-mounted cameras to imitate from 7 color patches. Frames are 0.33 seconds apart.

simulate noisy vision data by adding different random amounts of noise to each position in each frame (Test 2). Finally, we combine the first two tests by adding different amounts of noise to each point of each modified frame created in test 1, simulating both a different teacher geometry and noisy vision data (Test 3).

All three tests are run under two conditions: the first condition corresponds to added noise ranging from 0 to 1.6 centimeters in each position dimension; the second condition corresponds to added noise ranging from 0 to 2.4 centimeters. Results are obtained by measuring the mean absolute error between the recovered joint angles and the known solution for each DOF (Table 1).

In comparing tests 1 and 2, we see that model inaccuracy has a slightly smaller effect than noisy vision on the mean error for most DOFs in these trajectories. However, different model configurations may induce different types and amounts of error. What is more interesting is that in all cases these tests highlight which DOFs become harder to estimate under imperfect conditions. Here the right elbow is most vulnerable to error. This is not surprising, as the elbow joint angle estimation relies on a single data point on the lower arm. In contrast, the torso is very stable, with its joint angles being estimated from 3 data points on the chest. However, if significant torso error is present it may induce errors in the arm estimations.

Besides noisy vision and an inexact match between the model and the teacher's geometry, other sources of error include simplification of true human kinematics, assumptions about the rigidity of the triangle used to recover correspondence between the model and the teacher's coordinate systems, and the possibility of an inexact mapping of joint angles from the teacher to the robot due to different joint angle limits. Further, when using the head-mounted cameras, recovery of information has additional noise introduced from body oscillations affecting the head. However, even with these problems, we can still extract enough relevant information to enable imitative behavior. Movies of our results can be seen at http://www.his.atr.co.jp/~mriley/imitation.html.

4 Summary and Conclusions

We are exploring intuitive and efficient approaches for humanoids to acquire new behaviors. Several schemes have been proposed in the literature where imitation plays a key role. Many good frameworks exist, although few have yet been implemented for full-body imitation for robots. Here we present our work to implement and test such a method on a complex humanoid robot. Methods such as these are needed because humanoid robots will play an increasing role in our future.

Our work differs from much previous robotics work in that we emphasize full-body postures rather than the position of an end effector alone. We use 3D vision and a complex, yet still approximate, kinematic body model to enable real-time imitative behavior for a 30-DOF humanoid robot. We solve the inverse kinematics problem quickly by breaking it into sub-problems. This yields a joint angle representation of body postures, which is a more intuitive representation than motor velocities or forces, and one which facilitates further interaction. Finally, we test our method with real vision data from both external and head-mounted cameras for 7 DOFs, and with simulated vision data for up to 32 DOFs.

In future work, we wish to direct humanoid behaviors interactively through coaching with the ability to intervene and immediately modify behaviors as desired. Our work on humanoid imitation is the start of a low-level toolkit which will enable such higher-level interactions between people and robots.

Acknowledgments

This work was supported in part by the CRL Keihanna Human Info-Communication Research Center.

References

[1] P.Bakker and Y.Kuniyoshi, "Robot See, Robot Do: An Overview of Robot Imitation", *AISB'96 Workshop on Learning in Robots and Animals*, 1996.

[2] C. Breazeal and B. Scassellati, "Challenges in Building Robots That Imitate People", in *Imitation in Animals and Artifacts*, Kerstin Dautenhahn and Chrystopher Nehaniv, eds., MIT Press, 2000.

[3] G. Cheng and Y. Kuniyoshi, "Real-Time Mimicking of Human Body Motion by a Humanoid Robot", *The 6th International Conference on Intelligent Autonomous Systems*, Venice, Italy, July 2000, pp. 273-280.

[4] J.K. Hodgins, W.L. Wooten, D.C. Brogan, J.F. O'Brien, "Animating Human Athletics", *Proceedings of ACM SIGGRAPH 95*, August 1995, pp 71-78.

[5] Maja J. Mataric, "Getting Humanoids to Move and Imitate", *IEEE Intelligent Systems*, August 2000, Vol.15, no.4, pp.18-24.

[6] R. M. Murray, Z. Li, and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation*, CRC Press, Boca Raton, New York, 1994.

[7] J.Piaget, 1945. *Play, Dreams and Imitation in Childhood*, 1945, translated 1962, W. W. Norton, New York.

[8] N.S. Pollard, J. K. Hodgins, M.J. Riley, C. G. Atkeson, "Adapting Human Motion for the Control of a Humanoid Robot", *IEEE-RAS Conference on Robotics and Automation*, May 2002.

[9] M.Riley, A.Ude, and C.G. Atkeson, "Methods for Motion Generation and Interaction with a Humanoid Robot: Case Studies of Dancing and Catching", *AAAI and CMU Workshop on Interactive Robotics and Entertainment 2000*, April, 2000, pp.35-42.

[10] Hyun Joon Shin, Jehee Lee, Michael Gleicher, and Sung Yong Shin, "Computer puppetry: An importance-based approach", *ACM Transactions on Graphics*, April 2001.

[11] G. Tevatia and S. Schaal, "Inverse Kinematics for Humanoid Robots", *IEEE International Conference on Robotics and Automation*, April 2000, pp.294-299.

[12] A.Ude, "Robust Estimation of Human Body Kinematics from Video", *IEEE/RSJ Conference on Intelligent Robots and Systems*, October 1999, pp.1489-1494.

[13] A.Ude and C.G. Atkeson, "On line tracking and mimicking of human movements by a humanoid robot", *Journal of Advanced Robotics*, 2003, Vol.17, no.2.

[14] J. Zhao and N.I. Badler, "Inverse kinematics positioning using nonlinear programming for highly articulated figures", *ACM Transactions on Graphics*, October 1994.