Visual Based Human Motion Analysis: Mapping Gestures Using a Puppet Model

Jörg Rett and Jorge Dias

Institute of Systems and Robotics,
University of Coimbra, Polo II, 3030-290 Coimbra, Portugal
{jrett, jorge}@isr.uc.pt

Abstract. This paper presents a novel approach to analyze the appearance of human motions with a simple model i.e. mapping the motions using a virtual marionette model. The approach is based on a robot using a monocular camera to recognize the person interacting with the robot and start tracking its head and hands. We reconstruct 3-D trajectories from 2-D image space (IS) by calibrating and fusing the camera images with data from an inertial sensor, applying general anthropometric data and restricting the motions to lie on a plane. Through a virtual marionette model we map 3-D trajectories to a feature vector in the *marionette control space (MCS)*. This implies inversely that now a certain set of 3-D motions can be performed by the (virtual) marionette system. A subset of these motions are considered to convey information (i.e. gestures). Thus, we are aiming to build up a database which keeps the vocabulary of gestures represented as signals in the *MCS*. The main contribution of this work is the computational model of the *IS-MCS-Mapping*. We introduce the guide robot "Nicole" to place our system in an embodied context. We sketch two novel approaches to represent human motion (i.e. Marionette Space and Labananalysis). We define a gesture vocabulary organized in three sets (i.e. Cohens Gesture Lexicon, Pointing Gestures and Other Gestures).

1 Introduction

Robotics field is facing the challenge to develop robots that share an environment with humans. The two basic skills social robots need to have is to interact with the people and to navigate in the world. To study possible solutions and feasible techniques we started the development of the robot guide Nicole. Nicole will guide visitors through the Institute of Systems and Robotics (ISR), talk about the research and react on gestures performed by persons recognized as "godfathers". The interaction part as well as the navigation part will strongly rely on visual cues. This paper is concerned with *robot vision for human-machine-interaction* of "Nicole".

If the perceptual system of a robot is based on vision, interaction will involve *visual human motion analysis*. The ability to recognize humans and their activities by vision is key for a machine to interact intelligently and effortlessly with

C. Bento, A. Cardoso, and G. Dias (Eds.): EPIA 2005, LNAI 3808, pp. 398–409, 2005.

a human-inhabited environment [1]. Several surveys on visual analysis of human movement have already presented a general framework to tackle this problem [2], [1], [3] and [4]. Aggarwal and Cai point out in their survey [2] that one (of three) mayor areas related to the interpretation of human motion is motion analysis of the human body structure involving human body parts. The general framework consists of: 1. feature extraction, 2. feature correspondence and 3. high level processing. The architecture we present relates to this framework in that we define: 1. Perception- , 2a. Motor/Model-, 2b. Impression- and 3. Interpretation Level. As shown in fig. 5 e.g. the body part segmentation will be found in the *Perception Level* being part of a *Human Tracking Module*.

Research on human behavior suggest, that infants could compare the sensory information from his own unseen motor behavior to a *supramodal* representation of the visually perceived gesture and construct the match required [5,6]. In imitating, infants attempt to match the organ relations they see exhibited by the adults with those they feel themselves make [5]. Infants draw information from what they see by matching it to what they do. We like to further describe this process by proposing to simulate the motion through acting on a model inside our head and interpreting the (virtual) sensor signals. We name our concept: "*The Marionette in the Head*".

This article is about the model we use to generate human motions and the signals we extract to interpret a certain set of human gestures. To materialize the solution we contribute the mathematical concept to build a mapping between the 2-D image space (IS) of a monocular camera and the space of the signals for gesture interpretation (MCS). We introduce the guide robot "Nicole" to place our system in an embodied context. We sketch two novel approaches to represent human motion (i.e. Marionette Space and Labananalysis). We define a gesture vocabulary organized in three sets (i.e. Cohens Gesture Lexicon, Pointing Gestures and Other Gestures).

An interesting research on gesture recognition provides us with the first set of our vocabulary. In [7] Cohen et al. established a lexicon of 24 gestures which were captured by a human moving a flashlight against a black background. In our approach we are detecting and tracking the hands and the face automatically without using special device (markers). The second set of gestures was inspired by Kahn et al. [8] whose interface interprets pointing gestures. Similar to us they were using multiple cues to track the persons hand and heads. Our approach also incorporates face recognition for a personalized interaction.

Section 2, where we present our model, starts the extraction of the required 3-D data from the 2-D image and introduces entities of reference in the 3-D world. In the next part we introduce a vocabulary of gestures and relate simple commands for a mobile robot to it. In the third part we develop the puppet model and present the signals that can be extracted. Sect. 3 shows the implementation of our concept starting with a brief overview of the "Nicole" Roboticsystem, followed by the Gesture Perception system, the Visual-Inertial Sensor and the Human Tracking Module. Sect. 4 presents results on recorded gesture trajectories and Sec. 5 closes with a discussion and an outlook for future works.

2 Models

We have constrained the situation of interaction in the following way. The vision system is calibrated for a person acting at a certain distance and orientation towards the robot. In the current level of development the person needs to adopt this initial position to interact by himself. The interaction will start when the person is facing the robot and standing in a natural "at ease" position. The camera system will cover the entire "kinesphere" of the person while the person performs gestures using his Hands and Face. The position of the person's body is assumed to be static. Using our system in a situation as shown in Fig. 1 a) a motion of hands and face need to be tracked and transformed to what we will call "Marionette Space".

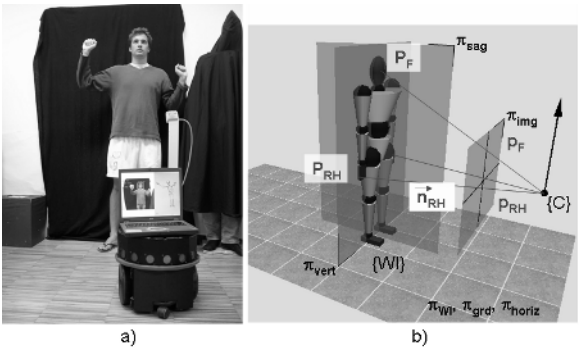

Fig. 1. a) Nicole in position to interact with Enguerran. b) Projection of 3-D point P.

2.1 Projection Space

The first step is the recovery of 3-D trajectories from 2-D images created by a projective camera. We start by defining the initial plane $\pi_{\{\mathcal{WI}\}}$ and relate it to the ground plane and the horizontal plane by $\pi_{\{\mathcal{WI}\}} = \pi_{horz} = \pi_{grd}$ (see Fig. 1 b)). We place the reference frame $\{\mathcal{WI}\}$ at the point of intersection of the vertical body plane π_{vert}, the sagittal plane π_{sag} and the ground plane π_{grd} shown in Fig. 1 b).

Any generic 3-D point $\mathbf{P} = [X\ Y\ Z]^\top$ and its corresponding projection $\mathbf{p} = [u\ v]^\top$ on an image-plane can be mathematically related using projective geometry and the concept of homogeneous coordinates through the following equation, the projective camera relation, where s represents an arbitrary scale factor [9]:

$$\begin{bmatrix} sv \\ su \\ s \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

Matrix $\mathbf{A}$ is called the projection matrix, and through its estimation it is possible to make the correspondence between any 3-D point and its projection

in a camera's image-plane. We can likewise express the matrix $\mathbf{A}$ by using the parameters of the projective finite camera model, as stated in [10].

$$\mathbf{A} = \mathbf{C}\left[{}^{\{\mathcal{C}\}}\mathbf{R}_{\{\mathcal{WI}\}}\ {}^{\{\mathcal{C}\}}\overrightarrow{\mathbf{t}}_{\{\mathcal{WI}\}}\right] \tag{2}$$

Where $\mathbf{C}$ is the camera's calibration matrix, more frequently known as the intrinsic parameters matrix, while the camera's extrinsic parameters are represented by the rotation orthogonal matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$ that relates the chosen $\{\mathcal{WI}\}$ to the camera frame.

The projective camera presents us, in fact, with the solution for the intersection of planes Π_{cam1} and Π_{cam2} which, assuming $\tilde{\mathbf{P}} = [X\ Y\ Z\ 1]^\top$ (i.e. homogeneous coordinates), can be proven from its projection expression to be given by 3) (see [9]).

$$\begin{cases} (\mathbf{a}_1 - u\mathbf{a}_3)^\top \mathbf{P} + a_{1,4} - u = 0 \\ (\mathbf{a}_2 - u\mathbf{a}_3)^\top \mathbf{P} + a_{2,4} - u = 0 \end{cases} \Longleftrightarrow \begin{cases} \mathbf{\Pi}_{cam1}\tilde{\mathbf{P}} = 0 \\ \mathbf{\Pi}_{cam2}\tilde{\mathbf{P}} = 0 \end{cases} \tag{3}$$

This solution is called the projection or projecting line, which can be alternatively represented by equation (4) [9].

$$\overrightarrow{\mathbf{n}} = (\mathbf{a}_1 - u\mathbf{a}_3) \times (\mathbf{a}_2 - u\mathbf{a}_3) \tag{4}$$

These relations indicate that all 3-D points on the projecting line correspond to the same projection point on the image-plane, which means that the projection equation is not unique. Thus, at least one additional restriction is needed to establish an unique correspondence between the 3D point and its projection on the image-plane. One possibility being restricting the locus of 3-D points to lie on a plane.

2.2 Gestures and Labananalysis

In our search for a suitable description of human motions we found the *Labananalysis*, named after the founder R. Laban [11]. In Labananalysis the kinematic chains are observed with relation to spatial shaping possibilities and the dynamic qualities (*Effort*) accompanying them. A pioneer in the attempt to re-formulate Labanotation in computational models is Norman Badler and his early works are summarized in his book on simulating humans [12]. He suggests to not implement Labanotation directly but use it as a good set of default values for normal human movements. More recently a computational model of gesture acquisition and synthesis to learn motion qualities from live performance has been proposed in [13].

We will investigate more the qualities of a gesture while trying to add an *Impression Level* to our system. For now we only like to address the problem of space or gesture plane. An interesting spatial concept is that of *Scales*. *Scales* are movement possibilities with reference to geometric shapes and sequences. *Scales* are related to the *kinesphere* which is defined as the reach space of the body. The simplest *Scale* is called 1-D (or defense) *Scale*. It is built around

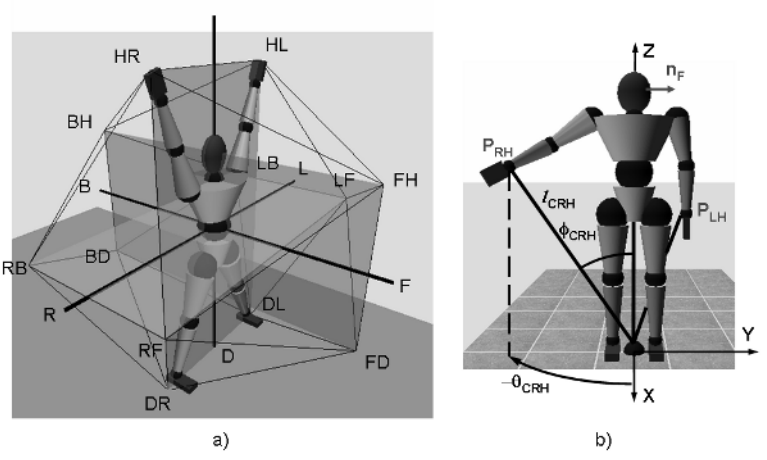

Fig. 2. Laban Scale: Icosahedron

the axes of the kinesphere (vertical, horizontal, sagittal). Figure 2 a) shows the axes and defined points (D = Deep, R = Right, L = Left, F = Front, B = Back, H = High). A 2-D *Scale* is created if the movement to six peripheral point are performed without returning to the center (e.g. a cycle around the three planes π_{vert}, π_{horz}, π_{sag}). As a suitable description for 3-D movements Laban used a Icosahedron see Fig. 2 a). The geometry of the Icosahedron can be developed from the three planes (π_{vert}, π_{horz}, π_{sag}) superimposed and their corners connected. Thus, twelve corners define maximal reach possibilities within the kinesphere. As *Scales* are ordered sequences for the most economical and expressive pathways between all the peripheral points (corners), Laban defined several different *Scales* for 3-D movements. Of particular interest is the one which goes along the outer edges of the icosahedron. Laban saw this *Scale* in many communicative gestures and dance forms, thus calling it *primary Scale*. This reflects that, although the sequences have been outlined as primarily total body movements, they are also identifiable in small movements.

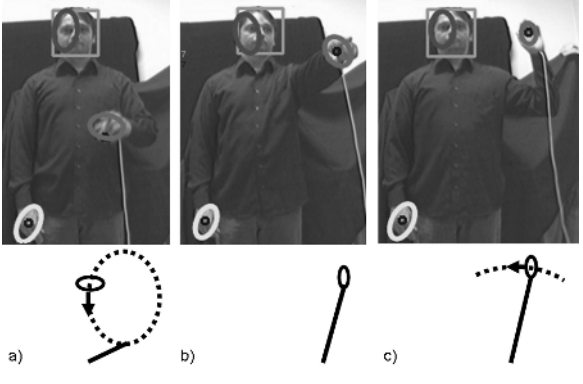

Fig. 3. a) Cohen's Gesture Lexicon. b) Pointing Gestures. c) Other Gestures.

To show the feasibility of our approach we have defined a vocabulary of gestures and recorded a group of people performing them (see fig. 3). Cohen et al. have already presented in [7] a gesture lexicon consisting of 24 planar oscillators to control an actuated mechanism. We use his lexicon as Set 1 and extend it by Pointing Gestures (Set 2) and Other Gestures (Set 3). The latter were gestures expressing information like "Speak louder!", "Be quiet!", "I am hungry!" and "Bye, bye!". We will show later in this article that Set 1 can be described by projecting the trajectory on a plane $\mathbf{\Pi}_{gest}\tilde{\mathbf{P}} = 0$ parallel to π_{vert}. Using the geometry involved in the perspective projection of the world onto the camera's image-plane and the gesture-plane's restriction, the 3D point in the scene can be uniquely related to its 2-D projection point in the image-plane of the camera using (5):

$$\left\{\mathbf{\Pi}_{cam1}\tilde{\mathbf{P}} = 0 \;\; \mathbf{\Pi}_{cam2}\tilde{\mathbf{P}} = 0 \;\; \mathbf{\Pi}_{gest}\tilde{\mathbf{P}} = 0\right\} \tag{5}$$

Our vocabulary of gestures also determines the future interaction of Nicole her "godfathers". Set 1 will be used as command primitives (e.g. turn left, move back), Set 2 to shift the focus of attention (e.g. look northwest) and Set 3 for any other form of communication (e.g. speak louder).

2.3 Marionette Space

There are some examples for the attention puppetry receives from the research community. The approaches involving marionettes are basically placed in the area of entertainment. Generally marionette figures are articulated by a set of servomotors to produce human-like motions. An early work was reported by Hoffmann [14] who used a human dancer to teach the coarse movements to the system. In [15] the marionette was used to produce gestures by superposition, inhibition and sequencing of motor primitives. The work was based on evidences for basic (innate) elementary neural motor programs from which all bodily movements are constructed [16]. The system was further developed in [17] which also gives a nice cultural summary on marionettes. Often human motion capture data is mapped to the marionette while dealing with inverse kinematics and physical constraints. In [18] the results compare the performance of two human actors and the marionette telling a story. Apart from applications in entertainment we also found comparisons of natural (human) and robot actuator systems. [19] presents a control strategy for stable movement of a marionette under a system of unidirectional muscle-like actuators. Analogies to monotonic function of the firing rate of natural muscles were drawn. The underlying question common to all contributions is: "What is the relationship between human and puppet movements?" Our answer to this questions is a model that synthesizes human movements by controlling a virtual puppet. The control vector associated to a certain gesture will be used later for gesture recognition. Our primary interest lies in the reduction of the parameters to describe the human motion i.e. to reduce the dimensionality of the parameterspace. Our secondary goal is to maintain an intuitive approach which can also be understood by non-engineers [20].

From the various types of puppets the *marionettes* have received the most scientific attention so far. We found are more promising concept in the rod puppets. The puppet hands are manipulated using (rigid) sticks see fig. 2. We will first invent a model of the puppet body considering a 3 DoF neck joint, 3 DoF shoulder joints, 1 DoF elbow joints and 3 DoF wrist joints. Next we will place three control joints at the origin of $\{\mathcal{WI}\}$ and connect them with sticks with the hand and the head. We connect the control joints by rigid control links to the wrist joints and neck joint (see fig. 2 b). The hands control joints will have two rotational and 1 translational DoF while the face control joint only has 1 rotational DoF. Thus, we have created a system with a 7-dimensional control space that is able to synthesize a certain set of movements in 3-D space. We are now able to express the relationship between the 3-D space and the control space. We establish a feature vector $\mathbf{F}$ consisting of the face normal (gaze) $\mathbf{n}_F(t)$ and the positions of the hands $\mathbf{P}_{RH}(t)$ and $\mathbf{P}_{LH}(t)$.

$$\mathbf{F}(t) = \begin{bmatrix} \mathbf{n}_F(t) \\ \mathbf{P}_{RH}(t) \\ \mathbf{P}_{LH}(t) \end{bmatrix} \tag{6}$$

We can express the components of the vector by using spherical coordinates. Omitting the dependence on (t) for the moment we get

$$\mathbf{n}_F = \begin{bmatrix} \cos\theta_{CF} \\ \sin\theta_{CF} \\ 0 \end{bmatrix} \text{ and } \quad \mathbf{P}_{RH} = \begin{bmatrix} l_{CRH}\ \cos\theta_{CRH}\ \sin\phi_{CRH} \\ l_{CRH}\ \sin\theta_{CRH}\ \sin\phi_{CRH} \\ l_{CRH}\ \cos\phi_{CRH} \end{bmatrix} \tag{7}$$

for $\mathbf{n}_F$ and $\mathbf{P}_{RH}$ the expression for $\mathbf{P}_{LH}$ goes accordingly. To make bimanual movements [21] more obvious we will count the azimuthal angle θ counterclockwise from the positive x-axis with $-\pi < \theta \leq \pi$.

Representing the human motion in such a way is very close to proposals made by researchers from physiology. In [22] Soechting and Flanders discuss how spatial parameters may be represented by the activities of neurons. They considered different motor tasks like postural responses, orienting movements and arm movements to a spatial target. Introducing frames of reference and coordinate systems they show that in all three motor tasks, one of the coordinate axes was defined by the gravitational vertical. Another coordinate was defined by the sagittal horizontal axis. They suggest that there is a common, earth-fixed frame of reference utilized for all motor tasks. For postural responses findings in research on human (and animal) motion suggests that bipedal posture can be described using limb angles and length from the center of gravity to the base of support.

3 Implementation

3.1 Architecture

As mentioned in section 1 the whole system of the robot "Nicole" needs to deal with navigation as well as interaction. The base of Nicole is a Nomad Scout

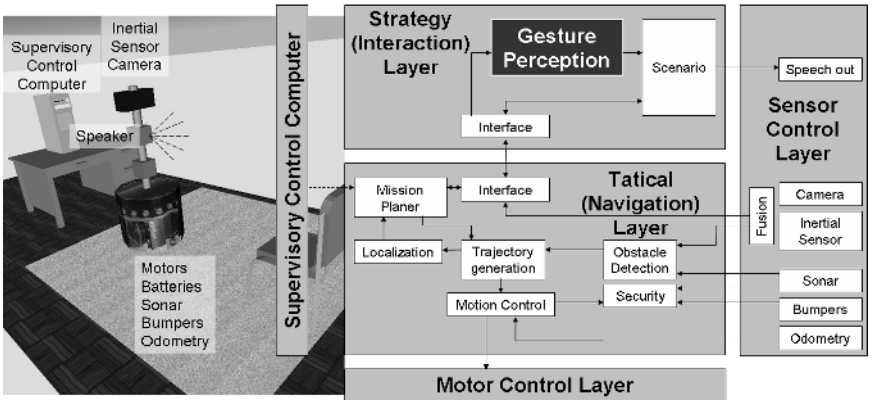

Fig. 4. a) Hardware architecture b) System Architecture

robot (see fig. 4 a)). The Motor Board Controller connects to the sonar, the bumpers and the odometry. With this data and the input from the camera and the inertial sensor the Navigation part will perform obstacle avoidance, global path planning and people tracking. The interaction part will also use the camera and the inertial sensor but additionally the loudspeaker as an output device. It will perform gesture and face recognition and use a speech synthesizer for speaking. The external hardware is supervisory control computer which is connected to the Navigation part via WLan to have the option to control Nicole and visualize her position. Fig. 4 b) shows the architecture of Nicole. In this paper we focus on the *Gesture Perception (GP)-System* which is part of the interaction layer.

3.2 The Gesture Perception (GP)-System

Fig. 5 shows the architecture of the GP-System. The system can be divided in six levels of visual perception and understanding. The Processing starts at the Perception Level with the Visual-Inertial Sensor dealing with Image Capture and Inertial Data registration. The image data is used by the Human-Tracking-Module to perform Face Detection, Face Recognition, Skin Color Detection and Object Tracking. The Projection-Module reconstructs from the 2D Image trajectory of Hands and Face the 3D trajectory. The output of the Perception Level will be used by the Motor/Model Level and Impression Level. The former transforming the trajectory to a feature vector in the *marionette control space (MCS)* the latter using the Labananlysis to create qualities related to *Effort*. The Interpretation Level will use both inputs to perform a *Emotional Tinted Gesture Recognition*. In the final step a Learning Level will refine the emotional and personal gesture vocabulary inside the database of the Knowledge level.

3.3 The Visual-Inertial Sensor

Again, we want to follow our belief that a successful perception of human motion is achieved best when the system is built in human manner. The inner ear

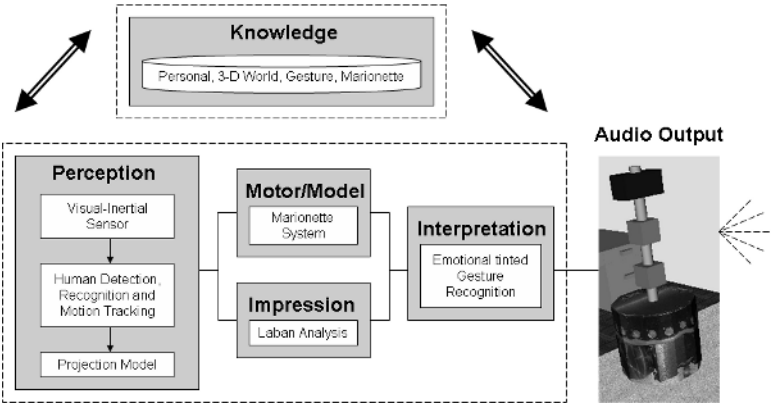

Fig. 5. Architecture of the GP-System

vestibular system in humans and in animals provides inertial sensing mainly for orientation, navigation, control of body posture and equilibrium. This sensorial system also plays a key role in several visual tasks and head stabilization, such as gaze holding and tracking visual movements [23]. Continuing the argumentation from section 2.2 about coordinate systems for human motion, Soechting and Flanders show in [22] that one of the coordinate axes was defined by the gravitational vertical. They point out the domination of gravitational force and the visual horizon and the primary role of the vestibular system as an indicator of the vertical direction.

From the practical point of view we need to establish a stable frame of reference (i.e. $\{\mathcal{WI}\}$) to infer the correct spatial trajectory. This includes also the correct initialization of the human body posture using hands and face position, anthropometric data and projective geometry. Assuming a person to start his interaction in a vertical body pose turns out to search and register the *gravity* in the image.

Recent work of Lobo and Dias present the successful integration and calibration of visual and inertial data [24] and the detection of vertical features [25]. When the system is not accelerating, gravity provides a vertical reference for the camera system frame of reference given by the sensed acceleration.

3.4 The Human Tracking Module

In brief our *Human Tracking Module* takes the images from the *Visual-Inertial Sensor* and creates three image trajectories from the head and both hands. As shown in fig. 5 the module contains of four mayor parts. The process starts with the detection if any human is present in the scene. We use a face detection module based on haar-like features as described in [26]]. In case a face is detected we try to recognize the person if belonging to the group of "godfathers" or not. This second part is based on eigen-objects and PCA as described in [27]. If the persons is identified as a "godfather". The third part will establish the communication by activating the skin color detection an the tracking of the hands. For the skin

detection and segmentation we use the CAMshift algorithm presented in [28]. To deal with hands and head occlusion we predict the positions and velocities based on a Kalman-filter [29].

4 Results

Figure 6 compares the tracking results from our *Human Tracking Module* (left) with a 3D magnetic tracker (miniBird, right). In the diagrams the image coordinate system is placed next to the ZY-plane of the miniBird. The ZY-plane of the miniBird reflects a projection of the motion trajectory on the vertical plane π_{vert} (see fig. 2). It can be seen, that the trajectories from our human tracking module fits well with the ground truth data from miniBird. Furthermore the gestures of Set 1 can be well distinguished from each other and due to their repetitive character a robust recognition should be possible. Though, our diagrams show mainly the left hand trajectories they also contain the head and right hand trajectories. Using the gestures of Set 2 we have the possibility (after the projection in 3D space) to indicate the pointing direction by generating a ray from the head to the hand position. Example b) shows a difficult situation for our Human Tracking Module. The gesture "Speak louder" is represented by moving the hand to the ear. Our skin detector melts the head and the hand together interpreting it as one (bigger) head. The prediction of the hand does not cover the fact that the hand is still and predicts a trajectory above the head.

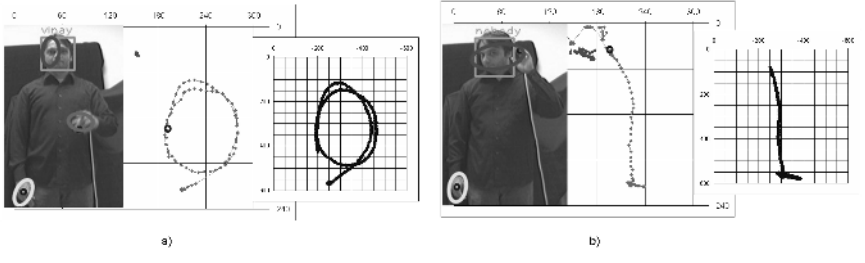

Fig. 6. Comparison of tracking results. a) Big Circle b) Speak louder.

5 Discussion and Conclusions

This article presented a framework towards a human-robot interaction based on gesture recognition. We presented the main architecture consisting of five hierarchical organized parts (i.e. Perception, Motor/Model, Impression, Interpretation and Knowledge Level). We introduced the guide robot "Nicole" to place our system in an embodied context. We sketched to novel approaches to represent human motion (i.e. Marionette Space and Labananalysis). We defined a gesture vocabulary organized in three sets (i.e. Cohens Gesture Lexicon, Pointing Gestures and Other Gestures). We presented experimental results from our

Human Tracking Module to show the feasibility of our gesture vocabulary and its representation in the vertical plane.

The future work will concerned with presenting results on 3D trajectory estimation and inertial sensor integration. To prove the benefits of a feature vector in the Marionette Control Space. Implement the Labananalysis module to produce Effort parameters.

Acknowledgements

The authors would like to thank Enguerran Boissier for his contributions to the Human Tracking Module. This work is partially supported by FCT-Fundação para a Ciência e a Tecnologia Grant #12956/2003 to J. Rett.

References

1. Gavrila, D.M.: The visual analysis of human movement: A survey. CVIU **73** (1999) pp. 82–98
2. Aggarwal, J.K., Cai, Q.: Human motion analysis: A review. CVIU **73** (1999) 428–440
3. Pentland., A.: Looking at people: Sensing for ubiquitous and wearable computing. IEEE Transactions on PAMI **22** (2000) 107–119
4. Moeslund, T.B., Granum, E.: A survey of computer vision-based human motion capture. CVIU **81** (2001) 231–268
5. Meltzoff, A.N., Moore, M.K.: Resolving the debate about early imitation. The Blackwell reader in developmental psychology, Oxford (1999) 151–155
6. Meltzoff, A.N., Moore, M.K.: Imitation of facial and manual gestures by human neonates. Science **198** (1977) 75–78
7. Cohen, C.J., Conway, L., Koditschek, D.: Dynamical system representation, generation, and recognition of basic oscillatory motion gestures. In: International Conference on Automatic Face- and Gesture-Recognition. (1996)
8. Kahn, R.E., Swain, M.J., Prokopowicz, P.N., Firby, R.J.: Gesture recognition using the perseus architecture. In: IEEE International Conference on Computer Vision and Pattern Recognition. (1996)
9. Dias, J.: Reconstrução Tridimensional Utilizando Visão Dinâmica. PhD thesis, University of Coimbra, Portugal (1994)
10. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press (2000)
11. Bartenieff, I., Lewis, D.: Body Movement: Coping with the Environment. Gordon and Breach Science, New York (1980)
12. Badler, N.I., Phillips, C.B., Webber, B.L.: Simulating Humans: Computer Graphics, Animation, and Control. Oxford Univ. Press (1993)
13. Zhao, L., Badler, N.I.: Acquiring and validating motion qualities from live limb gestures. Graphical Models **67** (2005) 1–16
14. Hoffmann, G.: Teach-in of a robot by showing the motion. In: IEEE International Conference on Image Processing. (1996) 529–532
15. Xing, S., Chen, I.M.: Design expressive behaviors for robotic puppet. In: International Conference on Control, Automation, Robotics And Vision. Volume 1. (2002) 378–383

16. Allot, R.: Gestural equivalence (equivalents) of language. Language Origins Society UCAL Berkeley (1994)
17. Chen, I.M., Tay, R., Xing, S., Yeo, S.H.: Marionette: From traditional manipulation to robotic manipulation. In: International Symposium on History of Machines and Mechanisms. (2004)
18. Yamane, K., Hodgins, J.K., Brown, H.B.: Controlling a motorized marionette with human motion capture data. In: IEEE International Conference on Robotics and Automation. (2003)
19. Hemami, H., Dinneen, J.A.: A marionette-based strategy for stable movement. IEEE Trans. on Systems, Man, and Cybernetics **23** (1993) 502–511
20. Loeb, G.E.: Learning from the spinal cord. Journal of Physiology **533.1** (2001) 111–117
21. Shamaie, A., Sutherland, A.: A dynamic model for real-time tracking of hands in bimanual movements. In: Gesture-based Communication in Human-Computer Interaction, LNAI 2915, Springer Verlag. (2003) 172–179
22. Soechting, J.F., Flanders, M.: Moving in three-dimensional space: Frames of reference, vectors, and coordinate systems. Annual Review of Neuroscience **15** (1992) 167–191
23. Carpenter, H.: Movement of the eyes. Volume 2nd ed. London Pion Limited, London (1988)
24. Lobo, J., Dias, J.: Inertial sensed ego-motion for 3d vision. Journal of Robotic Systems **21** (2004) 3–12
25. Lobo, J., Dias, J.: Vision and inertial sensor cooperation using gravity as a vertical reference. IEEE Trans. on PAMI **25** (2003) 1597–1608
26. Jose Barreto, P.M., Dias, J.: Human-robot interaction based on haar-like features and eigenfaces. In: IEEE International Conference on Robotics and Automation. (2004)
27. Paulo Menezes, J.B., Dias, J.: Face tracking based on haar-like features and eigenfaces. In: IFAC/EURON Symposium on Intelligent Autonomous Vehicles. (2004)
28. Bradski, G.R.: Computer vision face tracking for use in a perceptual user interface. Intel Technology Journal (1998) 15
29. Kalman, R.E.: A new approach to linear filtering and prediction problems. Trans. ASME—J.Basic Eng. **82** (1960) 35–45