Toward an affordable and user-friendly visual motion capture system

V. Bonnet, N. Sylla, A. Cherubini, A. Gonzáles, C. Azevedo Coste, P. Fraisse, and G.Venture

Abstract— **The present study aims at designing and evaluating a low-cost, simple and portable system for arm joint angle estimation during grasping-like motions. The system is based on a single RGB-D camera and three customized markers. The automatically detected and tracked marker positions were used as inputs to an offline inverse kinematic process based on biomechanical constraints to reduce noise effect and handle marker occlusion. The method was validated on 4 subjects with different motions. The joint angles were estimated both with the proposed low-cost system and, a stereophotogrammetric system. Comparative analysis shows good accuracy with high correlation coefficient (r= 0.92) and low average RMS error (3.8 deg).**

I. Introduction

Analysing and quantifying human motor act start with measuring motions as accurately as possible. Motion capture systems usually refer to stereophotogrammetric systems that are very accurate but also costly, require expert skills, and are not easily portable. Recently, inertial measurement units (IMUs) became a popular tool to quantify human motion outside the laboratory [1]. However, despite strong advantages in estimating global spatio-temporal parameters [1], the accurate estimate of 3D joint angles (JA) with IMU remains difficult due to non-linear drift occurring when integrating measured accelerations and angular velocities [2]. Consequently, to the best of our knowledge, no system can provide the absolute position of the sensor without a prior calibration phase and a strongly dependent model based approach [1]. These drawbacks might explain why such systems are rarely used in common clinical applications or for in-home rehabilitation programs. Recently, a new type of affordable sensor called RGB-D camera is bringing forward a number of breakthroughs in human motion analysis and robotics [3] by making human motion tracking and quantification available to a large public. The Kinect sensor (KS, Microsoft®) provides both RGB and depth information and embedded human motion tracking software. Several studies have analyzed the accuracy of this solution for rehabilitation purposes [4-5]. Their general conclusion is that the KS embedded marker-less methods are not reliable and accurate enough for quantitative evaluation of human motion. Other groups have developed their own algorithms using one or multiple KSs [6]. Marker-less pose estimation from multi-view video has been a long-standing problem in computer vision, and mature solutions exist [Buys]. However, they often require a complex calibration phase and a large volume to locate cameras. Besides, proposed solutions are not validated using a gold standard system, and thus are not accessible to the clinical community. Additionally, a marker-less skeleton-tracking algorithm will most probably fail if one or several individuals are in close interaction with the tracked patient. Finally, other researchers have proposed to fuse camera(s) and IMU data in order to handle occlusions and improve the general accuracy [7]. These approaches are often based on Kalman filter and appear to provide consistent JA [7]. However, the use of one or several additional IMU raises the overall cost of the system and synchronization between different low-cost elements, having a non-constant frequency rate, is not always straightforward [8].

Applications requiring a low-cost user-friendly motion capture system are numerous. Single arm grasping while seated for example, has been extensively used in post-stroke rehabilitation. This task can be performed in the plane or in 3D, and the variables of interest for clinicians are the wrist Cartesian trajectory and the evolution of the shoulder and elbow joints [9]. Low-cost solutions also allow patients to train in their home environment, so that the burden of mobilizing clinicians can be reduced. In industrial ergonomics, such as in car manufacturing, the need for portable systems that can be used directly in production lines is also increasing to evaluate the efforts endured by workers on a long-term basis, as musculoskeletal disorders can appear in worker performing repetitive over-head manipulation tasks [10].

In this context, a marker-based method to estimate arm's JA is proposed using a set of customized markers and a KS for arm motions. The accuracy and robustness of the proposed approach is validated with four human subjects over different motions using stereophotogrammetric data.

II. Method

A. Mechanical model

The human arm was modelled as a serial chain composed of two rigid segments articulated by four hinge joints (Fig. 1). The JA vector $\boldsymbol{\theta}=[\theta_1; \theta_2; \theta_3; \theta_4]^T$ consists in the shoulder flexion/extension, the shoulder abduction/adduction, the shoulder rotation and the elbow flexion/extension. Very low amplitudes were observed at the wrist joints, thus they were not considered for the investigated task. The Denavit -Hartenberg notation was used to calculate analytically the 3D positions of the elbow joint ${}^0P_4$ and of the hand ${}^0P_5$ in the shoulder frame supposed to be fixed. Analytical inverse dynamics model, was calculated to estimate the joint torques vector, $\boldsymbol{\Gamma}$. Inertial parameters were obtained from anthropo-

Research supported by the Japan Society for Promotion of Science and partially by the author's respective institutions.

V. Bonnet and G. Venture are with the GV laboratory, Tokyo University of Agriculture and Technology, Tokyo, Japan. (bonnet.vincent@gmail.com).
A. Cherubini, A. Gonzáles, N. Sylla and P. Fraisse are with the LIRMM, University of Montpellier 2, Montpellier, France.
N. Sylla is also with PSA Peugeot Citroen, Velizy Villacoublay, France.
C. Azevedo-Coste is with INRIA, Montpellier, France.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

metric tables.

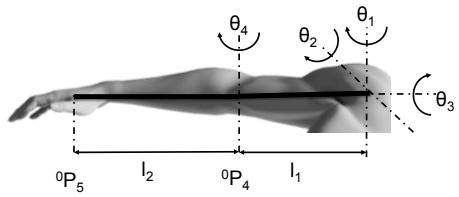

Fig. 1 Model of the human arm composed of 4 hinge joints.

B. Visual tracker

A Kinect sensor was located at the height of the subject's pelvis and roughly perpendicular with the sagittal plane at approximately 1.5 m. In ideal conditions, the resolution of the depth information is of 3 mm [3]. OpenNI middleware [3] was used to access the KS calibrated data expressed in the world frame coordinate. The embedded camera calibration process of OpenNi was used for the sake of simplicity since KS lenses exhibit low distortion [3]. As the native skeleton tracking of the KS performs poorly to JA estimation, a customized marker based motion capture system was developed. Three markers $Mk_{1,2,3}$, 5x5 cm, were printed on standard semi-rigid paper and attached to the body using double-side tape. This was chosen to maintain the accessibility of the system, price and ease-of-use, to comply with the material reflectivity sensitivity of the KS, and to be able to detect markers in spite of low resolution. The markers were selected from a markers list [11] as presented on Fig. 2 for their uniqueness, minimizing the risk of false positive. Accurate 3D pose estimation of such markers is possible with algorithms already used in virtual-reality applications [11]. However, they require a small motion of a large size marker and thus did not perform well in our case. A cascaded object detector based on the Viola-Jones algorithm [12] was trained over approximately ten thousands positives images per markers using videos collected during the system development. The images were collected from different motions in different light and environment conditions. Fig. 2 presents an overview of the tracking-detection system and of the marker trajectories processing. Once the markers are detected in the RGB image, the defined region of interest is used to extract the features of interest using a speeded up robust feature extraction [13]. A Kanade-Lucas-Tomasi (KLT) [14] feature-tracking algorithm tracked the detected markers over time. The KLT tracking algorithm, that estimates the consistency of affine transformations between two consecutive sets of tracked points and performs a multi-resolution tracking, is known to work well for tracking objects that exhibit a strong visual texture, under reasonable shape and lighting changes. In the considered application, the markers may be lost due to an important out of plane rotation. A marker is considered lost if the number of corresponding tracked points is inferior to three. When a marker is lost, the detector is re-run to re-initialize the KLT tracker. In case of detection fail, the human supervisor is asked to manually select the markers. If no marker is available the following image will be loaded until the marker can be detected again. Due to the KS's depth map estimation technology, no-measurement-points may be available in some regions in case of projection shadow.

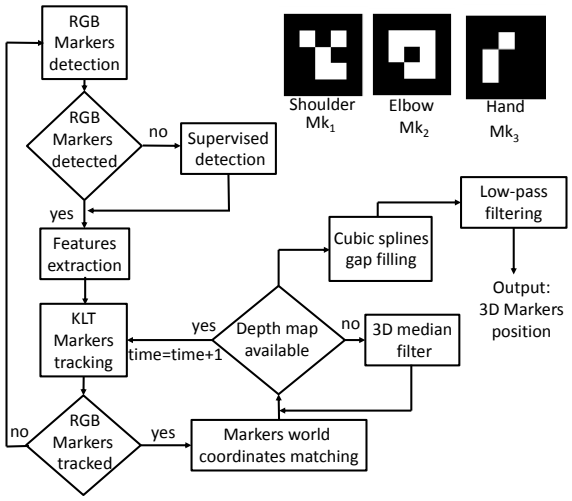

Fig. 2 Overview of the tracking-detection system and marker trajectories processing

In order to deal with a lack of depth information and noise at body segment edge, a median filter is used in 3D to fill the possible holes in the depth map. This filter is run only when depth information of a tracked marker is missing and it averages a 3x3 pixels matrix without taking into account the non-measured-depth points in the calculation. Since the KS does not provide a fixed sample rate, data were time stamped and resampled at a constant sampling rate of 30 Hz using a first order linear interpolation. Cubic splines gap filling was used over missing data lasting for less than 5 samples and all markers trajectories were filtered using a low-pass Butterworth zero-phase digital filtering.

C. Optimal joint angles calculation

Starting from the 3D positions of each marker, it is possible to calculate the JA of the arm model presented in Fig.1 using its inverse geometrical model. Unfortunately, if the markers' 3D positions are not well estimated, model singularities may occur. To cope with this issue, other inverse kinematics methods, inspired by the robotics field, enabling real-time estimate, and handling redundancy or singularities have been proposed [2]. However, handling occlusion cases, i.e. lack of data, or solving biomechanical constraints in real-time is still necessary. For these reasons, an off-line optimization process aiming at estimating the arm JA while managing physiological constraints, (cts), and occlusions (oc) was developed. If all markers positions are available, the optimization process aims at finding the four JAs for every sample of time, *t*, that minimizes the normalized weighted least square differences between measured and estimated positions of the markers as follows:

$$J = \frac{1}{N}\sum_{t=1}^{N}\left(\alpha \frac{\left({}^{0}Mk_1(t) - {}^{0}P_4(\theta(t))\right)^2}{\Delta M_1} + \beta \frac{\left({}^{0}Mk_2(t) - {}^{0}P_5(\theta(t))\right)^2}{\Delta M_2}\right) \quad (1)$$

where $\Delta M_{1,2} = \max(Mk_{1,2}) - \min(Mk_{1,2})$ and N are a normalization term and the number of considered samples, respectively. The shoulder marker position Mk_3 was sub-

tracted to the positions of the other markers $Mk_{1,2}$ in order to comply with the arm system of reference. α and β are weights giving more importance to the tracking of the elbow or of the hand marker, respectively. The marker attached at the hand is expected to move faster and to cover longer distance than the one attached at the elbow. Consequently, the resulting hand tracking might be less accurate than the elbow one. Therefore, segment lengths were used for the normalization of the fitting: $\alpha=1/l_1$ and $\beta=1/(l_1+l_2)$. A set of biomechanical constraints must be respected to produce feasible motions; first the estimated JA must be within the joint range of motion.

$$cts_1: \theta_{imin} < \theta_i < \theta_{imax} \quad \text{(2a)}$$

Secondly, estimated joint torques and joint velocities must be below representative physiological thresholds.

$$cts_2: |\Gamma| \leq \Gamma_{max};\ cts_3: |\dot{\theta}| \leq \dot{\theta}_{max} \quad \text{(2b;c)}$$

where Γ_{max} and $\dot{\theta}_{max}$ are the maximal values of the joint torques and velocity, set to Γ_{max}= 100 N.m and $\dot{\theta}_{max}$=90 deg.s^{-1}. These constraints were never active in the optimization process, but they had the effect of avoiding discontinuity and thus high instantaneous joint torques and velocities. Finally, as proposed by Tao et al. [7], the estimated total length of the arm should be equal to the initially measured one.

$$cts_4: ({}^0P_4(\theta(t)) + {}^0P_5(\theta(t)))^2 = l_1 + l_2 \quad \text{(2d)}$$

Occlusions were automatically detected and a specific cost function was used. As an example, the case of a missing elbow marker is taken. In this case, the following hybrid cost function, J_{oc}, that minimises the least square difference between the measured and estimated hand marker and a JA regularity term are used to estimate the JA.

$$J_{oc} = \frac{1}{N}\sum_{t=1}^{N_{oc}} \beta \frac{({}^0Mk_2(t) - {}^0P_5(\theta(t)))^2}{\Delta M_2} + \left(\frac{\boldsymbol{\theta}(t)-\boldsymbol{\theta_{oc}}}{\boldsymbol{\theta_{oc}}}\right) \quad (3)$$

where $\boldsymbol{\theta}_{oc}$ is the vector containing the last estimated JA before the occlusion started.
Finally, constraints forcing the elbow marker positions to be equal to the ones calculated just before, ${}^0P_4(oc)$, and just after, ${}^0P_4(oc_{end})$, the occlusion case was added to the optimisation process.

$$cts6:\ {}^0P_4(t) = {}^0P_4(oc);\ cts7:\ {}^0P_4(t) = {}^0P_4(oc_{end}) \quad \text{(4a, b)}$$

These optimisation problems were solved numerically using a gradient-based non-linear constrained sequential quadratic programming method [15].

C. Experimental validation

Four healthy right-handed volunteers (2 males and 2 females, age 31±5 years, stature 1.8±0.1 m, and mass 89±31 kg) participated to the study after giving informed consent. Volunteers were asked to perform two trials of five grasping-like motions with their right arm. As illustrated in Fig. 3, the hand was moved in 3D from a rest position, to five different positions on the medio-lateral axis localized at the upper chest height. An industrial under-car screwing task was simulated by volunteers by reaching with their hand a point located 0.3 m above their head. The three customized markers were located on the top of the upper arm segment, on the elbow center of rotation, and on the external side of the hand. Eight retro-reflective markers were located on the anatomical landmarks defined in the Plug-in-Gait template (VICON©) (see Fig 3). Their trajectories were recorded using a stereophotogrammetric system (8 Bonita cameras, VICON©, 100 Hz). The inverse geometric model was calculated using the retro-reflective marker positions to estimate relevant JA that will be considered references in the rest of the paper.

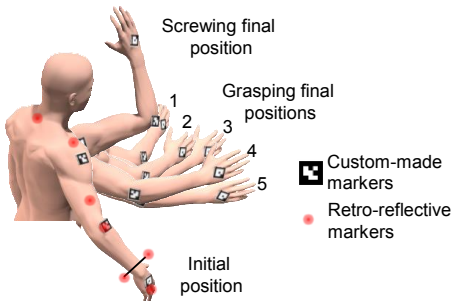

Fig. 3 Experimental paradigm used to validate the proposed method.

III. Results

TABLE I. SUMMARY RESULTS OF THE ACCURACY ASSESSMENT

		Tr 1; 2	θ_1	θ_2	θ_3	θ_4
Sb 1	RMS [°]	5.4; 6.0	2.0; 3.0	2.8; 4.1	5.2; 5.1
	NRMS [%]	5.5; 5.7	4.0; 6.0	3.9; 5.0	3.8; 4.8
	r	0.96; 0.95	0.95; 0.92	0.96; 0.85	0.96; 0.95
Sb 2	*RMS* [°]	3.5; 4.0	1.4; 1.5	3.2; 3.6	2.7; 3.6
	NRMS [%]	4.2; 5.1	4.6; 3.7	9.9; 9.0	3.1; 4.0
	r	0.97; 0.96	0.96; 0.96	0.87; 0.84	0.98; 0.98
Sb 3	*RMS* [°]	5.8; 6.2	4.9; 3.9	4.3; 3.8	5.9; 5.2
	NRMS [%]	6.4; 6.9	10.0; 8.9	6.9; 6.3	5.1; 5.0
	r	0.92; 0.92	0.83; 0.89	0.70; 0.91	0.92; 0.94
Sb 4	*RMS* [°]	3.6; 2.5	3.5; 3.9	3.9; 2.9	3.4; 5.1
	NRMS [%]	4.2; 2.6	10.2; 8.4	9.1; 8.5	4.0; 6.1
	r	0.96; 0.97	0.96; 0.92	0.94; 0.98	0.98; 0.99
Mean	*RMS* [°]	4.6±1.3	3.0±1.2	3.5±0.5	4.2±1.1
	NRMS [%]	5.0±1.3	6.9±2.7	7.3±2.1	4.4±0.9
	r	0.95±0.02	0.92±0.04	0.88±0.08	0.96±0.02

Initial detection of the three markers prior to starting the motion was successful in all investigated cases. Subsequently to this detection, the four subjects performed a total of (5+1)x2x4=48 grasping motions accounting for 5327 samples. From these collected samples the detection-tracking system asked to the human supervisor to select manually the markers twelve times, leading to a very small rate (inferior to 1 %) of missing data. Fig. 4 shows two representative grasping movements and one screwing task and their corresponding tracking, and JA estimates obtained with the proposed low-cost method and with the stereophotogrammetric system. The corresponding RMS difference and correlation coefficients, *r*, are reported in the Table 1 under the label "Sb 1 and Tr 2". Table 1 presents the results obtained for all the subjects and all the trials. The average RMS and correlation

coefficients show a good reproduction of JA with the highest error observed at the shoulder extension/flexion joint (θ_1). This can be explained by the large range of motion of this joint in the considered task. The RMS difference has been normalized by the range of motion of each joint and exhibits a relatively small difference at the shoulder level. Fig. 5 represents the prediction of the trajectories of a simulated missing marker of the elbow for an extended period of 150 samples (4.5 s). Despite the non-linear behavior of the elbow motion, the algorithm is able to reconstruct quite accurately the missing trajectory.

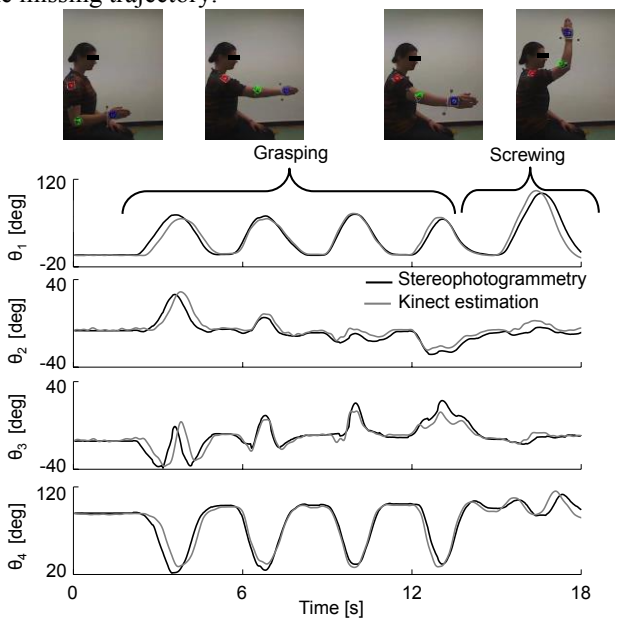

Fig. 4 Representative motions, marker tracking and joint angle estimates.

IV. Discussion

The proposed system based on a KS, a detection-tracking method and an off-line inverse kinematic process is able to estimate the three JA of the shoulder and of the elbow with an average RMS difference inferior to 5 deg and a normalized RMS difference inferior to 8 % for all considered JA. As shown in Fig. 5, a consequent part of the RMS difference is due to the low and variable sampling frequency of the KS and of its resampling to a constant frequency. Nevertheless, the temporal features within the JA estimated with the proposed approach are well respected, as are the amplitudes of the JA. One can see on Fig. 4 that the shoulder position changes during the motion. Shoulder displacement may be exaggerated by patients during reaching task and thus reduce the efficiency of the rehabilitation protocol [9]. With the proposed system, the shoulder position can be easily estimated, in contrast with system using, for example, IMU. Occlusions are handled by the presented approach, although experimentations in real clinical environments must be carried out to investigate the type and occurrence of the occlusions. In case of grasping, motion human motor control field provides numerous cost functions [16] that could be used to estimate missing markers. Future works will also focus on developing a six degree-of-freedom arm model in order to estimate wrist motions that were not of interest in the present study.

Acknowledgment

The authors would like to thanks Dr. Ilaria Pasciuto from the University of Rome Foro Italico (Italy) for her help during the data collection.

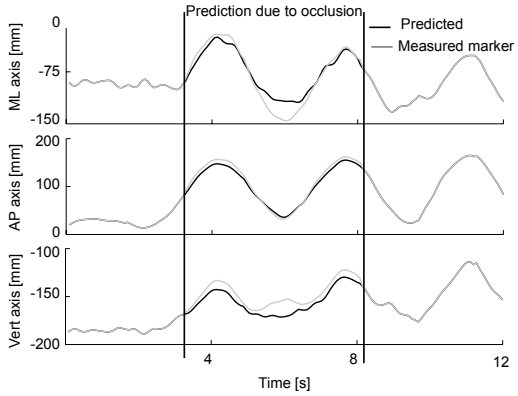

Fig. 5 representative prediction of the missing marker trajectory.

References

[1] D.T.P. Fong, and Y.Y.Chan, "The Use of Wearable Inertial Motion Sensors in Human Lower Limb Biomechanics Studies: A Systematic Review", Sensors, 2010, vol. 10, pp. 1556-1565.

[2] V. Bonnet, C. Mazzà, P. Fraisse, A. Cappozzo, "A least-squares identification algorithm for estimating squat exercise mechanics using a single inertial measurement unit", 2012, vol. 45, pp. 1472-1477

[3] V. Ren, D. Konolige, "RGB-D cameras for 3D modeling and recognition", Robotics & Automation Magazine, vol. 20, 2013.

[4] A. Fernandez-Baena, A. Susin, X. Ligadas, "Biomechanical Validation of Upper-body and Lower-body Joint Movements of Kinect Motion Capture Data for Rehabilitation Treatments", In Proc. of Int. Conf. on Intelligent Networking and Collaborative Systems, 2012, pp. 656-661.

[5] S. Obdrzalek, et al., "Accuracy and Robustness of Kinect Pose Estimation in the Context of Coaching of Elderly Population", In proc. of the Int. Conf. of the IEEE EMBS, 2012, pp. 1188-1193.

[6] K. Buys, et al. "An adaptable system for RGB-D based human body detection and pose estimation", Journal of Visual Communication and Image representation, 2014, vol. 25, pp. 39-52.

[7] Y. Tao and H. Hu, "A Novel Sensing and Data Fusion System for 3-D Arm Motion Tracking in Telerehabilitation", IEEE trans. on instrumentation and measurement, 2008, vol. 57, pp. 1029-1040.

[8] G. Chelius, et al. "A Wearable Sensor Network for Gait Analysis: A Six-Day Experiment of Running Through the Desert", IEEE/ASME Transactions on Mechatronics, 2011, vol. 16, pp. 878 – 883.

[9] L. Oujamaa, et al. "Rehabilitation of arm function after stroke. Literature review", Ann. Phys. Rehabil. Med., 2009, vol. 52, pp. 269-293.

[10] N. Sylla, et al. "Ergonomic contribution of able exoskeleton in automotive industry," Int. J. of Industrial Ergonomics, in press.

[11] S. Garrido-Jurado et al., "Automatic generation and detection of highly reliable fiducial markers under occlusion ", Pattern Recognition, 2014, vol. 47, pp. 2280-2292.

[12] Viola, Paul and Michael J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, vol.1, pp.511–518.

[13] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, "Speeded Up Robust Features", Computer Vision and Image Understanding, 2008, vol. 110, pp. 346-359.

[14] C. Tomasi, T. Kanade. "Detection and Tracking of Point Features", Computer Science Department, Carnegie Mellon University, 1991.

[15] Powell, M.J.D., "A Fast Algorithm for Nonlinearly Constrained Optimization Calculations," Numerical Analysis, ed. G.A. Watson, Springer Verlag, vol. 630, 1978.

[16] B. Berret, E. Chiovetto, F. Nori, T. Pozzo, "Evidence for Composite Cost Functions in Arm Movement Planning: An Inverse Optimal Control Approach", Plos Computational biology, 2011, vol.7.