Computers and Electrical Engineering 000 (2017) 1–20

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Vision-based human grasp reconstruction inspired by hand postural synergies☆

Ritwik Chattaraj[a,d,*], Siladitya Khan[b], Deepon Ghose Roy[a], Bikash Bepari[c], Subhasis Bhaumik[a]

[a] School of Mechatronics and Robotics, Indian Institute of Engineering Science and Technology, Shibpur, 711103 Howrah, India
[b] Department of Applied Electronics and Instrumentation Engineering, RCC Institute of Information Technology, 700015 Kolkata, India
[c] Department of Production Engineering, Haldia Institute of Technology, 721657 Haldia, India
[d] Department of Mechatronics Engineering, Manipal Institute of Technology, Karnataka

A R T I C L E I N F O

Article history:
Received 29 March 2017
Revised 18 October 2017
Accepted 20 October 2017
Available online xxx

Keywords:
Human hand
Dimensionality reduction
Principal component analysis
Isomap
Postural synergies
Grasp reconstruction

A B S T R A C T

The human hand exhibits enormous versatility and dexterity, to a degree paralleled by few gripping assemblies. Although several hand posture animators have been discovered, vision-based trackers have retained the research focus owing to their compactness, cost-effectiveness and ease-of-installation. The present investigation explores a marker-based hand pose-tracking solution, using a Kinect depth-capture device. It exploits the inherent synergism within the finger linkages through a novel motion capture algorithm for grasp reclamation. The tracked data-set is analysed for an optimal number of condensed primitives which yielded an effectively reclaimed grasp-pose. Isomap based dimensional reduction followed by Principal Component Analysis (PCA) back-projection, drives the reconstruction of thirty-three Feix-grasps solely from Index, Thumb and three edges of a Palm marker. The derivatives were observed to contribute across grasp initiation to final posture assumption, with a scope for future investigations on their direct correlations to cortical motor impulses.

1. Introduction

Human hand articulations have always remained a prominent fore-runner in gesture recognition and imitation of human behaviour. The hand structure possesses natural agility and manipulability that allows assumption of various postures with little effort. Photogrammetric systems have been preferred over other motion capture methods, owing to economical hardware setups, non-invasive mode of operation as well as absence of labour intensive post-processing. However, recovery of full articulations of human hands from a single camera is cumbersome, due to immense complexities involved in background segmentation and obtrusions caused by individual fingers. The advent of Kinect in the domain of human motion tracking has enabled 3D pose estimation, by taking into account the depth variation along with color based segmentations [1–3]. The problem of deriving correspondences between the actual 3D feature points and their pixel counterparts is the key to this context. Establishing direct associations between the detected feature locations and their corresponding set of 3 dimensional landmarks, (henceforth referred to as control points) have been in crucial focus over decades. The issue asso-

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Area Editor Dr. E. Cabal-Yepez.
* Corresponding author.
E-mail address: ritwik.chattaraj@manipal.edu (R. Chattaraj).

https://doi.org/10.1016/j.compeleceng.2017.10.018
0045-7906/

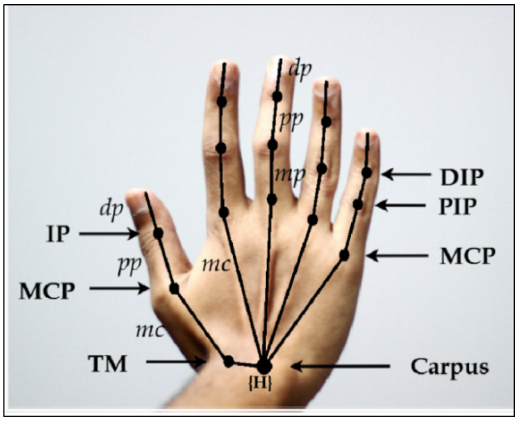

Fig. 1. Human hand kinematic model.

ciated with determining the relative location of control points whose image has been acquired, forms the basic perspective transformation problem in a pin-hole camera model (referred to as the Location Determination Problem (LDP)) [4]. A framework based on the Random Sample Consensus (RANSAC) algorithm, as a method of solving the LDP was presented in [5], for dynamic hand gesture recognition with optical cues from a single camera. Similar attempts were undertaken with analytical tools like Hidden Markov Models (HMM), Unscented Kalman Filters, Dynamic Time Warping and Weak Perspective Transformations [6–12]. Although such methodologies saw a remarkable increase in operational accuracy, yet each of them failed to cope up with high computational burdens.

In the context of grasp reclamation, tracking all the fingertip locations for reconstruction of the 3D hand proves to be a computationally costly affair. Also the fact that grasping occludes a large part of the hand structure, thereby leading to notable tracking errors, which ultimately results in a distorted estimation of the final posture. This has been eliminated by the introduction of standard dimensional reduction techniques on human grasps [13,14], which may be explored to reclaim the complete hand stance by tracing reduced number of marked hand features. The lower dimensional representation of the human hand kinematic chain, presents a unique method of visualizing the same as a coupled system, with the existential relationship accounted by the Synergy Matrix [15–17].

The present investigation is targeted at tracking the hand articulations with the help of coloured markers placed on the finger-tips and the palm. A revisit to the classical Perspective-3-point problem in RANSAC paradigm is presented in the article, and sought to be resolved by modern depth capture devices like Kinect. Both RGB as well as the depth sensor of the module, is used collectively to present a computationally inexpensive solution to recreate a complete hand grasping action in 3D space. However intermittent obtrusions produced while performing a grasping act, calls for the selective tracking of a few feature locations from which the complete posture may be derived. In order to investigate the underlying correlations intrinsic in a hand kinematic architecture, thereby leading to reduction in the complexity of control, dimensional reduction techniques have been used to predict the complete hand motion from the tip locations of Index and the Thumb. However as an improvement over the prevalent techniques targeting hand posture reclamation in terms of reduced synergy derivatives, the proposed method does not impose additional kinematic constraints other than the ones already levied on the model in order to ensure its operational similarity to that of a human hand. In contrast to previous investigations which catered to a generic overview of reclaimed grasp stances, this article seeks to narrate the optimal synergy explanation in accomplishing specific grasping acts. Such an optimum postural Eigen subspace, ensures simultaneous reduction in reclamation complexity for various task specific hand configurations, yet at the same time retains considerable similarities between the assumed and modelled stances.

Section 2 describes the kinematic description of the hand model employed in the investigation followed by a discussion on the motion capture scheme from RGB + D sequence in Section 3. The formulation of grasp synergies are described in Section 4, from which the technique of synergy inspired motion capture is derived in Section 5. Finally the associated results and comparisons are discussed in Section 6, with a brief conclusion and scope for future improvements in Section 7.

2. Kinematic representation of a human hand in space

The present investigation uses the paradigmatic hand model [18] for establishing correlations between the various joints. The model possesses twenty Degrees of Freedom (DoFs), as shown in Fig. 1 with each finger representing a single kinematic linkage-considered to have 4 DoFs individually, signifying the parameterized human hand state. The various ostial links shown in the figure represent the skeletal framework namely (*dp*: Distal Phalanx, *pp*: Proximal Phalanx, *mp*: Middle Phalanx, *mc*: Metacarpal bone) along with the joints viz. (DIP: Distal Interphalangeal, PIP: Proximal Interphalangeal, MCP: Metacarpophalangeal, IP: Interphalangeal, TM: Trapeziometacarpal), which symbolizes a system of connected rigid bodies.

The Denavit-Hartenberg (D-H) parameters for the hand, representing the coordinate transformations between the frames attached to all the links and forming the overall kinematic description in a recursive manner has been tabulated in Table 1.

Please cite this article as: R. Chattaraj et al., Vision-based human grasp reconstruction inspired by hand postural synergies, Computers and Electrical Engineering (2017), https://doi.org/10.1016/j.compeleceng.2017.10.018