

Pose estimation algorithm for hand assessment

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

Abstract—In order to evaluate therapy efficacy in restoring impaired sensori-motor functions, patient performance assessment should be as much as possible precise and independent of the examiner subjective judgment. In this paper, a method for continuously monitoring patient activity in order to correct wrong movements and to follow patient improvements is proposed. In particular, a novel low-cost method for hand pose estimation by using a monocular motion sensing device and a robust marker-based pose estimation approach based on the Unscented Kalman Filter is presented. The hand kinematics is used to enclose geometrical constraints in the estimation process. The approach is applied for evaluating some significant kinematic parameters necessary for understanding human hand motor improvements during rehabilitation. In particular, the estimated performance indicators are joint positions, angles, Range of Motion (RoM) and trajectory for the fingers and position, orientation and velocity for the wrist.

I. INTRODUCTION

Cerebrovascular diseases, such as stroke, are the third leading cause of death in industrialized countries and the leading cause of permanent disability [1]. This leads to a remarkable demand of healthcare services with consequently increasing public expenses. The aim of neuro-rehabilitation is to promote and accelerate the functional recovery of motor ability compromised by brain damages of various nature, exploiting the intrinsic neural plasticity: motor patterns are relearned through repeated execution of predefined movements [2]. Patient monitoring is needed to evaluate the quality of the performed movements, modify the therapy if needed, apply corrective actions and assess patient performance. Quantitative functional evaluation often suffers from limitations due for instance to the subjective judgment of the examiners. In order to overcome these limitations, strummental evaluations are preferred. For instance, the application of human motion tracking techniques to rehabilitation is finding more and more consensus. Systems for human movement tracking applied to rehabilitation [3] are usually divided into two categories: non-visual tracking systems and visual tracking systems. In this paper the attention is focused on systems belonging to the latter class, which are in turn classified as marker-based and marker-less systems. Marker-based motion analysis systems use optoelectronic cameras

and reflective markers: although these systems provide an accurate estimation of joints [4], they are expensive and cumbersome. Further, they require a completely structured environment to perform calibration and acquisition. Marker-less systems rely on Computer Vision algorithm that are sensitive to environmental conditions, but usually use one or two cameras making the system cheap and space-saving.

Vision-based techniques for estimating the hand pose are usually grouped into two categories [5]: 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 [6] in the matching. In particular, Bayesian filtering techniques using Monte Carlo methods, such as particle filters [7], [8], [9] 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 [10] and motion constraints [11], [12] are viable solution.

The hand pose estimation approach presented in this paper tries to merge computer-vision and marker-based techniques proposing a cheap system (that facilitates a fundamental step for hand pose estimation: the triangulation process of the visual features) using a monocular camera, with reduced computational cost, easy to implement and robust. It performs the visual analysis of human hand motion and records hand joint kinematics during movements in a robust and repeatable way making the system adapt for home based rehabilitation.

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

II. HAND KINEMATICS

In this work, the hand wrist is modeled as a system with 6 Degrees of Freedom (DoFs), consisting of 3 components of translation and 3 angles of rotation, i.e. Adduction/Abduction (A/A), Flexion/Extension (F/E), Pronation/Supination (P/S). These angles correspond to the Euler angles in configuration ZYX. The palm is composed by rigid segments linked to the wrist and its anatomy is assumed known. For the purposes 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

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

F. Di Corato is with Research Center "E. Piaggio", Università di Pisa, largo Lucio Lazzarino, 1, 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

wrist joints only. The long fingers are four kinematic chains, each composed by MetaCarpo-Phalangeal (MCP) joint (i.e. a 2 DoFs joint), Proximal Inter-Phalangeal (PIP) and Distal Inter-Phalangeal (DIP) joints (that are 1 DOF joint each). It has been assumed a coupling between PIP and DIP joints ($\theta_{DIP} = \frac{2}{3}\theta_{PIP}$) [13]. The thumb is modeled with 5 DOFs [14]: the TrapezioMetacarpal (TM) and the MCP are 2 DOF joints with A/A and F/E axes orthogonal and incident, whereas the interphalangeal (IP) joint has 1 DOF. Denavit-Hartenberg (DH) parameters for the index finger and for the thumb are shown in Tabs. I and II, respectively.

TABLE I
DENAVIT-HARTENBERG PARAMETERS OF THE INDEX FINGER

Joint #	a_i	α_i	d_i	θ_i
1	0	-90°	0	θ_{MCP2}
2	L_P^{index}	0°	0	θ_{flex}^{MCP2}
3	L_M^{index}	0°	0	θ_{PIP2}
4	L_D^{index}	0	0	θ_{DIP2}

TABLE II
DENAVIT-HARTENBERG PARAMETERS OF THE THUMB

Joint #	a_i	α_i	d_i	θ_i
1	0	-90°	0	θ_{TM}
2	L_P^{thumb}	90°	0	θ_{flex}^{TM}
3	0	-90°	0	θ_{MCP1}
4	L_M^{thumb}	0°	0	θ_{PIP1}
5	L_D^{thumb}	0°	0	θ_{DIP1}

The 5 kinematic chains of the fingers have the origin in common (i.e. the wrist). Fig. 1 shows the joint reference frames for the index finger and for the thumb.

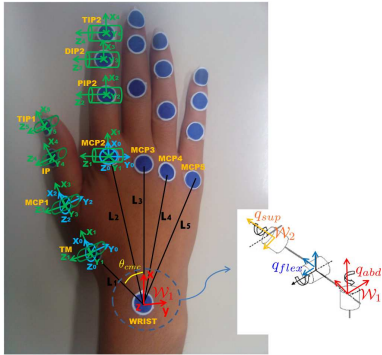


Fig. 1. Protocol used for marker positioning and joint reference frames in the hand starting position. The system reference frame, positioned on the hand wrist, is outlined in red.

DH parameters are evaluated in such a way as to obtain a generic algorithm valid for different hand sizes. Therefore, the algorithm envisages an initial calibration phase, where marker centers are detected manually in the first image acquired by the camera and the link lengths are measured by means of the depth information provided by the vision system. It is assumed that the camera focal axis is perpendicular to the plane where the hand lies.

III. ALGORITHM FOR POSE DETECTION AND ESTIMATION

The pose of the hand has been estimated by positioning 21 blue paper markers on the subject hand in a configuration chosen in such a way to minimize artifacts, due for example to skin movements or marker occlusion, and for obtaining information about the wrist (Fig. 1). Markers have been recognized by a fast detector based on color histogram and a connected component labeling algorithm [15]. The Asus Xtion Pro live motion sensing device working at 30 *fps* and consisting of an InfraRed (IR) laser emitter, an IR camera for measuring depth information and a RGB camera, with a resolution of 640×480 , has been used. In [15], the authors outlined that the simple detection algorithms used makes it difficult to associate visual measurements to physical markers and to decide whether a given measurement is an outlier or a valid marker projection. Therefore, since a model of the hand is available, the marker tracking problem has been reformulated into a stochastic optimization problem. This renders the proposed algorithm robust with respect to outliers and markers entering and exiting from the field of view.

The pose parameters – position, $T(t)$ and orientation, $R(t)$ – of the wrist with respect to its initial pose (corresponding to the first image), together with the kinematics of the 17 finger joints can be modeled according to the following discrete-time kinematic model:

$$\begin{cases} T(t+1) = T(t) + v(t) dt \\ v(t+1) = v(t) + \eta_v(t) dt \\ R(t+1) = R(t) e^{(\Omega(t)dt)} \\ \theta_i(t+1) = \theta_i(t) + \eta_{\theta_i}(t) dt, \quad i = 1, \dots, 17 \end{cases} \quad (1)$$

where $\Omega(t) = \eta_\omega(t) \wedge$, being $\wedge$ the skew-symmetric operator, $\eta_v(t)$, $\eta_\omega(t)$ and $\eta_{\theta_i}(t)$ are zero-mean white noises with constant variance, modeling the hand motions as random walks, and dt is the base sample time, chosen coincident with the sampling rate of the camera. The rotation matrix $R(t)$ is parametrized via Euler angles and encodes the current value of wrist joint angles. The output model is represented by the projection of the visible markers on the image space.

$$y_i(t) = \pi(g_{w_0c}(t) T_{m_iw}(\Theta)) + \nu_i(t), \quad i \in \mathcal{V}(t) \subseteq \{1, 2, \dots, 21\} \quad (2)$$

where $g(t) = \{R(t), T(t)\} \in SE(3)$ and $T_{m_iw}(\Theta)$ is the 3D position of the i -th marker with respect to the wrist reference frame. This position is a function of the hand kinematic parameters Θ , i.e. the joint angles θ_i and the DH parameters, and can be obtained via direct kinematic. 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 and $\nu_i(t)$ is a zero-mean white noise with variance R_i , assumed constant among features. 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.

According to the kinematic model (1) and the output model (2) a nonlinear estimation scheme has been designed. The aim of the filter is to estimate the state $x(t)$ of the system, consisting of: i) the motion variables, $T(t)$, $v(t)$ and

the Euler angles parametrization of the rotation matrix $R(t)$, and ii) the joint angles $\theta_i(t)$ of the fingers. 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 [16] has been used. The peculiarity of the adopted estimation scheme, compared with the classical UKF approach [17], 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 [17]. For the purposes of this work, the outputs given by the blob detection algorithm, for the image at the time t , are considered as a random sequence of M_t measurements $\mathbf{y}_t = \{y_1(t), y_2(t), \dots, y_{M_t}(t)\}$ of blob candidates. In general the condition $M_t \neq 21$ holds, which means that the sequence $\mathbf{y}_t$ does contain projections of visible markers and clutters. It implies that: i) the associations between measurement h and marker j or with a clutter cannot be decided a priori and has to be estimated; ii) each sequence of measurements for each image 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. Because of the above hypotheses, the filtering problem is solved by using a probabilistic technique. To this end, consider a latent variable $a_i(t)$, modeling the measurement-to-marker association [15]:

$$a_i(t) = \begin{cases} 0, & \text{if } y_i(t) \text{ is a clutter} \\ j, & \text{if } y_i(t) \text{ is the projection of marker } j \end{cases} \quad (3)$$

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

$$p(a_i(t) | y_i(t), \mathbf{y}_{0:t-1}) \propto p(y_i(t) | a_i(t), \mathbf{y}_{0:t-1}) p(a_i(t)) \quad (4)$$

given the current measurement $y_i(t)$ and the whole history of the measurements up to the previous step. The *prior* $p(a_i(t))$ is determined by the a priori knowledge of clutter and marker association event probabilities [15], while the density $p(y_i | a_i, \mathbf{y}_{0:t-1})$ is the likelihood that the current measurement is associated to a given marker or to a clutter. Fixing a certain guess for the association, $a_i(t) = j$, $j \neq 0$, the density $p(y_i | a_i, \mathbf{y}_{0:t-1})$ is the Kalman Filter likelihood of the measurement $y_i(t)$, given the prediction of the marker j . Thus the probability of the association $a_i = j$ can be computed as:

$$p(a_i = j | y_i, \mathbf{y}_{0:t-1}) \propto \mathcal{N}(y_i - \hat{y}_j^-, P_{yy,j}^-) p(a_i = j), \quad j = 0, \dots, 21 \quad (5)$$

being $\mathcal{N}()$ the multivariate normal distribution of proper mean value and covariance, while $\hat{y}_j^-$ is the predicted projection of the marker j and $P_{yy,j}^-$ its covariance, obtained as in the standard UKF algorithm [16]. When testing the association to a clutter, $a_i = 0$, Equation (5) is written as $p(a_i = 0 | y_i, \mathbf{y}_{0:t-1}) \propto (1/RES) p(a_i = 0)$, where RES

is the image resolution, meaning that a clutter can happen everywhere in the image. The set of possible associations is discrete, thus the (discrete) value of the association posterior distribution can be computed by inspecting all the possible values of the associations [15]. Selecting the maximum probability among the ones in equation (5) gives 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.

IV. EXPERIMENTAL VALIDATION AND RESULTS

During a rehabilitation session motor enhancements are evaluated by extracting quantitative indicators. In particular, measure of RoM, A/A and F/E angles of the fingers, wrist orientation and velocity, finger trajectories provide an indication of the ability of a person to perform a movement [18] [19]. The proposed algorithm has been experimentally tested for tracking the whole hand and extracting the above mentioned kinematic parameters during F/E and A/A movements of the fingers and of the wrist and during reach and grasp action. These are standard movements used for understanding the behaviour of each hand joint during a common rehabilitation session. The paper wants to provide a proof-of-concept of the pose estimation approach for evaluating those parameters; hence, the study is still preliminary and is based on the experimental tests on one subject. The participant (a healthy woman of 34 years old) was seated in front of a table with the right hand placed on it. She was asked to perform reach and grasp actions and finger movements for evaluating joint RoMs paying special attention not to rotate the arm.

Fig. 2 shows the first and final instant of the action performed for evaluating joint trajectories during the reach and grasp action. It is possible to note that some markers disappear due to occlusions or illumination problems. With the proposed pose estimation algorithm, it has been possible to reconstruct the position of each hand joint (on the bottom of Fig. 2). Moreover, the introduction of the hand kinematic model in the tracking algorithm improves robustness with respect to outliers and to the possibility of confusing markers among them due to possible occlusions and rapid movements. The acquired data have also been used for analyzing the wrist behaviour in the reaching phase. The wrist velocity is an indicator of movement fluidity. Fig. 3 shows the wrist trajectory and velocity during the reach and grasp action. Reaching movement is characterized by a single velocity peak, as usual in healthy subjects. Fig. 4 shows finger A/A angles and wrist joint angles behaviour. The plotted results are reasonable, in fact it is possible to note that the measured A/A RoMs respect the values of published data on human beings [20]. The previously listed indicators are also extracted but are not reported for the sake of brevity. In conclusion, the approach could be easily used for patient performance evaluation during a rehabilitation session.

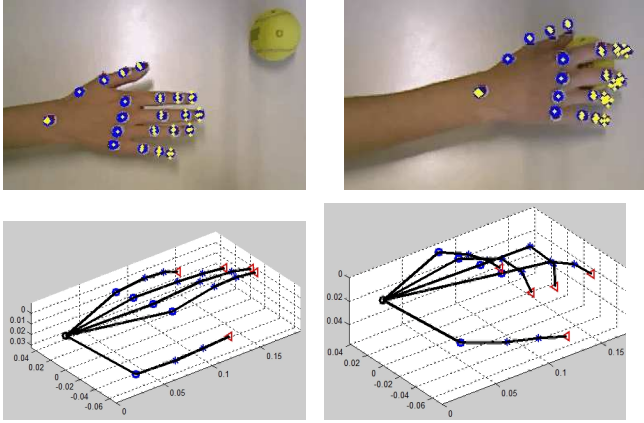


Fig. 2. First and final instant of the grasping action. Human hand (top) and the corresponding reconstructed hand (bottom) configurations are reported. Note that the detector failed with some markers on the fingers due to partial occlusion and shadows, however the pose estimation is still coherent.

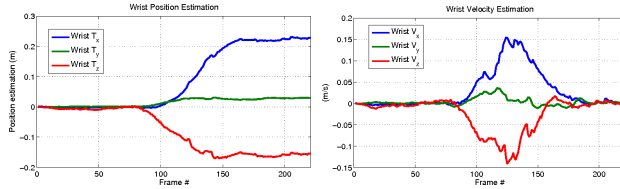


Fig. 3. Position and velocity over time of the wrist during the reaching and grasp experiment.

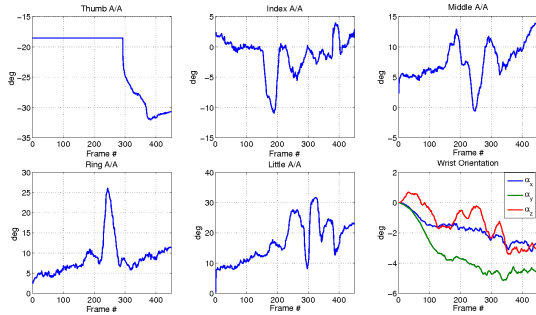


Fig. 4. Fingers A/A motion and wrist angle components.

V. CONCLUSIONS

In this paper a novel low-cost hand pose estimation method for quantitative functional evaluation of patient motor enhancement during hand rehabilitation has been proposed. The non-linear hand pose estimation problem has been solved by using UKF and introducing the hand kinematic model for considering the markers interdependence. Performance indicators such as joint orientation, position, trajectory and velocity have been extracted demonstrating the possible adoption of the proposed approach in rehabilitation. It provides quantitative information about the performed task, such as the measurement of joint motion. Further improvements will be devoted to verify the accuracy of the approach by means of a comparison with a ground truth obtained with an optoelectronic system and to test the approach on real patients.

ACKNOWLEDGMENT

This work was supported in part by the National PRIN Project ROCOCO – COoperative and COLlaborative Robotics (CUP E61J11000300001) and partly by the National PRIN Project HandBot (CUP: B81J12002680008).

REFERENCES

- [1] 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).
- [2] 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*, 128, 152?-158 (2006).
- [3] H. Zhou, H. Hu, "Human motion tracking for rehabilitation—A survey". *Biomedical Signal Processing and Control* vol. 3, pp. 1–18, 2008.
- [4] 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.
- [5] A. Erol, G. Bebis, M. Nicolescu, R.D. Boyle, X. Twombly, "A Review on Vision-Based Full DOF Hand Motion Estimation", *Conf. on Computer Vision and Pattern Recognition*, pp. 75–82, 2005.
- [6] 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.
- [7] 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.
- [8] J.Y. Lin, Y. Wu, T.S. Huang, "3D model-based hand tracking using stochastic direct search method", *Conf. on Automatic Face & Gesture Recognition*, pp. 693–698, 2004.
- [9] 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.
- [10] B. Stenger, A. Thayananthan, P.H.S. Torr, R. Cipolla, "Hand pose estimation using hierarchical detection", *Workshop on Human-Computer Interaction*, 2004.
- [11] C.S. Chua, H.Y. Guan, Y.K. Ho, "Model-based finger posture estimation", *Asian Conference on Computer Vision*, 2000.
- [12] J. Lee, T. Kunii, "Constraint-based hand animation", *Models and Techniques in Computer Animation*, pp. 110?-127, 1993.
- [13] J. Lin, Y. Wu, T. S. Huang, "Modeling the Constraints of Human Hand Motion?", *Proceeding Workshop on Human Motion*, pp. 121–126, 2000.
- [14] S. Parasuraman, C.C.S. Zhen, "Development of Robot Assisted Hand Stroke Rehabilitation System", *Conf. on Computer and Automation Engineering*, pp. 70–74, 2009.
- [15] F. Cordella, F. Di Corato, L. Zollo, B. Siciliano, P. van der Smagt, "Patient performance evaluation using Kinect and Monte Carlo-Based finger tracking", *Conf. on Biomedical Robotics and Biomechanics*, pp. 1967–1972, 2012.
- [16] R. Kandeup, 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.
- [17] E.A. Wan, R. Van der Merwe, "The unscented Kalman filter for nonlinear estimation", *Symposium on Adaptive Systems for Signal Processing, Communications, and Control*, pp. 153–158, 2000.
- [18] 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.
- [19] D. Formica, H.I. Krebs, S.K. Charles, L. Zollo, E. Guglielmelli, N. Hogan, "Passive wrist joint stiffness estimation", *Journal of Neurophysiology*, 2012.
- [20] R.A. Berger, A.P.C. Weiss, "Hand surgery", 2004.