2013 IEEE/RSJ International Conference on
Intelligent Robots and Systems (IROS)
November 3-7, 2013. Tokyo, Japan

Robust Pose Estimation Algorithm for Wrist Motion Tracking

F. Cordella, F. Di Corato, G. Loianno, B. Siciliano, L. Zollo

Abstract— **The wrist plays a fundamental role in reaching and grasping actions, i.e. it guides the hand to the grasp position and adjusts its orientation on the basis of the grasping type and task. This paper proposes a novel, low-cost method for wrist pose estimation by using the Asus Xtion Pro Live motion sensing device and a robust marker-based tracking approach based on Unscented Kalman Filter (UKF). The hand palm kinematic model is also considered. The applicability of the approach to evaluate some interesting kinematics parameters, such as position, orientation, Range Of Motion, angular and linear velocity and trajectory has been proved. In particular, since the nature of the paper is to present a novel approach for wrist pose estimation, only initial validation for wrist kinematic measurement will be reported.**

I. INTRODUCTION

The wrist plays a fundamental role in hand movements since it guides the hand determining its position and orientation during reaching and preshaping. In particular, wrist pose estimation plays a key role in many application fields. For instance, it is essential for learning by demonstration tasks, in which a robotic hand learns how to generalize an action performed by a human demonstrator [1]. Other important applications of wrist pose estimation are related to rehabilitation [2].

Aim of this work is to propose a vision-based algorithm for estimating the pose of the human wrist during preshaping. The developed approach possesses some interesting peculiarities: it relies on passive markers placed onto the hand palm. It is robust with respect to outliers contamination in the measured data, while keeping the detection phase very simple and fast. This is possible since it relies on a model of the anatomy of the palm, thus making it possible to adaptively identify the measurement of each visible marker. It is cheap, compared with the usual tracking visual systems, since it relies on only one RGBD camera. It is higlhy adaptable to different subjects, since it is able to estimate the anatomy of the palm, under some acceptably relaxed assumptions, as it will be clarified in Section II-C. Finally it is easy to implement. For the purpose of this paper, the wrist is regarded a 6 active Degrees of Freedom (DoFs) system, with 3 DoFs related to translation (due to arm motion) and 3 DoFs related to orientation (consisting of abduction/adduction, flexion/extension, pronation/supination). The approach proposed for wrist pose estimation is useful for motion analysis of the wrist joint. In particular, it could be applied in rehabilitation to evaluate, for instance, the joint Range of Motion (RoM) or trajectory in the reaching phase.

Different methods have been proposed for capturing human wrist motion: magneto-inertial sensors, optoelectronic cameras, computer vision-based techniques. Magneto-inertial sensors mounted on the human wrist can provide wrist acceleration and orientation; however, they suffer from drift problems (especially if velocity and position are desired outcomes [20], [4]). Marker-based motion analysis systems use optoelectronic cameras and reflective markers: although these systems provide accurate estimation of joints [6], they are expensive and cumbersome. Further, they require a completely structured environment to perform calibration and acquisition. Vision-based estimation techniques are sensitive to environmental conditions, but usually uses one or two cameras making the system cheap and space-saving. The approach presented in this paper tries to merge vision-based and marker-based techniques proposing a cheap system with reduced computational cost, easy to implement and robust. By using the Asus Xtion sensing system for motion analysis, a blob-detection algorithm and a filtering technique for tracking six markers positioned on the hand palm, it is possible to estimate the pose of the wrist and some kinematic parameters selected on the basis of literature findings.

In order to estimate human wrist pose, vision approaches commonly used can be classified in two main categories [7]: Model-based and single-frame pose estimation. Model-based visual pose estimation consists of finding the best matching between a group of features characterizing the input image and a group of model features. In order to reduce the computational cost of searching, a prediction step is considered. Multiple hypothesis around the prediction are considered to avoid local minima and discontinuities [8] in the matching. In particular, Bayesian filtering techniques using Monte Carlo methods, such as particle filters [9], [10] are applied. Single-frame pose estimation does not make assumptions on time coherence, making the problem very hard to solve. Global search over a database of templates [11] and motion constraints [12], [13] are viable solution.

In this paper, the wrist tracking problem has been formulated as a non-linear estimation problem solved by an Unscented Kalman Filter (UKF) [14]. With respect to other non-linear estimation techniques, such as Extended Kalman Filters (EKF), the UKF is proven to improve the estimation performance, is very simple to implement, improving the

F. Cordella, G. Loianno, B. Siciliano are with PRISMA Lab, Department of Electrical Engineering and Information Technology, Università di Napoli Federico II, via Claudio 21, 80125 Napoli, Italy {francesca.cordella, giuseppe.loianno, bruno.siciliano}@unina.it

F. Di Corato is with Research Center "E. Piaggio", Università di Pisa, largo Lucio Lazzarino, 56100 Pisa, Italy francesco.dicorato@for.unipi.it

L. Zollo is with Laboratory of Biomedical Robotics and Biomicrosystems, Università Campus Bio-Medico, via Alvaro del Portillo 21, 00128 Roma, Italy l.zollo@unicampus.it

software modularity, and does not need the computation of Jacobian matrices, as required by EKF. The main contribution of the paper is to show that the adoption of the proposed robust pose estimation algorithm gives promising results in wrist motion analysis. Future developments will be devoted to further improvements that better exploit the peculiar characteristics of the proposed pose estimation algorithm.

The paper is structured as follows: in Section II the kinematic and the motion model of the wrist are introduced; in Section III the wrist pose estimation algorithm is explained; results about the wrist motion tracking are presented in Section IV. Finally, conclusion and future work are proposed in Section V.

II. Modeling wrist motion and pose

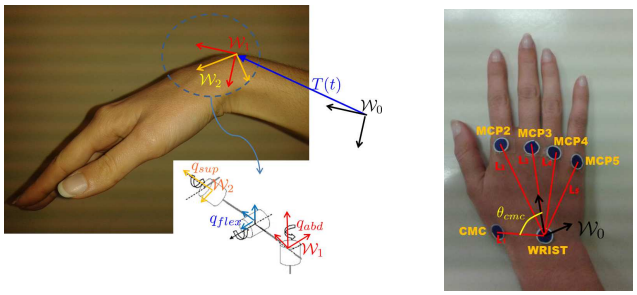
A. Notation

In this paper the classical notation of the Computer Vision and Robotics community is used [17]: the pose (rotation R_{ij} and translation T_{ij}) of frame $\mathcal{I}$ with respect to frame $\mathcal{J}$ is denoted with the group transformation $g_{ij} = \{R_{ij}, T_{ij}\} \in SE(3)$, which maps a vector expressed in frame $\mathcal{I}$ into a vector expressed in frame $\mathcal{J}$. The sole exception is made for the pose of the wrist frame with respect to the fixed frame (that will be formally defined), for which the subscripts are dropped, for cleaner notation, and it is denoted simply as g. The inverse transformation is indicated with the notation $g_{ij}^{-1} \triangleq \left\{R_{ij}^T, -R_{ij}^T T_{ij}\right\} \in SE(3)$. The action of the group transformation g_{jk} on g_{ij}, usually denoted with the symbol $\circ$, to indicate function composition, is indicated with a simple product, i.e. $g_{ik} = g_{jk} g_{ij} \triangleq g_{jk} \circ g_{ij}$, being by definition: $g_{ik} \triangleq \{R_{jk}R_{ij},\ R_{jk}T_{ij} + T_{jk}\}$. Finally the action of scaling by a certain amount α is defined as: $\alpha g_{ij} \triangleq \{R_{ij},\ \alpha T_{ij}\}$.

B. Wrist kinematic model

The wrist is modeled as a system with 6 DoFs, consisting of 3 components of translation and 3 angles of rotation, and the palm is assumed to be composed of rigid segments linked to the wrist. Therefore, the palm arch is not considered. In order to model wrist motion and pose, reference frames reported in Fig. 1(a) are defined: frame $\mathcal{W}_0$ is centered in the hand starting position, when the wrist is in a neutral position with 0 degrees for flexion/extension and 0 degrees for the radial-ulnar deviation and is fixed all along the wrist motion. Frame $\mathcal{W}_1$ is assumed to be coincident with $\mathcal{W}_0$ at the beginning of the motion. This frame is rigidly attached to the wrist and moves jointly with the arm. For the purpose of this work, the arm is supposed not to change its orientation during motion, thus it can be assumed that changes in hand orientation are due to actuation of the wrist joints only. The third frame, called $\mathcal{W}_2$, is defined at the end of the kinematic chain composed by the 3 wrist joints responsible for the rotation. Namely, $\mathcal{W}_2$ corresponds to the reference frame of the supination joint and its orientation with respect to $\mathcal{W}_1$ is parametrized via Euler angles in configuration ZYZ.

In order to evaluate the wrist angular component, the markers on the MCP joint are considered. The 5 angles θ_{cmc}, θ_{mcp2}, θ_{mcp3}, θ_{mcp4}, θ_{mcp5} are fixed and depend on the hand anatomy (Fig. 1(b)).



(a) Reference frames (b) Protocol

Fig. 1. a) The system reference frame, $\mathcal{W}_0$, has the X-axis along the line connecting the marker $WRIST$ with the marker placed on the middle finger MCP joint, the Z-axis perpendicular to the palm plane and the Y-axis defined with the right hand rule. b) Protocol used for marker positioning. Denavit-Hartenberg parameters are shown.

C. Acquisition Procedure

In order to have information about the wrist pose (position and orientation) during motion, six coloured markers have been positioned on the human hand, as shown in Figure 1(b). The markers are made of blue paper of diameter 1.2 cm. DH parameters are evaluated in order to make the approach applicable to different hand sizes. Therefore, the algorithm envisages an initial user-specific calibration phase, where marker centers are detected manually in the first frame acquired by the camera and the link lengths are measured. It is assumed that the camera focal axis is perpendicular to the plane where the hand lies. A minimum distance of 60 cm from the hand plane must be guaranteed.

D. Motion model

With the assumptions made in Section II-B, the frame $\mathcal{W}_2$ experiences a motion with respect to the fixed frame $\mathcal{W}_0$ according to the following continuous-time kinematic model:

$$\begin{cases} \dot{T}(t) = v(t) \\ \dot{v}(t) = \eta_v(t) \\ \dot{R}(t) = R(t)\,\Omega(t) \\ \Omega(t) = \eta_\omega(t)\,\wedge \end{cases} \tag{1}$$

where $\Omega(t)$ is the skew symmetric matrix of the wrist angular velocity $\eta_\omega(t)$ expressed in the coordinates of $\mathcal{W}_2$ (being $\wedge$ the cross-product operator), $T(t), v(t)$ and $R(T)$ are respectively the position, linear velocity and rotation matrix of $\mathcal{W}_2$ with respect to $\mathcal{W}_0$. Finally $\eta_v(t)$ and $\eta_\omega(t)$ are zero-mean white noises with constant variance, modeling the wrist linear accelerations and angular velocities as random walks. This choice is justified by the fact that it is assumed to not have any prior information regarding the nature of the wrist motion. The variables $T(t)$ and $R(t)$ can be put together to define the group transformation $g(t) \triangleq \{R(t), T(t)\} \in SE(3)$, which fully describe the 6 -DoF localization problem of the wrist with respect to the defined fixed frame.

Given the known positions $T_{wi} \in \mathbb{R}^3$ of the markers, expressed in the coordinates of the local frame $\mathcal{W}_2$, the

generic projection of the i-th marker on the image plane can be written as:

$$y_i(t) = \pi\left(g_{w_0c} g(t) T_{wi}\right) + \nu_i(t) \tag{2}$$

where $\pi() : \mathbb{R}^3 \rightarrow \mathbb{R}\mathrm{P}^2$ denotes the projective operator, according to the pinhole model, and $\mathbb{R}\mathrm{P}^2$ represents the projective space, see [17]. The group transformation $g_{w_0c} \in SE(3)$ is the pose (translation and rotation) between the camera frame and the fixed frame $\mathcal{W}_0$, which is assumed known. A procedure for the estimation of the foregoing transformation is given in Section III-D. Moreover, $\nu_i(t)$ is a zero-mean white noise with variance R, assumed constant among features. The measurement equations can thus be written as:

$$y(t) = \begin{bmatrix} \pi\left(g_{w_0c} g(t) T_{w1}\right) + \nu_1(t) \\ \pi\left(g_{w_0c} g(t) T_{w2}\right) + \nu_2(t) \\ \vdots \\ \pi\left(g_{w_0c} g(t) T_{w6}\right) + \nu_6(t) \end{bmatrix}. \tag{3}$$

Note that in the case of the marker on the wrist, $y_1(t) = \pi\left(g_{w_0c} g(t) T_{w1}\right) + \nu_1(t)$, the projection on the image plane can be simplified as:

$$y_1(t) = \pi\left(g_{w_0c} T(t)\right) + \nu_1(t) \tag{4}$$

being by definition $T_{w1} = 0$.

III. Wrist pose estimation algorithm

A. Detection and tracking

The first step in the tracking algorithm is the features detection to extract the Regions of Interest (RoIs) from the image. This can be done in two steps: by locating the hand (via skin detection) and then by recognizing colored markers in each input frame of the obtained RoI. The blobs whose histogram is as close as possible (in the sense of Bhattacharyya similarity coefficient) to the (blue) reference color histogram are extracted. Then, a connected component labeling algorithm is used. Once the blobs on the scene have been determined, the 2D coordinates of every blob center are computed and the markers are tracked during the whole video. The use of simple detection algorithms like the one used may render the task of associating visual measurements to physical markers or deciding whether a given measurement is an outlier or a valid marker projection difficult. For this reason, since a wrist model was employed in this work, it was found convenient to reformulate the tracking problem in a stochastic optimization problem, embedded into the estimation task. This showed very good and accurate results.

B. Filtering motion and pose

According to the motion parameter dynamics in (1), given the image-space measurements (3), a non-linear estimation scheme, based on the Unscented Kalman Filter [14], was designed and tested with real datasets. The aim of the filter is to estimate the state $x(t)$ of the system, consisting of the motion variables, $T(t), v(t)$ and the Euler angles parametrization of the rotation matrix $R(t)$, which reflects the current value of wrist joint angles. For the purpose of real-time implementation, the kinematic equations (1) have been time-discretized using the Euler integration method. The base sample time has been chosen coincident with the sampling rate of the camera, i.e. $dt \approx 0.03s$. Therefore, the resulting discrete time equations of the estimator are:

$$\begin{cases} T(t+1) = T(t) + v(t)\, dt \\ v(t+1) = v(t) + \eta_v(t)\, dt \\ R(t+1) = R(t)\, e^{\Omega(t) dt} \\ \Omega(t) = \eta_\omega(t) \wedge \\ y_i(t) = \pi\left(g_{w_0c} g(t) T_{wi}\right) + \nu_i(t), \ i \in \mathcal{V}(t) \subseteq \{1, 2, ..., 6\} \end{cases} \tag{5}$$

Note the exponential approximation used to numerically integrate the Euler angles rotation matrix and the fact that the attitude-related noise term enters non-linearly into the mathematical model. The set $\mathcal{V}(t)$ denotes the group of visible markers at the current time (omitting the clutters). It incorporates the time index since the markers may move out of the field of view or be occluded.

In this paper, given the non-linearity of the model with respect to the state and the orientation noise terms, the Augmented Unscented Kalman Filter algorithm presented in [14] has been used. The peculiarity of the adopted estimation scheme, compared with the classical UKF approach [15], is the possibility to easily deal with non-affine noise terms in the state/measurement model. For the remaining part, the technique is a classical UKF as in [15]. The full algorithm can be found in [14] and it will be omitted for brevity.

C. Dealing with occlusions and outliers: the association problem

The outputs given by the blob detection algorithm, for the image at the time t, are given by a random sequence of M_t measurements $\mathbf{y}_t = \{y_1(t), y_2(t), ..., y_{M_t}(t)\}$ of blob candidates. In general the condition $M_t \neq 6$ holds, which means that the sequence $\mathbf{y}_t$ does contain projections of visible markers and clutters. The randomness of the measurement sequences is a fundamental issue in this framework, since it implies some important consequences: i) the associations between measurement h and marker j or with a clutter cannot be decided a priori and has to be set; ii) each sequence of measurements for each frame can be considered conditionally independent of every other sequence in the past; iii) once the current sequence of associations has been defined, it can be considered conditionally independent of the past history of associations as well. A direct consequence is that predicting the order in which markers and clutters are detected, for each image, is not allowed (as experimental tests showed). Because of the above hypotheses, it is claimed that the best way to solve the filtering problem, while ensuring robustness to clutters, is by using probabilistic techniques. For this aim, consider a latent variable $a_i(t)$, modeling the measurement-to-marker association [3]:

$$a_i(t) = \begin{cases} 0, \text{ if } y_i(t) \text{ is a clutter} \\ j, \text{ if } y_i(t) \text{ is the projection of marker } j \end{cases} \tag{6}$$

It is desired to find the most probable value of the variable $a_i(t)$, $\forall i = 1, \ldots, M_t$, that is for every measurement collected at the current time step.

Introducing the latent variable is the same as considering the non-linear measurement model (3), in compact form $h(x(t))$, as a conditional function of the variable $a_i(t)$. In fact, it is possible to condition the output function over a certain value of the latent variable, i.e. $h\left(x(t)\left|a_i(t) = j \neq 0\right.\right)$, meaning to select the rows corresponding to the projection of the marker j from the function $h(x(t))$.

The introduction of the latent variable, allows to solve the association problem, which can be recast as maximizing the belief that the current measurement $y_i(t) \in \mathbf{y}_t$ is either the projection of a visible marker or a clutter. Formalizing, the aim is to find the maximum of the posterior distribution:

$$p\left(a_i(t)\left|y_i(t), \mathbf{y}_{0:t-1}\right.\right) \propto p\left(y_i(t)\left|a_i(t), \mathbf{y}_{0:t-1}\right.\right) p\left(a_i(t)\right) \tag{7}$$

given the current measurement $y_i(t)$ and the whole history of the measurements up to the previous step. The previous equation was obtained via application of Bayes' rule. The *prior* $p(a_i(t))$ is assumed to be independent of the previous measurements and is determined by the a priori knowledge of clutter and marker association event probabilities. Even though better choices could exist, in this work the association priors are assigned as in [3], which showed acceptable results. In the following the time index will be dropped for simplicity, when its disambiguation is straightforward. The density $p\left(y_i\left|a_i, \mathbf{y}_{0:t-1}\right.\right)$ is the likelihood that the current measurement is associated to a given marker or to a clutter. This distribution can be obtained via marginalization of a proper joint density:

$$p\left(y_i\left|a_i, \mathbf{y}_{0:t-1}\right.\right) = \int p\left(y_i\left|x, a_i, \mathbf{y}_{t-1}\right.\right) p\left(x\left|a_i, \mathbf{y}_{t-1}\right.\right) dx. \tag{8}$$

Fixing a certain guess for the association, $a_i(t) = j,\ j \neq 0$, Equation (8) is the Kalman Filter likelihood of the measurement $y_i(t)$, given the prediction of the marker j (i.e. given the conditioning of the measurement model over that value of the latent variable). Thus, given the predicted state-related Sigma-Points [14], $\mathbf{X}^x_{n,t/t-1},\ n = 1, \ldots, N$, computed by employing the non-linear state model, their transformation through the conditioned measurement function can be obtained, as in the classical UKF:

$$\mathbf{Y}^j_{n,t/t-1} = h\left(\mathbf{X}^x_{n,t/t-1}\left|a_i = j\right.\right). \tag{9}$$

The superscript j on the transformed Sigma-Points of the output, indicates that $Y^j_{n,t/t-1}$ refers to the predicted projection of the marker j, for which the association is being tested. The mean and covariance of the measurement vector are calculated as:

$$\widehat{y}^-_j = \sum_{n=0}^{N} W^n_m \mathbf{Y}^j_{n,t/t-1} \tag{10}$$

$$P^-_{yy,j} = \sum_{n=0}^{N} W^n_c \left(\mathbf{Y}^j_{n,t/t-1} - \widehat{y}^-_j\right)\left(\mathbf{Y}^j_{n,t/t-1} - \widehat{y}^-_j\right)^T + R \tag{11}$$

where W^n_m and W^n_c are the weights associated to the Sigma-Points [14], $\widehat{y}^-_j$ is the predicted projection of the marker j and $P^-_{yy,j}$ its covariance. The probability of the association $a_i = j$ can be thus computed as:

$$p\left(a_i = j\left|y_i, \mathbf{y}_{0:t-1}\right.\right) \propto \mathcal{N}\left(y_i - \widehat{y}^-_j, P^-_{yy,j}\right) p\left(a_i = j\right), \tag{12}$$

being $\mathcal{N}()$ the multivariate normal distribution of proper mean value and covariance. It is worth to mention that, when testing the association to a clutter, $a_i = 0$, Equation (12) is written as $p\left(a_i = 0\left|y_i, \mathbf{y}_{0:t-1}\right.\right) \propto (1/RES)\, p(a_i = 0)$, where RES is the image resolution, meaning that a clutter can happen everywhere in the image. Selecting the maximum probability among the ones in (12), will give the most probable value of the variable $a_i(t)$, corresponding to the measurement $y_i(t)$. The association problem is solved by repeating the above procedure for all the measurements in the set $\mathbf{y}_t$. Then, the correction step can take place, employing the visible markers and the associated image projections, as in the classical UKF.

D. Filter initialization

The initialization phase is responsible of the estimation of the relative pose between the camera and the fixed reference frame, and needs to be reasonably accurate. For this reason, the estimation is formulated as a Least-Squares optimization problem. During this phase, the markers are required to be visible, such that the association between markers and measurements can be made without effort, after the detection phase. Therefore, no probabilistic optimization needs to be carried out. Finally, the hand must be in neutral configuration, with null angles for the wrist joints. The measurements employed during the initialization phase are the projection of the markers on the image plane and the measurement of their depth, relative to the camera, which are obtained via the available IR camera. The equation mapping the available measurements into the estimation variables are:

$$\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} \pi\left(g_{w_0 c}(\theta)\, T_{wi}\right) \\ \vdots \\ e_3^T\left(g_{w_0 c}(\theta)\, T_{wi}\right) \\ \vdots \end{bmatrix} = \begin{bmatrix} h_y\left(g_{w_0 c}(\theta)\right) \\ h_z\left(g_{w_0 c}(\theta)\right) \end{bmatrix} \tag{13}$$

$$\bar{y} = \bar{h}\left(g_{w_0 c}(\theta)\right). \tag{14}$$

Note that in this case the relative transformation $g_{w_0 c}(\theta)$ is parametrized via $\theta \in \mathbb{R}^6$, which encodes the unknown pose parameters (translation and angular parametrization) to be estimated. y_i and z_i are respectively the measured projections and depths of the markers. Finally it is $e_3 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$. The locally optimal estimation of the foregoing transformation is found by minimizing the 2-norm cost function:

$$\min_{\theta} \|\bar{y} - \bar{h}\left(g_{w_0 c}(\theta)\right)\|^2. \tag{15}$$

The linearization of the non-linear function $\bar{h}\left(g_{w_0c}\left(\theta\right)\right)$ around an initial estimation of the pose parameter θ_0, gives:

$$J_\theta = \|\bar{y} - \bar{h}\left(g_{w_0c}\left(\theta\right)\right)\|^2 \quad (16)$$

$$\approx \|\bar{y} - \bar{h}\left(g_{w_0c}\left(\theta_0\right)\right) - H_{\theta_0}\left(\theta - \theta_0\right)\|^2 \quad (17)$$

$$= \|\widetilde{y} - H_{\theta_0}\theta\|^2 \quad (18)$$

In the previous equation, it is $H_{\theta_0} = \frac{\partial \bar{h}}{\theta}\Big|_{\theta_0}$. Equation (18) is a well-known linear quadratic cost function, whose minimum is obviously given by:

$$\widehat{\theta} = H_{\theta_0}^{\dagger}\widetilde{y}. \quad (19)$$

That is, expanding the solution, the optimal estimation of the relative pose between the camera and the wrist reference frames, at the initial time, is given by:

$$\widehat{\theta} = R_H\theta_0 + H_{\theta_0}^{\dagger}\left(y - \bar{h}\left(g_{w_0c}\left(\theta_0\right)\right)\right) \quad (20)$$

where $R_H = H_{\theta_0}^{\dagger}H_{\theta_0}$ is the range-space projector of the matrix H_{θ_0}.

IV. Experimental Validation of Wrist Tracking Algorithm

A. Experimental Setup

In order to evaluate the behaviour of the human wrist, it has been chosen to measure: the Range of Motion (RoM), the orientation and the velocity, which provide an indication of the ability of a person to perform a movement [18], [19].

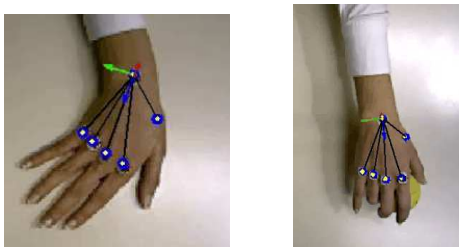
In order to detect and track markers, the Asus Xtion Pro-live has been used. It is a motion sensing device consisting of an InfraRed (IR) laser emitter, an IR camera for measuring depth information, and a RGB camera. It captures depth and color images simultaneously at a frame rate of about 30 frames per second (fps). The resolution of the RGB camera is 640×480. The IR camera and the IR emitter form a stereo pair. OpenNI library has been used in order to make the sensing device work on the PC and the whole algorithm has been implemented under ROS (Robotic Operating System) for ensuring a real-time approach.

B. Preliminary experimental results

The proposed algorithm has been experimentally tested for tracking the wrist and extracting the above mentioned wrist kinematic parameters in two conditions: RoM movements and reach-and-grasp action. The paper aims at providing a proof-of-concept of the pose estimation approach for evaluating those parameters; hence, the study is still preliminary and only one subject has been involved in the experimental tests. The participant, a woman without pathological diseases of the wrist, was seated in front of a table with the right hand placed on it. In the starting configuration of the hand, the four fingers are fully extended, the thumb is adducted and the wrist is in a neutral position. The subject was asked to perform the wrist movements necessary for measuring its RoM. Further, she was asked to reach and grasp an object located on the table. In both the experiments, the subject paid special attention at not rotating the arm.

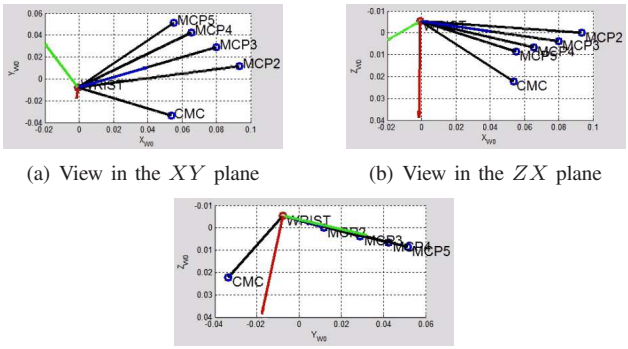
Figure 2(a) shows one frame of the action performed for evaluating the range of motion. The links connecting the wrist with the MCP joints are outlined in black. Their distance with respect to the wrist and their mutual positions are constant, depending only on the hand anatomy. The introduction of this information in the tracking algorithm improves its robustness with respect to outliers and to the possibility of confusing markers among them due to possible occlusions.

Figure 3 illustrates the wrist frame orientations obtained by using the rotation matrix estimated by the filter. In particular, their view from the YX, ZX and ZY planes are shown.



(a) Abduction/Adduction (b) Reaching

Fig. 2. a) One frame of the adduction/abduction movement. b) Last frame of the reaching phase. Frame $\mathcal{W}_2$ is shown.



(a) View in the XY plane (b) View in the ZX plane

(c) View in the ZY plane

Fig. 3. Wrist frame orientation in the plane a) XY, a) ZX, a) ZY.

Figure 4 shows an example of measurement of the three wrist angular components during abduction/adduction movements. From it, the RoM of the rotation performed around the z-axis can be extracted. As expected, it varies from 25^o to -25^o that represents the wrist abduction/adduction range.

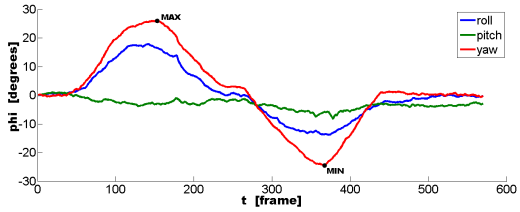
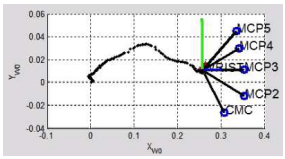
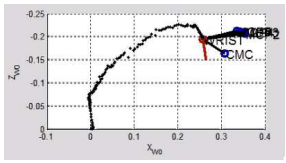


Fig. 4. Angular components obtained during wrist abduction/adduction motion. As outlined in the legend, the 3 components are drawn with different colours.

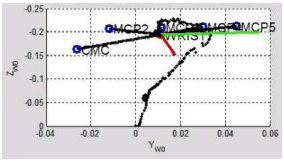
The acquired data have also been used for analyzing the wrist behaviour in the reaching phase. In particular, the trajectories implemented by the subject (Fig. 5) during the reach-and-grasp action (Fig. 2(b)) have been recorded as well as the velocities (Fig. 6).



(a) View in the YX plane



(b) View in the ZX plane



(c) View in the ZY plane

Fig. 5. Wrist trajectory, outlined in black, in the plane a) XY, a) ZX, a) ZY.

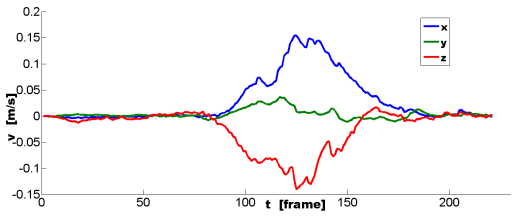


Fig. 6. Wrist velocity profile along the whole reaching movement.

These preliminary results show some reasonable estimation capabilities and they seem very promising for further development.

V. CONCLUSION

In this paper, a novel and low-cost method for wrist pose estimation has been proposed. The Asus motion sensing device has been used in order to track the wrist behaviour during different movements. In particular, the wrist tracking problem has been formulated as a non-linear estimation problem solved by the Unscented Kalman Filter. The interdepedence of the markers has been taken into account by introducing the palm kinematic model. Information about the joint orientation, position, trajectory and velocity has been extracted in order to demonstrate that the proposed pose estimation algorithm can be adopted for finding kinematic parameters about the wrist. The approach can have useful applications in rehabilitation providing quantitative information about the performed task, such as the measurement of joint motion. Further improvements will be devoted to extend the approach for estimating the pose of the whole hand (fingers included) and to test the accuracy of the approach by means of a comparison with a ground truth obtained with an optoelectronic system.

VI. ACKNOWLEDGEMENTS

This work was supported in part by the National PRIN Project ROCOCÒ – COoperative and COllaborative RObotics (CUP E61J11000300001) and partly by the National PRIN Project HandBot (CUP: B81J12002680008).

REFERENCES

[1] A. Sahbania, S. El-Khoury, P. Bidaud, “An overview of 3D object grasp synthesis algorithms”, *Robotics and Autonomous Systems*, vol. 60, no. 3, pp. 326–336.

[2] A. Karime, E. Mohamad, W. Gueaieb, A. El Saddik, “Determining wrist reference kinematics using a sensory-mounted stress ball”, *IEEE International Symposium on Robotic and Sensors Environments*, pp. 109–114, 2012.

[3] F. Cordella, F. Di Corato, L. Zollo, B. Siciliano, P. van der Smagt, “Patient performance evaluation using Kinect and Monte Carlo-Based fnger tracking”, *IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics*, pp. 1967–1972, 2012.

[4] L. Zollo, L. Rossini, M. Bravi, G. Magrone, S. Sterzi, E. Guglielmelli, “Quantitative evaluation of upper-limb motor control in robot-aided rehabilitation”, *Medical and Biological Engineering and Computing*, vol. 9, no. 49, pp. 1131–1144, 2011.

[5] E. Foxlin, “Motion tracking requirements and technologies”, *Handbook of Virtual Environments: Design, Implementation, and Applications*, Lawrence Erlbaum Associates, vol. 35, pp. 1989–2002, 2007.

[6] P. Cerveri, E. De Momi, N. Lopomo, G. Baud-Bovy, R.M. Barros, G. Ferrigno, “Finger kinematic modeling and real-time hand motion estimation”, *Annals of Biomedical Engineering*, vol. 35, pp. 1989–2002, 2007.

[7] A. Erol, G. Bebis, M. Nicolescu, R.D. Boyle, X. Twombly, “Vision-based hand pose estimation: A review.” *Computer Vision and Image Understanding*, vol. 108, no. 1, pp. 52–73, 2007.

[8] J. Deutscher, A. Blake, I. Reid, “Articulated body motion capture by annealed particle filtering”, *Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 126–133, 2000.

[9] M. Bray, E. Koller-Meier, L. Van Gool, “Smart particle filtering for 3D hand tracking”, *Conf. on Automatic Face & Gesture Recognition*, pp. 675–680, 2004.

[10] W.Y. Chang, C.S. Chen, Y.P. Hung, “Appearance-guided particle filtering for articulated hand tracking”, *Conf. on Computer Vision and Pattern Recognition*, vol. 1, pp. 235–242, 2005.

[11] B. Stenger, A. Thayananthan, P.H.S. Torr, R. Cipolla “Hand pose estimation using hierarchical detection”, *Workshop on Human-Computer Interaction*, 2004.

[12] C.S. Chua, H.Y. Guan, Y.K. Ho, “Model-based finger posture estimation”, *Asian Conference on Computer Vision*, 2000.

[13] J. Lee, T. Kunii, “Constraint-based hand animation”, *Models and Techniques in Computer Animation*, pp. 110–127, 1993.

[14] R. Kandepu, B. Foss, L. Imsland, “Applying the unscented Kalman filter for nonlinear state estimation”, *Journal of Process Control*, vol. 18, no. 7–8, pp. 753–768, 2008.

[15] E.A. Wan, R. Van der Merwe, “The unscented Kalman filter for nonlinear estimation”, *IEEE Symposium on Adaptive Systems for Signal Processing, Communications, and Control*, pp. 153–158, 2000.

[16] B. Siciliano and O. Khatib, *Springer Handbook of Robotics*, Springer, Heidelberg, 2008.

[17] Y. Ma, S. Soatto, J.vKosecka, S.S. Sastry, *An Invitation to 3-D Vision: From Images to Geometric Models*, Springer, New York, 2003.

[18] G. Pellegrino, L. Tomasevic, M. Tombini, G. Assenza, E. Gallotta, S. Sterzi, V. Giacobbe, L. Zollo, E. Guglielmelli, G. Cavallo, F. Vernieri, F. Tecchio, “Interhemispheric coupling changes associate with motor improvements after robotic stroke rehabilitation”, *Restorative Neurology and Neuroscience*, 2012.

[19] D. Formica, L. Zollo, E. Guglielmelli, “Torque-dependent compliance control in the joint space of an operational robotic machine for motor therapy”, *ASME Journal of Dynamic Systems, Measurement, and Control*, vol. 128, pp. 152–158, 2006.

[20] D. Formica, H.I. Krebs, S.K. Charles, L. Zollo, E. Guglielmelli, N. Hogan, “Passive wrist joint stiffness estimation”, *Journal of Neurophysiology*, 2012.