Synergy-Driven Performance Enhancement of Vision-Based 3D Hand Pose Reconstruction

Simone Ciotti[1,2(✉)], Edoardo Battaglia[1], Iason Oikonomidis[3], Alexandros Makris[3], Aggeliki Tsoli[3], Antonio Bicchi[1,2], Antonis A. Argyros[3], and Matteo Bianchi[1]

[1] Research Center E. Piaggio, University of Pisa, Largo L. Lazzarino 1, 56126 Pisa, Italy
{simone.ciotti,e.battaglia,bicchi,matteo.bianchi}@centropiaggio.unipi.it
[2] Department of Advanced Robotics (ADVR), Istituto Italiano di Tecnologia (IIT), via Morego 30, 16163 Genova, Italy
[3] Institute of Computer Science, Foundation for Research and Technology, Hellas, Heraklion, Greece
{oikonom,amakris,aggeliki,argyros}@ics.forth.gr

Abstract. In this work we propose, for the first time, to improve the performance of a Hand Pose Reconstruction (HPR) technique from RGBD camera data, which is affected by self-occlusions, leveraging upon *postural synergy information*, i.e., *a priori* information on how human most commonly use and shape their hands in everyday life tasks. More specifically, in our approach, we ignore joint angle values estimated with low confidence through a vision-based HPR technique and fuse synergistic information with such incomplete measures. Preliminary experiments are reported showing the effectiveness of the proposed integration.

1 Introduction

In recent years, the need for accurate 3D Hand Pose Reconstruction (HPR) has gained an increasing attention in many application fields such as virtual reality, ambulatory human motion/activity monitoring, biomechanics, rehabilitation and human robot interaction [1]. Different methods proposed for HPR can be classified as glove-based HPR (e.g., [2–4]) and vision-based HPR (e.g., [5]). The first type of approaches relies on the usage of wearable resistive, inertial or piezoelectric sensors [1,6,7] to measure quantities related to joint angles. Vision-based methods employ data acquired from cameras (typically RGBD) to reconstruct hand kinematics information. However, both these approaches can be affected by constraints that arise from the complexity in modeling the biomechanics of the human hand, measurement noise, sensor resolution and visual occlusion, among others [4]. To improve HPR performance, an important asset is the work laid out in [1,4,8–12], where the existence of postural synergies was exploited to enhance kinematic and joint angle reconstruction performance and to design optimized gloves with a limited number of sensing elements. The basic idea was to interpret *postural synergies*, that is, goal-directed kinematic

P. Perego et al. (Eds.): MobiHealth 2016, LNICST 192, pp. 328–336, 2017.
DOI: 10.1007/978-3-319-58877-3_42

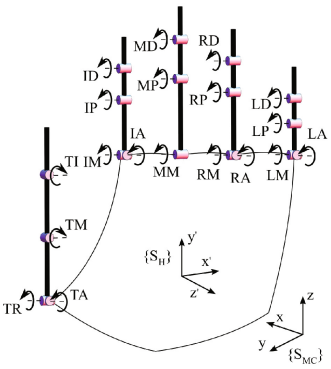
activation or inter-joint covariation patterns [13,14], in terms of statistical *a priori* information on the probabilistic distribution of human poses in common tasks like grasping. This information can be fused with incomplete and possibly inaccurate measurements provided by an HPR to increase its performance [4] and can be used for optimal placement of sensors on a glove for HPR in order to reconstruct hand posture, especially with a limited number of sensors [10].[1] In this work, we push forward our investigation and apply, for the first time, synergy-inspired performance enhancement to a vision-based HPR method [5]. This vision-based technique proposes to recover and track in real-time 3D position, orientation and full articulation of a human hand from marker-less visual observations obtained by the commercial RGBD camera Xtion PRO[2] following the optimization approach described in [5,17]. Despite its simplicity and effectiveness, such a reconstruction procedure is affected by some intrinsic limitations. For example, no-matter how the camera is placed w.r.t the hand, there are always self-occlusions that limit the reconstruction accuracy. In this paper we propose to discard joints estimated with low confidence and to complete HPR through synergistic information, integrating techniques reported in [4,5].

2 Synergy-Based Hand Pose Reconstruction

For the sake of clarity, let us summarize the definitions and results from [4,8]. Let us consider an n degrees of freedom kinematic hand model, with $y \in \mathbb{R}^m$ measures provided by an HPR system. In this case, the joint variables $x \in \mathbb{R}^n$ and measurements y are related by the equation $y = Hx + \nu$, with $H \in \mathbb{R}^{m \times n}$ $(m < n)$ the full row rank matrix and $\nu \in \mathbb{R}^m$ the vector of measurement noise, with a zero mean and Gaussian distribution with covariance matrix R. Our objective is to determine hand posture, which can be represented by joint angles x in a hand model (Fig. 1), from a reduced set of measures y. This objective can be achieved by using postural synergy information. Hand synergies are goal-directed, combining muscle and kinematic activation, leading to a reduction of the dimensionality of the motor and sensory space. Furthermore, in robotics, hand synergies have represented a highly effective solution for the fast and simplified design and control of artificial systems (see [14] for a comprehensive review on hand synergies and their applications). From a kinematic point of view, hand synergies can be defined in terms of inter-joint covariation patterns, which were observed both in free hand motion and object manipulation [14]. In [4], following the approach introduced in [13], we embedded synergy information in an *a priori* set of imagined grasped object poses N, defining a $X \in \mathbb{R}^{n \times N}$ matrix. This information can be summarized in a covariance matrix $P_o \in \mathbb{R}^{n \times n}$, i.e., $P_o = \frac{(X - \bar{x})(X - \bar{x})^T}{N-1}$, where $\bar{x}$ is a matrix $n \times N$ whose columns contain the mean

[1] It is worth to mention that robotics research has leveraged upon neuroscientific insights on synergies to inform the design and control of artificial hands, see e.g. [14–16].

[2] Images and depth maps are captured at 640 × 480@24 bit and 640 × 480@16 bit, respectively.



DoFs	Description
TA	Thumb Abduction
TR	Thumb Rotation
TM	Thumb Metacarpal
TI	Thumb Interphalangeal
IA	Index Abduction
IM	Index Metacarpal
IP	Index Proximal
ID	Index Distal
MM	Middle Metacarpal
MP	Middle Proximal
MD	Middle Distal
RA	Ring Abduction
RM	Ring Metacarpal
RP	Ring Proximal
RD	Ring Distal
LA	Little Abduction
LM	Little Metacarpal
LP	Little Proximal
LD	Little Distal

Fig. 1. Hand Model used for the synergy-based Hand Pose Reconstruction.

values for each joint angle arranged in vector $\mu_o \in \mathbb{R}^n$. According to [4], the hand pose reconstruction can be obtained through the minimum variance estimation (MVE) technique as:

$$\hat{x} = \mu_o - P_o H^T (H P_o H^T + R)^{-1} (H\mu_o - y)\,. \tag{1}$$

3 Visual Tracking of Hand Posture

The visual tracking of hand posture described in this paper relies on the technique described in [5][3]. Within the vision-based HPR methods, a distinction can be made between discriminative and generative: the former rely on large datasets of hand poses to learn a mapping from the visual input to the hand pose space [18,19], while the latter rely on 3D models of the kinematics and appearance of a human hand, and try to match these models to the visual input by rendering the 3D hand model and comparing it to the visual observations. This explicit handling of a 3D hand model allows the calculation of the occluded parts of the hand and, indirectly, the level of estimation confidence. The approach presented in [5] falls in the second category. We use as input to our method the Xtion PRO, a camera that captures RGBD data, and the appearance of the hand model is approximated by appropriately transformed and positioned cylinders and spheres as shown in the left panel of Fig. 2, while the kinematics is illustrated in the right panel of the same figure. To perform tracking-based HPR for each input frame, we begin from an initial hand pose, which is used to start a search using our own variant of Particle Swarm Optimization (PSO) [17], as described in [5]. During tracking, a hand pose x is sought that matches the observations $O = (o_s, o_d)$, respectively the silhouette and the depth map of the observed hand. The core of optimization involves

[3] Implementation available online at: http://cvrlcode.ics.forth.gr/handtracking/.

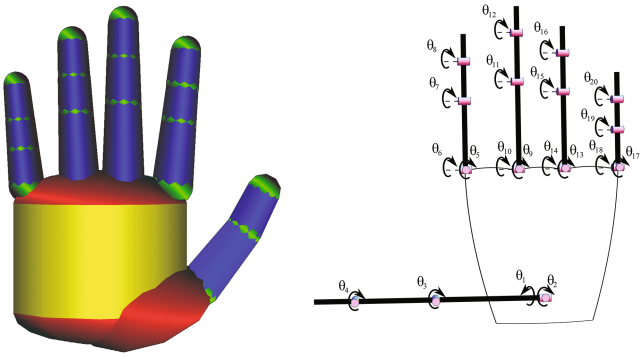


Fig. 2. Appearance (left) and kinematics (right) model of the hand used for tracking-based HPR. The hand appearance is built out of appropriately positioned cylinders and spheres. Colors encode different types of geometric primitives.

the comparison of a candidate hand pose x against O, with a matching error computed from per-pixel depth differences. Specifically, a hand pose hypothesis x, given the camera calibration information C yields a rendered depth map $r_d(x, C)$. We compare this map with the respective observation o_d, computing a "matched depths" binary map $r_m(x, C)$. This map in turn is compared to the observed silhouette, so that the objective function exhibits an optimum when the hypothesized and observed silhouettes match. Overall, the D is computed as: $D(O, x, C) = \frac{\sum min(|o_d - r_d|, d_M)}{\sum(o_s \vee r_m) + \epsilon} + \lambda \left(1 - \frac{2\sum(o_s \wedge r_m)}{\sum(o_s \wedge r_m) + \sum(o_s \vee r_m)}\right)$, where D_M serves as a maximum penalty in depth difference, used to smooth out the behavior of the objective function around the optimum and λ is an experimentally determined weight factor. We formulate the final objective function by adding to the quantity D an appropriately weighted penalty term to prevent configurations in which hand parts (e.g., palm, fingers) occupy the same physical space. Based on the values of this objective function, PSO improves the candidate hand poses, eventually coming up with a hand pose that matches the input data. For more details on the employed tracking-based HPR method, the reader is referred to [5].

4 Assessing the Confidence of Vision-Based Joint Angles Estimation

Obtained a hand pose, in order to select the most trusted joints to be used in the synergy-based HPR stage (Sect. 2), an estimation of the confidence for each of them is required. We determine this confidence by capitalizing on occlusion information. Intuitively, the confidence of each estimated joint angle is at least partly determined by the level of occlusion of the two parts on either side of the joint. Therefore, we first compute the occlusion for each rigid part of the hand. Then, the confidence for each joint is computed as the product of the occlusion levels of the two rigid hand parts linked by that joint. More specifically, given

a hand pose x we can count the number of pixels that are drawn using each geometric primitive of the hand model shown in Fig. 2, left. In order to provide a normalized estimate of the visibility for each hand part, a reference pose x_r is used, in which all the hand parts are visible. In our implementation this pose is an open hand with the palm facing the camera (see Fig. 3, left). The reference pose is rendered offline and the area of each hand part is calculated (in pixels). Assuming this information, we then compute the percentage of each part within the reference pose as the ratio of its rendered pixels over the total number of rendered pixels. Specifically, let a be the area in pixels of a hand part. Then the i-th hand part in the reference pose x_r has an area $a_i(x_r)$ and respective ratio $r_i(x_r) = \frac{a_i(x_r)}{\sum_{k \in P} a_k(x_r)}$, where P denotes the set of all part indices. The vector of precomputed ratios for all the hand parts $r_i(x_r), i \in P$ is stored for use during the visibility check of arbitrary hand poses. Provided an arbitrary hand pose x, a similar computation is carried out. We first count the number of pixels $a_i(x)$ per hand part in that pose. An area percentage $r_i(x)$ is again computed as the ratio of hand part pixels over total occupied. The final visibility score per part $v_i(x)$ is obtained as the ratio of its two percentages, the one computed using the target pose over the precomputed one on the reference pose $v_i(x) = \frac{r_i(x)}{r_i(x_r)}$. Computed the visibility score for each hand part, the final decision per DoF is taken by checking the two hand parts corresponding to it. If the product of its visibility ratios is over a given threshold then we retain the estimation for this DoF, otherwise we discard it. Figure 3 illustrates this process. The posture is then completed with the synergy-based techniques described in Sect. 2. Note that the visual tracking HPR method has negligible measurement noise ($< 0.2°$) and hence we have not considered it within the synergy-based reconstruction, which was proven to be robust to noise [8].

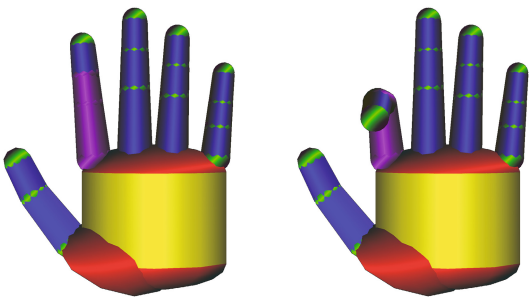


Fig. 3. Illustration of the main idea for assessing the confidence of joint angles estimation. The proximal interphalangeal joint of the index finger connects the proximal phalanx to the intermediate one (highlighted in purple). The areas of the highlighted phalanges in the reference pose (left) are computed offline and stored. Assuming an arbitrary hand pose (right), we perform the same computation and compare the visibility ratios v_i for each of the two phalanges. In this example, the intermediate phalanx is almost completely occluded, lowering the confidence in the estimation for the examined joint. (Color figure online)

5 Integration

As detailed in the previous sections, the tracking-based HPR and the synergy-based HPR rely on two different hand models. The model used in tracking-based HPR is naturally induced by the parametrization of the fingers as a succession of 3D points clusters, while the model used in synergy-based HPR was derived from a biomechanical model of the hand [13]. For this reason, the two models, while being fairly similar overall, differ substantially in the description of the thumb. In order to provide a rough mapping to feed thumb angle joints from the tracking-based HPR model to the synergy-based HPR model, a sparse sampling of direct kinematics for the tracking-based HPR model was performed, for a large number K of values of joint angles. For each of these instances, inverse kinematics was performed in the model used for synergy-based HPR, imposing a minimal distance between the position of joint centers with a least-squares approach, assuming the same length for each phalanx. At the end of this process we obtained a matrix of joint angles $\Theta_S \in \mathbb{R}^{4\times K}$ in the synergy-based HPR, for which each column corresponds to joint angles in the tracking-based HPR hand model $\Theta_V \in \mathbb{R}^{4\times K}$. From these values a matrix $Q = \Theta_S((\Theta_V^T)^{-1})^T$ can be obtained which gives an approximate mapping of joint angles for the thumb from one model to the other, and an estimate of joint values in the synergy-based HPR model can be obtained from values in the tracking-based HPR model as $\theta_S = Q\theta_V$. For the other fingers this conversion is straightforward, as it is simply a change in sign. Figure 4 shows a block-diagram with the different phases of integration of the tracking-based HPR system with the synergy-based HPR algorithm. The tracking-based HPR is implemented with a python script, while the synergy-based HPR is implemented in C++. This, together with the fact that it is a closed form formula, ensures that the additional computational cost introduced by the synergy-based HPR is negligible. Python script, C++ code, and hand pose visualization all run from the same PC through UDP. The whole loop is executed with a frequency of 2 Hz (±0.1 Hz). Both input (from the tracking-based HPR) and output (from synergy-based HPR) joint angles are filtered with a smoothing filter.

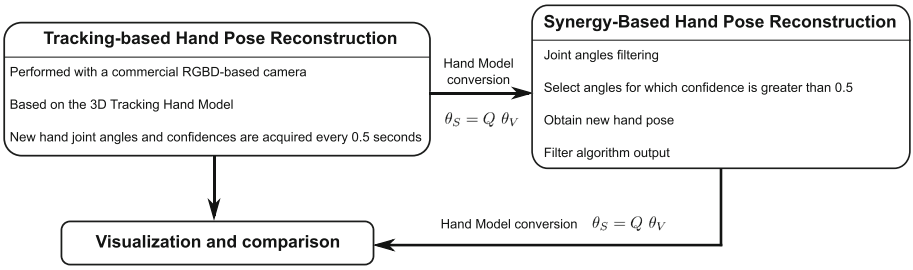


Fig. 4. Integration of visual-based HPR system and synergy-based HPR algorithm.

6 Preliminary Results

Preliminary experiments were performed with one 28 years old male subject and 3 objects (bag, book, and key). In the experiments the subject was provided a picture of the object to grasp, and asked to perform a grasp action for the target object (similarly to what was done in [13]). The subject was asked to maintain the hand posture for 5 s. In Fig. 5 we show results of the hand pose reconstruction from different points of view.

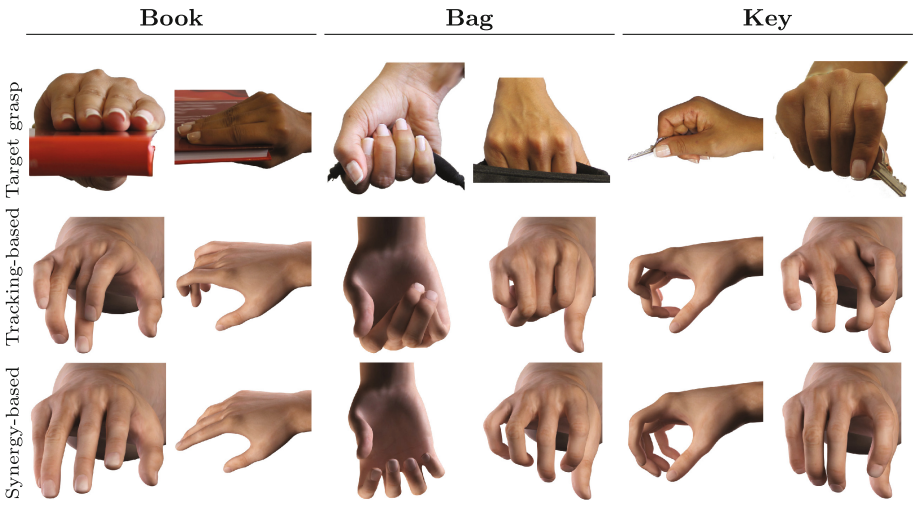


Fig. 5. Hand Pose Reconstruction results.

Referring to Fig. 1, we chose to use as input to the synergy-based HPR algorithm the following angles from the visual tracking outcomes: TA and MP for the bag; TA, IA, IM, IP, RA, RM, LA, and LM for the book; TA, MM, MP, RA, RM, RP, LA and LM for the key. We selected these angles since their confidence (whose assessment is detailed in Sect. 4) is greater than 0.5. The rest of kinematics is obtained by fusing these data with synergy-based a priori information (as described in Sect. 2). What is noticeable is that the integration of measurements and postural synergy information on the most probable human poses enables a more human-like and reliable posture reconstruction, in cases where the visual tracking of hand posture needs to deal with low-confidence estimated angles.

7 Conclusions and Future Work

In this work we have presented an integrated approach that combines optimal HPR based on *a priori* synergistic information on probabilistic distribution of human hands, and an optimization procedure to accurately track and reconstruct hand pose from visual data provided by a commercial RGBD camera.

More specifically, hand joint values estimated through such an optimization procedure are discarded if the confidence in their estimation is low. Hand posture is then completed by fusing synergy information with the remaining estimates in a Bayesian optimal fashion. Preliminary qualitative results show that the integrated approach provides more realistic and accurate 3D hand tracking than the original optimization techniques. This is particularly true in those conditions where occlusions of parts of the tracked hands can be observed. While the performed experiments considered human grasping, there is no inherent limitation that prevents the applicability of our approach to other types of hand activities. Future works will further develop the integration of different sensing modalities as done e.g. in [12]. The idea is to push forward under-sensing approach for wearable sensors [1] and synergy-based performance enhancement, taking advantage from both visual (unobtrusive, usable) and non-visual (wearable) HPR to increase performance in ambulatory monitoring, virtual reality and human-robot interaction. Finally, we will perform a more quantitative evaluation of the results in real-time hand tracking tasks, investigation how synergy information can be used to reduce the search space for the methods described in [5].

Acknowledgment. This work is supported in part by the European Research Council under the Advanced Grant SoftHands (No. ERC-291166), by the EU H2020 projects SoftPro (No. 688857) and SOMA (No. 645599), and by the EU FP7 project WEARHAP (No. 601165).

References

1. Ciotti, S., et al.: A synergy-based optimally designed sensing glove for functional grasp recognition. Sensors **16**(6), 811 (2016)
2. Sturman, D.J., et al.: A survey of glove-based input. IEEE Comput. Graphics Appl. **14**(1), 30–39 (1994)
3. Dipietro, L., et al.: A survey of glove-based systems and their applications. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. **38**(4), 461–482 (2008)
4. Bianchi, M., et al.: Synergy-based hand pose sensing: Reconstruction enhancement. Int. J. Robot. Res. **32**(4), 396–406 (2013)
5. Oikonomidis, I., et al.: Efficient model-based 3D tracking of hand articulations using kinect. In: British Machine Vision Conference (BMVC 2011), vol. 1, no. 2, pp. 1–11. BMVA, Dundee (2011)
6. Muth, J.T., et al.: Embedded 3D printing of strain sensors within highly stretchable elastomers. Adv. Mater. **26**(36), 6307–6312 (2014)
7. Hsiao, P.-C., et al.: Data glove embedded with 9-axis imu and force sensing sensors for evaluation of hand function. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 4631–4634. IEEE (2015)
8. Bianchi, M., et al.: On the use of postural synergies to improve human hand pose reconstruction. In: 2012 IEEE Haptics Symposium (HAPTICS), pp. 91–98. IEEE (2012)
9. Bianchi, M., et al.: Synergy-based optimal design of hand pose sensing. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 3929–3935, October 2012

10. Bianchi, M., et al.: Synergy-based hand pose sensing: optimal glove design. Int. J. Robot. Res. **32**(4), 407–424 (2013)
11. Bianchi, M., et al.: Exploiting hand kinematic synergies and wearable under-sensing for hand functional grasp recognition. In: 2014 EAI 4th International Conference on Wireless Mobile Communication and Healthcare (Mobihealth), pp. 168–171, November 2014
12. Bianchi, M., et al.: A multi-modal sensing glove for human manual-interaction studies. Electronics **5**(3), 42 (2016)
13. Santello, M., et al.: Postural hand synergies for tool use. J. Neurosci. **18**(23), 10 105–10 115 (1998)
14. Santello, M., et al.: Hand synergies: integration of robotics and neuroscience for understanding the control of biological and artificial hands. Phys. Life Rev. **17**, 1–23 (2016)
15. Catalano, M.G., et al.: Adaptive synergies for the design and control of the Pisa/IIT softhand. Int. J. Robot. Res. **33**(5), 768–782 (2014)
16. Matrone, G.C., et al.: Principal components analysis based control of a multi-dof underactuated prosthetic hand. J. Neuroeng. Rehabil. **7**(1), 1 (2010)
17. Kennedy, J., et al.: Particle swarm optimization. In: International Conference on Neural Networks, vol. 4, no. 3, pp. 1942–1948. IEEE, January 1995
18. Sun, X., et al.: Cascaded hand pose regression. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 824–832 (2015)
19. Keskin, C., Kıraç, F., Kara, Y.E., Akarun, L.: Hand pose estimation and hand shape classification using multi-layered randomized decision forests. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7577, pp. 852–863. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33783-3_61