Patient Performance Evaluation Using Kinect and Monte Carlo-Based Finger Tracking

Francesca Cordella, Francesco Di Corato, Loredana Zollo, Bruno Siciliano and Patrick van der Smagt

***Abstract*— The growing use of Virtual Reality (VR) in rehabilitation is justified by a number of advantages, such as an increase of patient motivation, repetitiveness of learning trials, possibility to tailor treatment to individual subject, safety of the environment, quantitative patient improvement assessment, and remote data access. This paper proposes a novel low-cost evaluation method of patient performance in task-oriented hand rehabilitation grounded on two key elements: a Virtual Environment (VE) which the patient has to interact with, and the Microsoft Kinect motion sensing device, which is used to fully interact with the VE and to feed back patient movements in order to perform an off-line analysis. To this purpose, the VE is equipped with a virtual hand and virtual objects the patient has to interact with. In order to make the interaction between patient and VE possible, a robust marker-based finger tracking algorithm has been developed by using Bayesian estimation methods.**

In the proposed framework, the hand movements involved in daily activities are performed off-line by the therapist and are tracked by using the Kinect camera. The estimated hand joint trajectories are provided in input to a virtual hand model developed with the Matlab Virtual Reality Toolbox. The virtual hand reproduces the movements performed by the therapist and the patient is asked to imitate them. User motor improvements can be monitored by the Kinect camera, superimposing the therapist finger trajectories on the patient finger trajectories. The error between the two trajectories can be used for evaluating the patient residual mobility. The proposed system can be easily applied to home-based rehabilitation.

I. Introduction

The main objective of neurorehabilitation is to help patient relearn sensori-motor capabilities by exploiting the plasticity of the neuromuscular system: motor patterns are relearned through repeated execution of predefined movements. Medical studies [1], [2] have evidenced that, as a consequence of a neuromuscular damage, the human motor system has to relearn correct spatio-temporal muscle activations. In order to improve the affected limb function, essential features are patient motivation and intensive, goal-oriented repetitive therapy [3]. In a task-oriented rehabilitation therapy for the upper limb, the patient should perform functional tasks, such as grasping actions most commonly used in daily living activities [4], and/or constrained-induced movements, in which the use of the paretic limb is forced by restraining the contralateral limb [5].

The introduction of VR in rehabilitation helped improve patient motivation, record the patient exercises in order to reduce the human errors in data interpretation, reduce therapy costs and perform tele-rehabilitation, so that the patient is not obliged to always reach the clinical setting [6]. Moreover, the possibility to analyse patient behavioral data enables therapists to evaluate patient performance and, thus, improve rehabilitation efficiency.

VR-based rehabilitation uses different therapeutic approaches, e.g. teaching by imitation, video game-like and exposure therapy. In the former approach, the required motion trajectory is visualized while the patient tries to perform the required action. In [7], the patient arm motion is tracked by a haptic workstation and is reproduced in the virtual environment where the reference motion trajectory is constantly visualized in order to help the patient. In [8] a tool for evaluating patient performance during rehabilitation therapy is proposed: the patient hand centroid is tracked and used as a computer mouse for following reference trajectories made of curves and linear segments.

The objective of this paper is to introduce techniques and tools necessary for realizing a low-cost system for patient performance evaluation during task-oriented hand rehabilitation therapy. Repetitive actions along predefined path used in the robot-assisted rehabilitation therapy showed improvements in stroke patients [9]. Therefore, this paper uses a reference trajectory approach for the rehabilitation therapy. The proposed system is composed of (i) a VE, in which the therapist and the patient hand avatars interact with some virtual objects in order to perform different grasping actions, and (ii) the Kinect® [10] motion sensing device (developed by Microsoft), used for tracking both therapist and patient hand finger movements. The VE inputs are therapist and patient hand joint positions obtained by a marker-based finger tracking algorithm. The novelty of the proposed system is that it allows monitoring the behavior of each joint of the patient hand by comparing it with the reference movements performed by the therapist. The proposed system could be useful for evaluating each hand joint performance. With respect to classical optoelectronic motion analysis devices, the proposed system is very cheap and compact.

During hand tracking, the problems to deal with are essentially due to the segmentation of the hand from the

F. Cordella, B. Siciliano are with PRISMA Lab, Dipartimento di Informatica e Sistemistica, Università di Napoli Federico II, via Claudio 21, 80125 Napoli, Italy {francesca.cordella, bruno.siciliano}@unina.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

F. Di Corato is with Department of Energy and Systems Engineering, Università di Pisa, largo Lucio Lazzarino, 1, 56100 Pisa, Italy francesco.dicorato@for.unipi.it

P. van der Smagt is with Institut für Robotik und Mechatronik, Deutsches Zentrum für Luft- und Raumfahrt (DLR), Wessling, Germany smagt@dlr.de

background and to the large number of hand degrees of freedom (DoFs). Skin color offers a possible way for locating the hand in image sequences, but precludes the tracking of single fingers. As a countermeasure, it is necessary to extract some features, such as points, contours, or silhouettes [11] to be tracked. A commonly used approach for tracking objects characterized by a high number of DoFs, such as the human hand, is to use a prior 3D model on which the observed data are projected. The best match between the image feature observations and the projected 3D model renders the joint angle values. Therefore, it is possible to formulate the hand tracking problem as an optimization problem [12], [13]. In order to solve the optimization problem, sequential Monte Carlo methods, such as particle filters, have been used [14], [15], [16], [17].

In this work, a Monte Carlo-based approach for tracking the hand joints has been developed. In particular, a set of markers have been placed on the hand joints of the therapist and the marker detection during motion has been made by using a blob-detection algorithm. A central problem appeared to be tracking the markers along the image video sequence, taking into account the presence of noise, outliers due to clutters and the possibility that markers appear and disappear during motion. Bayesian estimation techniques with suitable filtering techniques have been used to this effect. An advantage of the developed approach is mostly the reduced computational cost and the ease of filter implementation and tuning. Moreover, as it will be clear in the following, there is no need of prior training the entire system. By using the Kinect, it has been possible to perform the visual analysis of human hand motion and to record the hand joint positions during movements in a reliable and repeatable way. The obtained finger joint positions and trajectories during hand movements have been used as input to a virtual hand that reproduces the therapist movements. According to the teaching-by-imitation approach, the patient visualizes the reference trajectory and tries to follow it. The movements have been monitored by a Kinect system and hand trajectories have been given as input to another virtual hand model. The error between the two trajectories is computed in order to measure patient performance and to improve rehabilitation therapy.

The paper is structured as follows: Section II is devoted to a brief description of the Kinect technology; in Section III the finger tracking algorithm is explained; in Section IV, the application of the developed algorithm to the field of graphical interfaces for hand rehabilitation devices is presented. Finally, conclusions and directions for future work are proposed in Section V.

II. The Kinect motion sensing device

The Kinect (Figure 1) is a motion sensing device consisting of an InfraRed (IR) laser emitter, an IR camera for measuring depth information, and an 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.

Fig. 1. The Kinect motion sensing device.

Depth is evaluated by using a CMOS sensor device, produced by PrimeSense, by using the light coding method [18].

The device can supply depth measurements of images formed by the RGB camera when the vision system is appropriately calibrated.

LibFreenect library by OpenKinect has been used in order to make the Kinect work on the PC.

III. Hand joint tracking algorithm

In order to obtain information about the hand joint positions during motion, colored markers have been placed on the human hand as shown in Fig. 2, taking into account the literature [19], [20], [21], [22] and the necessity of tracking hand joints. The markers are made of paper, have diameter 1.2 cm, and consist of a central blue circle surrounded by a white ring. This external white part is useful to make the detection algorithm robust when finger projections overlap.

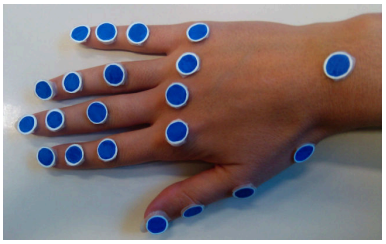


Fig. 2. Protocol used for marker positioning.

Marker detection and tracking are the core problems in reconstructing the hand joint positions during hand movements. In this work, the ideal perspective camera model [23] is used to project 3D scene points into their corresponding image–space 2D features.

A. Detection

The first step in the tracking algorithm is the detection of distinctive features, to extract our regions of interest from the image. This is done by detecting the colored markers in each input frame. Blobs whose histogram is as close as possible (in the sense of Bhattacharyya similarity coefficient) to the reference color histogram are extracted. In order to improve and speed up the extraction process, the image is thresholded, thus yielding to a binary image where the regions whose histogram is closer to the reference one are represented by white pixels. The threshold value is kept constant along the whole video and it has been set up by

making some experimental tests over the available video set. Then, a connected component labeling algorithm is used: the image is scanned by rows reading the gray level of each pixel. If the gray level of a pixel exceeds the given threshold (i.e., the pixel is identified as a white pixel), its position is stored, because it is possible that such pixel belongs to a candidate blob. The above procedure is done for each row. If two white pixel sequences on consecutive rows are neighboring, they are considered members of the same blob and a same label is assigned to them. Only blob candidates composed by more than N_p pixels are considered valid. Once the blobs on the scene have been determined, the 2D coordinates of every blob center are computed. Figure 3 shows an example of the algorithm output.

B. Tracking

Once the center of each colored marker has been detected, it is necessary to track the markers during the whole video. Tracking is the process that, assigned a frame at a given time, aims at making coherent correspondences between visible markers on successive frames, while time is passing. Since a robust blob extraction has proved to be difficult, in the literature, some detection ambiguities could raise. In particular, the blob detector could fail in some regions due to local illumination changes or shadows. This can lead to the inability to detect some visible markers or to the possibility that the features set may be contaminated by outliers. Moreover some markers may disappear from the field of view due to hand and finger movements. For these reasons, a robust tracking scheme must be implemented. In the following, the developed algorithm is illustrated in detail.

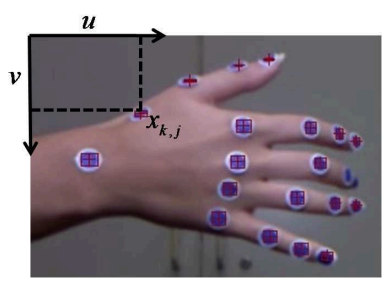


Fig. 3. Frame k of the video sequence: output of the blob detection algorithm. The center of each marker is outlined with a red cross. The image space coordinates of the j-th marker, $x_{k,j}$, are outlined. Notice that one marker disappears from the field of view and its projection is temporarily not available. Moreover, one visible marker is not recognized by the blob detector. The tracking algorithm is required to be robust with respect to such kind of situations.

Let N markers be given, where N is the maximum number of the markers expected in the scene. In the absence of a hand kinematic model, the best we can do is to assume the marker projections $j = 1, 2, \ldots, N$ to have linear-state Markovian dynamic and measurement models, driven by zero-mean white noise. The latter assumption is usually considered realistic within the Computer Vision community.

If we write $u_{k,j}, v_{k,j}$ as the image space coordinates of the j-th marker at the time when the k-th frame has been processed, the marker state can be defined as

$$x_{k,j} = \begin{bmatrix} u_{k,j} & v_{k,j} \end{bmatrix}^T, \tag{1}$$

and thus

$$\begin{cases} x_{k,j} = x_{k-1,j} + w_{k,j} \\ y_{k,j} = x_{k,j} + \nu_{k,j}, \end{cases} \tag{2}$$

where

$$w_{k,j} \sim N(0, Q_j) \tag{3}$$

$$\nu_{k,j} \sim N(0, R_j) \tag{4}$$

are white Gaussian noises with null mean value and covariance matrix Q_j and R_j, respectively. Figure 3 shows the projection $x_{k,j}$ of the j-th marker on the image plane corresponding to the k-th frame, according to the pinhole model. The output of the blob detection algorithm, for the k-th image, is a random sequence of M_k measurements $\mathbf{Y}_k = \{y_{k,1}, y_{k,2}, \ldots, y_{k,M_k}\}$ of blob candidates. In general, the condition $M_k \neq N$ will hold, which means that sequence $\mathbf{Y}_k$ contains projections of visible markers and clutters, as in the following example:

$$\begin{aligned} y_{k,1} &\rightarrow \text{marker } 6 \\ y_{k,2} &\rightarrow \text{marker } 7 \\ y_{k,3} &\rightarrow \text{clutter} \\ &\vdots \\ y_{k,M_k} &\rightarrow \text{marker } h. \end{aligned}$$

In designing the tracker, the following main hypotheses have been considered:

- The associations between measurement h and marker j or with a clutter cannot be a priori decided and has to be set.
- Each sequence of measurements for each frame can be considered conditionally independent on every other sequence in the past.
- Once the current sequence of associations has been defined, it can be considered conditionally independent of the past history of associations.

Given the intrinsic randomness of the problem and the discussed ambiguities, the best way to solve it is using a general probabilistic approach, via robust Bayesian filtering [24], [25], [26].

The current version of the tracking algorithm does not include an explicit kinematic hand model, and thus the markers are treated as mutually independent entities. Measurements are serially processed, leading to a simplification of the tracking algorithm. Obviously the lack of a kinematic model has to be compensated for by a more accurate tuning required for the filter parameters and for the noise matrices of the markers and measurement models. This has been found necessary to avoid swapping situations between markers in some configurations where their projections were very close to each other or overlapping.

As can be seen in the video output, the model-free approach leads to acceptable and accurate tracking results. Higher accuracy can still be obtained by including a model.

With the above assumptions, let $a_k \in \{0, 1, \ldots, N\}$ be a discrete indicator variable, defined as follows:

$$a_k = \begin{cases} 0 & \text{if } y_{k,i} \text{ is an outlier} \\ j & \text{if } y_{k,i} \text{ is associated to marker } j \end{cases} \tag{5}$$

where $y_{k,i}$ is the i-th measurement at the k-th frame. According to this definition, the markers model can be converted into a new stacked model. By introducing the augmented markers vector

$$\mathbf{x}_k = \left[\begin{array}{cccc} x_{k,1}^T & x_{k,2}^T & \ldots & x_{k,N}^T \end{array} \right]^T \tag{6}$$

and in view of (5), the following Conditional Dynamic Linear Model (CDLM) [26] can be defined:

$$\begin{cases} \mathbf{x}_k = F(a_k = j)\, \mathbf{x}_{k-1} + E(a_k = j)\, \mathbf{w}_k \\ y_{k,j} = H(a_k = j)\, \mathbf{x}_k + \nu_{k,j}, \end{cases} \tag{7}$$

where $F(a_k = j) \in \mathbb{R}^{(2N)\times(2N)}$ and $E(a_k = j) \in \mathbb{R}^{(2N)\times(2N)}$ are selection matrices with $\mathcal{I}_{2\times 2}$ in the (2×2) block (j, j) and 0 elsewhere and $H(a_k = j) \in \mathbb{R}^{2\times(2N)}$ is a selection matrix with $\mathcal{I}_{2\times 2}$ in the $(1, j)$ block and 0 elsewhere. By conditioning the system on a_k, the CDLM (7) becomes a linear Gaussian dynamic model. Thus, the state of the j-th marker can be updated once the measure corresponding to that marker has been selected. In this work, a "per-measurement" association is used. In the framework of optimal Bayesian filtering, the tracking problem can be defined by estimating the filtering (posterior) distribution

$$p(\mathbf{x}_k | y_{1:k}) = p(x_{k,1}, ..., x_{k,N} | y_{1:k}) \tag{8}$$

being the estimation of the current augmented marker state, given the measurement history up to the current time. With the introduction of the latent variable a_k, the posterior probability $p(\mathbf{x}_k | y_{1:k})$ can be evaluated by marginalization with respect to a_k of the joint probability $p(\mathbf{x}_k, a_k | y_{1:k})$, that is:

$$\begin{aligned} p(\mathbf{x}_k | y_{1:k}) &= \sum_{a_k} p(\mathbf{x}_k, a_k | y_{1:k}) \\ &= \sum_{a_k} p(\mathbf{x}_k | a_k, y_{1:k})\, p(a_k | y_{1:k}). \end{aligned} \tag{9}$$

The first equality in (9) is obtained by applying the definition of marginalization; the second equality is obtained via Bayes' rule on the joint posterior probability $p(\mathbf{x}_k, a_k | y_{1:k})$.

In the last equality of (9) two terms are highlighted:

- $p(a_k | y_{1:k})$ is the posterior distribution of the data association, whose shape is, in general, not known a priori, depending on the detection algorithm. A weighted m-sample approximation has to be determined by using Monte Carlo Sequential Importance Resampling (SIR) techniques. In this work it has been found convenient to use particle filtering as SIR approximation, making it possible to represent the association posterior with a set of particles which are updated and reweighted recursively.
- $p(\mathbf{x}_k | a_k, y_{1:k})$ is the posterior (updated) distribution of the markers projections, conditioned on the association a_k. Given the definition of the CDLM (7), it is known that such model becomes a linear Gaussian dynamic model, once the data association is fixed: that is, the posterior $p(\mathbf{x}_k | a_k, y_{1:k})$ can be solved using a Kalman filter.

From the above analysis, the following relationship holds:

$$p(\mathbf{x}_k | y_{1:k}) \approx \sum_{i=1}^{m} w_k^i\, N\left(\hat{\mathbf{x}}_k\left(a_k^i\right),\, P_k\left(a_k^i\right)\right) \tag{10}$$

where $\hat{\mathbf{x}}_k\left(a_k^i\right)$ and $P_k\left(a_k^i\right)$ are respectively the mean vector and the error covariance matrix of the Kalman filter associated with the i-th sample (which infers the association a_k^i). The notation $N(\cdot, \cdot)$ indicates the multivariate normal distribution of order 2. According to (10), a set of m particles can be defined, which contains the augmented state mean $\hat{\mathbf{x}}_k^i$, the error covariance matrix P_k^i and the weight w_k^i associated with each particle:

$$\mathcal{S}_k = \left\{ \left(\left\{ \hat{\mathbf{x}}_k^1,\ P_k^1 \right\}, w_k^1 \right),\ \left(\left\{ \hat{\mathbf{x}}_k^2,\ P_k^2 \right\}, w_k^2 \right), \ldots, \left(\{ \hat{\mathbf{x}}_k^m,\ P_k^m \}, w_k^m \right) \right\} \tag{11}$$

Equation (9) together with (10) give a direct information about how to solve the filtering distribution $p(\mathbf{x}_k | y_{1:k})$: by generating all the possible hypotheses $a_k = i$, $i = 0, 1, \ldots, N$ and evaluating each of them together with the current observation (by running a Kalman filter), it is possible to evaluate the most likely hypothesis a_k^* which gives the highest score. This will give the most likely marker state according to (10).

By using Bayes' rule, the posterior $p(a_k | y_{1:k})$ can be decomposed as follows:

$$p(a_k | y_{1:k}) \propto p(y_k | a_k, y_{1:k-1})\, p(a_k) \tag{12}$$

where $p(y_k | a_k, y_{1:k-1})$ is the likelihood of the measurements and $p(a_k)$ is the association prior, which is assumed to be independent of the previous measurements. The latter is determined by the a priori knowledge of clutter and marker association event probabilities.

In the case of association hypothesis with marker j:

$$\begin{aligned} &p(y_k | a_k, y_{1:k-1}) = \\ &\quad \int N\left(H\left(a_k^i = j\right) \mathbf{x}_k, R_j\right) N\left(\hat{\mathbf{x}}_k^{i-},\ P_k^{i-}\right) d\mathbf{x}_k(j) \end{aligned} \tag{13}$$

where $N\left(\hat{\mathbf{x}}_k^{i-},\ P_k^{i-}\right)$ is the Kalman prediction step for the i-th particle, being $\mathbf{x}_k^{i-}$ the prediction of the marker positions and $\mathbf{P}_k^{i-}$ the prediction of the error covariance. Moreover, $N\left(H\left(a_k^i = j\right) \mathbf{x}_k, R_j\right)$ is the probability density function of the measurement y_k, assuming that it has been generated by the j-th marker. In the case of clutter association,

$$\begin{aligned} &p(y_k | a_k = 0, y_{1:k-1})\, p(a_k = 0) = \\ &\qquad = \frac{1}{RES_u \times RES_v}\, p(a_k = 0) \end{aligned} \tag{14}$$

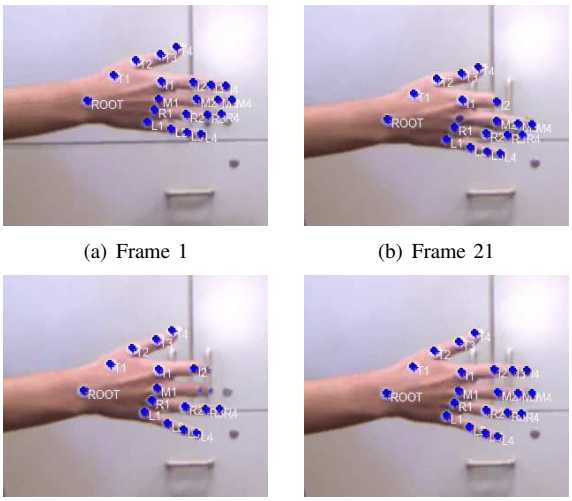
where RES_u, RES_v are the numbers of pixels in the u, v direction (image resolution) and $p(a_k = 0)$ is known.

Once the expressions of the association prior and of the measurement likelihood have been derived, the importance sampling of the optimal distribution for the association

problem can be computed for each particle in the set $\mathcal{S}$. Thus, an approximation of the posterior

$$p\left(\mathbf{x}_k|y_{1:k}\right) \approx \sum_{i=1}^{m} w_k^i \, N\left(\hat{\mathbf{x}}_k\left(a_k^i\right), \, P_k\left(a_k^i\right)\right)$$

can be found. The algorithm is explained in Tab. I.



(a) Frame 1 (b) Frame 21

(c) Frame 50 (d) Frame 55

Fig. 4. Results of the proposed filter based tracking.

Figure 4 shows a sequence of hand movements where the blue markers on the hand have to be tracked. At the beginning, there are 25 markers in the scene (Frame 1), at Frame 21 the marker on the MCP joint of the middle finger and the markers on the TIP and DIP joints of the index finger disappear; at Frame 50 the marker on the middle finger MCP joint reappears, but markers on DIP and TIP joint of the index finger are still not present and markers on PIP, DIP and TIP joints of the middle finger have disappeared. In Frame 55 all the markers reappear. Those results prove the capacity of the algorithm to track the markers in spite of the disappearing marker problem.

IV. Representation of finger trajectories in VE

The 3D position of each visible marker (with respect to the camera) can be reconstructed by multiplying the marker pixel coordinates coming from the tracking algorithm by the RGB camera calibration matrix and by the measured depth value. The tracked finger joint positions of the therapist hand have been given in input to a VE developed by means of the Matlab Virtual Reality Toolbox. The VE is made of a virtual human hand, designed by taking into account the physical characteristics of the patient hand, and by different objects suitably chosen for rehabilitating the user hand to perform daily life activities.

With the finger tracking algorithm described in Sect. III we are able to track the pinch grasp and the tripod grasp. In fact, for the proper operation of the algorithm, it is necessary that, at the first and last frame, all the markers are visible. In such a way, an interpolation process can reconstruct the movement of each single joint during the whole grasping action, even if in the reaching phase some markers disappear from the image.

TABLE I

PSEUDO-CODE OF THE TRACKING ALGORITHM

INIT

- Marker state vector initialization: $\mathbf{x}_0 = \left[\begin{array}{ccc} x_{0,1} & \dots & x_{0,N} \end{array}\right]^T$
- Particles initialization:
  $\mathcal{S}_0 = \{(\{\mathbf{x}_0^1, P_0^1\}, w_0^1), ..., (\{\mathbf{x}_0^m, P_0^m\}, w_0^m))\}$

CICLE

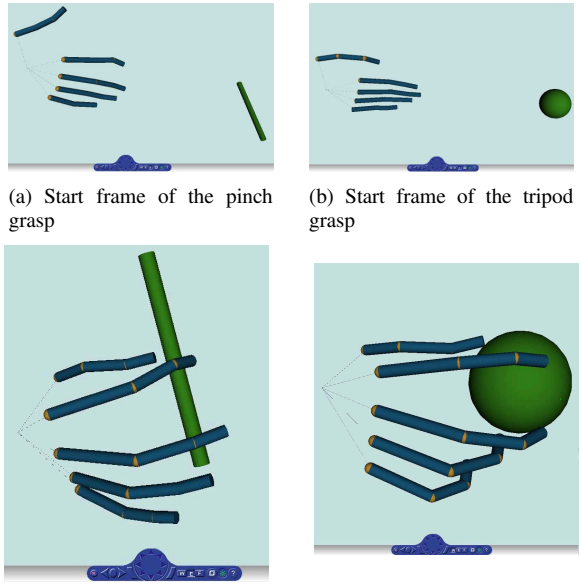
- For each measure $(n = 1 : M_k)$:
  - For each particle $(i = 1 : m)$:
    - ∗ **MKF prediction:** For each hypothesis a_k, run a Kalman prediction step on the couple $\{x_k^i(a_k), P_k^i(a_k)\}$
    - ∗ Calculate the likelihood of the current measure with respect to predictions and to possible associations:
      $$\begin{cases} p_0^i = p(y_k^n|a_k = 0, \mathbf{x}_k^i, \mathbf{y}_{1:k-1})p(a_k = 0) \\ p_1^i = p(y_k^n|a_k = 1, \mathbf{x}_k^i, \mathbf{y}_{1:k-1})p(a_k = 1) \\ \vdots \\ p_N^i = p(y_k^n|a_k = N, \mathbf{x}_k^i, \mathbf{y}_{1:k-1})p(a_k = N) \end{cases}$$
    - ∗ Sample a new association hypothesis a_k^i (particle) from the importance distribution $p\left(a_k|y_k^n\right)$:
      draw $a_k^i = 0$ with normalized probability p_0^i
      draw $a_k^i = 1$ with normalized probability p_1^i
      $\vdots$
      draw $a_k^i = N$ with normalized probability p_N^i
    - ∗ **MKF update:** if $a_k^i \neq 0$, run a Kalman correction step, using the new sampled association variable a_k^i: $\{x_k(a_k^i), P_k(a_k^i)\}$;
    - ∗ compute the new particle weight: $w_k^i = w_{k-1}^i \sum_h p_h^i$;
  - End for each particle
  - Resample, if necessary, the particle set;
  - Extract the couple $\{x_k^i, P_k^i\}$ from particles with highest probability.
- End for each measure.

In case of power grasp and lateral grasp, when the user grasps the object, some markers may be hidden by other hand parts or by the object itself. In order to be able to record the trajectories followed by both therapist and patient in such grasp types, it is necessary to introduce the kinematic model of the hand in the tracking algorithm. Future work will be oriented to this direction.

In Fig. 5, the start and final frames of the grasping action simulation carried out by the therapist are shown, in case of tripod grasp and pinch grasp. Only the fingers are shown since in the precision grasp the palm is not involved. In the future extension of the proposed approach to power and lateral grasps, the VE will be completed with the integration of the palm. When looking at the movements in Fig. 5, the patient should try to reproduce them while the Kinect is tracking his/her motion. The trajectories implemented by the patient will be recorded and supplied to the VE. The user can have a feedback on his/her movements by looking at the display. Patient performance will be evaluated by quantitatively comparing patient and therapist trajectories.

V. Conclusions

Basic techniques and tools for the realization of a low-cost system for evaluating patient performance in hand rehabilita-



(a) Start frame of the pinch grasp

(b) Start frame of the tripod grasp

(c) Final frame of the pinch grasp

(d) Final frame of the tripod grasp

Fig. 5. Simulation of the grasping action performed by the therapist, in case of pinch and tripod grasp.

tion have been proposed. The Kinect motion sensing device has been used in order to track therapist and patient hand movements during predefined grasping tasks. In particular, an algorithm based on Bayesian estimation theory and on filtering techniques has been implemented. The resulting hand joint trajectories have been recorded and provided in input to a VE in order to move a virtual hand. The patient motor improvements will be monitored by another Kinect camera, superimposing the therapist finger trajectories on the patient finger trajectories. Patient performance could be evaluated on the basis of the error between trajectories. At the moment, only two of four basic grasp types can be performed. Future work will be devoted to introduce the hand kinematic model in the tracking algorithm, in order to apply the system to all types of grasp. Further, the system will be tested on patients.

VI. Acknowledgements

This work was partly supported by the European Project THE—The Hand Embodied (FP7-IST-248257), partly by the National PRIN Project OPENHAND—OPEN neuroprosthetic HAND platform for clinical trials (prot. PMZT2Z, CUP B81J10000050008) and partly by the National PRIN Project ROCOCÒ—COoperative and COllaborative RObotics (prot. 20094WTJ29_003, CUP E61J11000300001).

References

[1] M.W. O'Dell, C.D. Lin, V. Harrison, "Stroke rehabilitation — Strategies to enhance motor recovery", *Annual Review of Medicine*, vol. 60, pp. 55–68, 2009.

[2] R. Teasell, N.A. Bayona, J. Bitensky, "Plasticity and reorganization of the brain poststroke", *Topics in Stroke Rehabilitation*, vol. 12, no. 3, pp. 11–26, 2005

[3] A. Merians, D. Jack, R. Boian, M. Tremaine, G.C. Burdea, S. Adamovich, M. Recce, H. Poizner, "Virtual reality-augmented rehabilitation for patients following stroke", *Physical Therapy*, vol. 82, no. 9, pp. 898–915, 2002.

[4] Y.Y. Huang, K.H. Low, H.B. Lim, "Initial analysis of EMG signals of hand functions associated to rehabilitation tasks", *IEEE Conference on Robotics and Biomimetics*, pp. 530–535, 2009.

[5] M. Rensink, M. Schuurmans, E. Lindeman, T.. Hafsteinsdottir, "Task-oriented training in rehabilitation after stroke: systematic review", *Journal of Advanced Nursing*, vol. 65, no. 4, pp. 737–754, 2009.

[6] G.C. Burdea, "Virtual rehabilitation - Benefits and challenges", *Methods of Information in Medicine*, vol. 42, no. 5, pp. 519–523, 2003.

[7] M.K. Holden, E. Todorov, "Use of virtual environments in motor learning and rehabiliation", *Handbook of Virtual Environments: Design, Implementation and Applications*, chap. 49, pp. 999–1026, 2002.

[8] G. Boato, N. Conci, M. Daldoss, F.G.B. De Natale, N. Piotto, "Hand tracking and trajectory analysis for physical rehabilitation", *IEEE Internation Workshop on Multimedia Signal Processing*, pp. 1–6, 2009.

[9] H.I. Krebs, N. Hogan, M.L. Aisen,B.T. Volpe, "Robot-aided neurorehabilitation",*IEEE Transactions on Rehabilitation Engineering*, vol.6, pp. 75-87, 1998.

[10] "Kinect motion sensing device webpage", *http://mirror2image.wordpress.com/2010/11/30/how-kinect-works-stereo-triangulation*.

[11] J.J. Kuck, T.S. Huang, "Vision-based hand modeling and tracking for virtual teleconferencing and telecollaboration", *IEEE International Conference on Computer Vision*, pp. 666–671, 1995.

[12] S. Lu, D. Metaxas, D. Samaras, J. Oliensis, "Using multiple cues for hand tracking and model refinement", *IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 443–450, 2003.

[13] J. Wu, T.S. Huang, "Capturing articulated human motion: A divide-and-conquer approach", *IEEE International Conference on Computer Vision*, pp. 606–611, 1999.

[14] M. Bray, E. Koller-Meier, L. Van Gool, "Smart particle filtering for 3D hand tracking", *IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 675–680, 2004.

[15] M. Isard, A. Blake, "CONDENSATION-conditional density propagation for visual tracking", *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.

[16] J.Y. Lin, Y. Wu, T.S. Huang, "3D model-based hand tracking using stochastic direct search method", *IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 693–698, 2004.

[17] W.Y. Chang, C.S. Chen, Y.P. Hung, "Appearance-guided particle filtering for articulated hand tracking", *IEEE International Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 235-242, 2005.

[18] K Khoshelham, "Accuracy analysis of Kinect depth data", *ISPRS Workshop Laser Scanning*, 2011.

[19] F.C. Su, Y.L. Chou, C.S. Yang, G.T. Lin, K.N. An, "Movement of finger joints induced by synergistic wrist motion", *Clinical Biomechanics*, vol. 20, pp. 491–497, 2005.

[20] N. Miyata, M. Kouchi, T. Kurihara, M. Mochimaru, "Modeling human hand link structure from optical motion capture data", *IEEE/RSJ Internation Conference on Intelligent Robots and Systems*, vol. 3, pp. 2129–2135, 2004.

[21] I. Carpinella, P. Mazzoleni, M. Rabuffetti, R. Thorsen, M. Ferrarin, "Experimental protocol for the kinematic analysis of the hand: definition and repeatability", *Gait & Posture*, vol. 23, pp. 445–454, 2006.

[22] C.D. Metcalf, S.V. Notley, P.H. Chappell, J.H. Burridge, V.T. Yule, "Validation and application of a computational model for wrist and hand movements using surface markers", *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 1199–1210, 2008.

[23] E.Trucco, A. Verri, *Introductory Techniques for 3D Computer Vision*, Prentice Hall, 1998.

[24] S. Sarkka, A. Vehtari, J. Lampinen, "Rao-blackwellized particle filter for multiple target tracking", *7th International Conference on Information Fusion*, vol. 8, no. 1, pp. 2-15, 2007.

[25] S. Sarkka, A. Vehtari, J. Lampinen, "Rao-blackwellized Monte Carlo data association for multiple target tracking", *International Conference on Information Fusion*, vol. 7, pp. 583-590, 2004.

[26] R. Chen, J.S. Liu, "Mixture Kalman filter", *Journal of the Royal Statistical Society: Series B*, vol. 62, no. 3, pp. 493-508, 2000.