

3D Motion Capture for Indian Sign Language Recognition (SLR)

E. Kiran Kumar, P.V.V. Kishore, A.S.C.S. Sastry and D. Anil Kumar

Abstract A 3D motion capture system is being used to develop a complete 3D sign language recognition (SLR) system. This paper introduces motion capture technology and its capacity to capture human hands in 3D space. A hand template is designed with marker positions to capture different characteristics of Indian sign language. The captured 3D models of hands form a dataset for Indian sign language. We show the superiority of 3D hand motion capture over 2D video capture for sign language recognition. 3D model dataset is immune to lighting variations, motion blur, color changes, self-occlusions and external occlusions. We conclude that 3D model based sign language recognizer will provide full recognition and has a potential for development of a complete sign language recognizer.

1 Introduction

Motion analysis is the way toward capturing real life gestures and movements of a subject as arrangements of Cartesian facilitates in 3D space. A motion capture framework has applications in various domains such as surveillance [1–4], assistive interaction of human with computer technologies [5, 6], deaf sign word recognition [7–9], computational behavioral science [10, 11] and consumer behavior analysis [12], the focus is the detection, recognition and analysis of human movements and behavioral actions.

E. Kiran Kumar (✉) · P.V.V. Kishore · A.S.C.S. Sastry · D. Anil Kumar
Department of Electronics and Communication Engineering,
K L University, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh, India
e-mail: kiraneepuri@kluniversity.in

P.V.V. Kishore
e-mail: pvvkishore@kluniversity.in

A.S.C.S. Sastry
e-mail: ascssastry@kluniversity.in

D. Anil Kumar
e-mail: danilmurali@kluniversity.in

Motion capture systems are classified into magnetic, mechanic, and optical. In magnetic system, the electromagnetic sensors connected to a computer used to produce the real-time 3D data at lower processing cost. But some of the movements are restricted due to the cabling. The mechanical motion capture system uses suits with integrated sensors which records the real-time movements as 3D data.

Optical motion capture utilizes cameras to recreate the body stance of the entertainer. One approach utilizes an arrangement of numerous synchronized cameras to catch markers set in vital areas on the body.

As introduced newly in [13], the community of computer vision describes a skeleton as a schematic model of the human body. The skeletal parameters and motion attributes can be used as an illustration of the gestures commonly known as actions and, consequently, the human frame pose is described by means of the relative joint locations in the skeleton. Application domains such as gaming and human-computer interfaces are greatly benefiting of this new innovative technology.

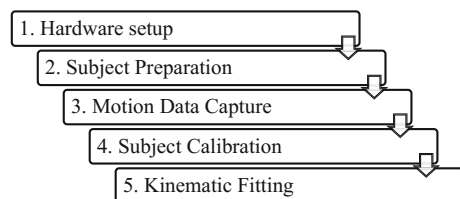
In recent years, the motion capture technology is introduced to capture the hand motions, which in turns greatly boost the field of sign language recognition in 3D environment. This increases the efficiency and accuracy of sign language recognition system. The biomechanical anatomy of human hand incorporates many number of degrees of freedom (DoF) which making the mapping of external measurements with functional variables complex. Hence the finest assessment of human hand kinematics is a complex task [14].

The present trends in sign language recognition technologies are on the primary basis on virtual reality, unable to produce animations on their own in which the signer associates locations in space with entities under discussion. The automatic spatial association of signs locations is closely related to the sign recognition process. Without extracting the signature of the sign word through motion analysis, the spatial location of signs cannot be modified as the change in location of entities under discussion.

Kinematic movements of a human hand are increasingly demanded in sign language recognition. The appropriateness of the finger movements is greatly prompting to identify the accurate sign. From the past few decades, the development is confined to human gait analysis in clinical research field. In clinical gait analysis, the positional changes of the markers attached to the skin greatly affect the analysis. Performing the sign, which provides small displacements in the marker positions due to skin moments will not effects our sign recognition.

The Optical motion capture of a subject flow is shown in Fig. 1.

Fig. 1 3D motion capture model



Hardware Setup: As an initial stage the capture volume must be defined accordingly the camera positions to be fixed such a way that at least any two cameras to see the markers, which is known as the field of view. Using the software interface nexus, the cameras to be calibrated and a global coordinate system is set to produce reliable 3D data. A special wand is used to calibrate the cameras.

Subject Preparation: The signer is a subject attaches a passive retro reflective marker set on to the surface of the hand and a static trial captured. The skeletal structure of the subject and the marker set is described in the nexus and stored as a Vicon skeleton template (.vst) file.

Motion Data Capture: Now in the dynamic trial the cameras capture the radiated light from the markers and produces blobs at exact locations. The cameras will use the calibration information to reconstruct and locate the markers on to the 3D coordinates.

Subject Calibration: The VST file is not necessary to be 100% accurate all the times. The lengths of the fingers vary with subject to subject. In this case the subject calibration fixes the problem.

Kinematic Fitting: All captured trials are labeled and fit the kinematic model to compute the joint angles. These joint angles are treated as outputs of motion capture and necessary to operate the 3D model.

2 Motion Capture Setup

The finger movements of a signer can be recorded using motion capture setup. We conducted motion capture using a 6-camera Vicon model at 100 Hz. Each hand is represented using 22 markers of size 6.4 mm. A 14 mm sized markers were used for head. A total of 54 markers were used to represent a signer. The below Fig. 2 shows the camera arrangement on a 3D Cartesian plane.

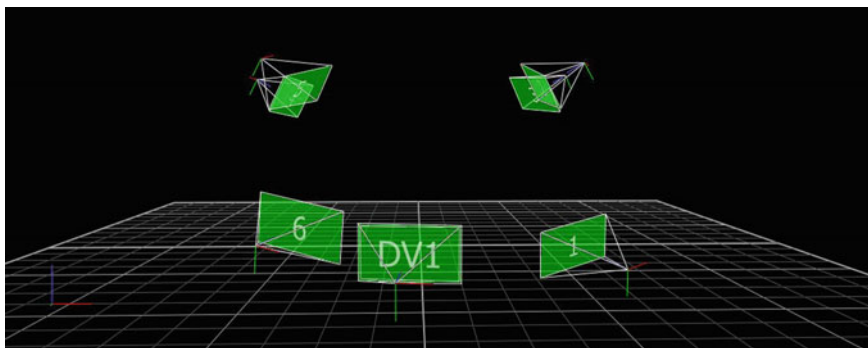


Fig. 2 Camera positions on a 3D cartesian plane

2.1 Marker Placement for Hand Motion Capture

Several methods of marker placements were proposed by many researchers. Miyata [15] have introduced a model with 25 markers for each hand to measure wrist, fingers joint angles. Carpinella [16] developed a hand model with less number of markers but wrist movements are not included. In [17], three linear markers used at metacarpal, proximal and distal interphalangeal joints for each finger.

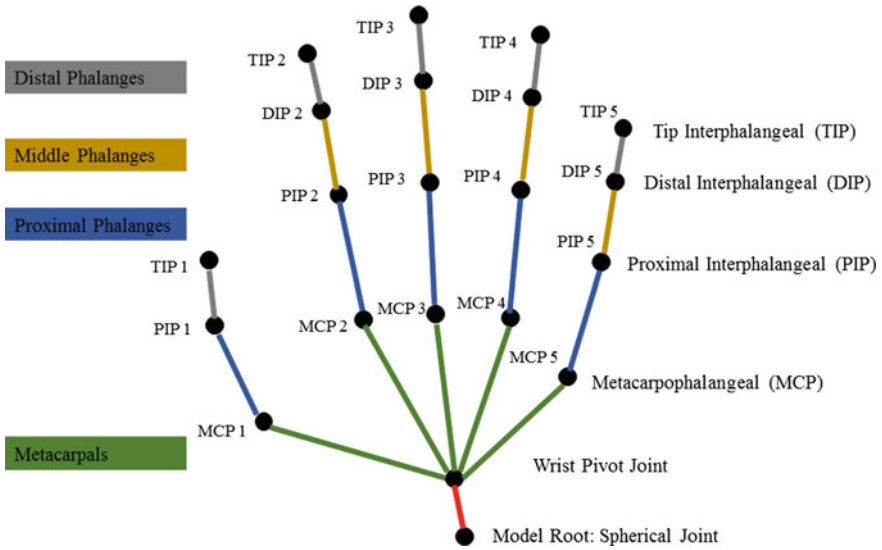


Fig. 3 The hand anatomy



Fig. 4 Marker placement design and actual markers used for capturing the sign language

Even though several models were proposed by many researchers the sign language requires a sophisticated hand model in order to produce accurate data. Figure 4 shows the proposed hand model with marker placements based on the hand anatomy as shown in Fig. 3.

The model in Fig. 4 can capture every sign language character which can be easily recognized accurately. Whereas in 2D the visualization is only in one direction and the information which can only be seen in other direction can be lost leading to inappropriate classification. As shown in Fig. 5 blurring will also affect the recognition in 2D which leads to false classification.

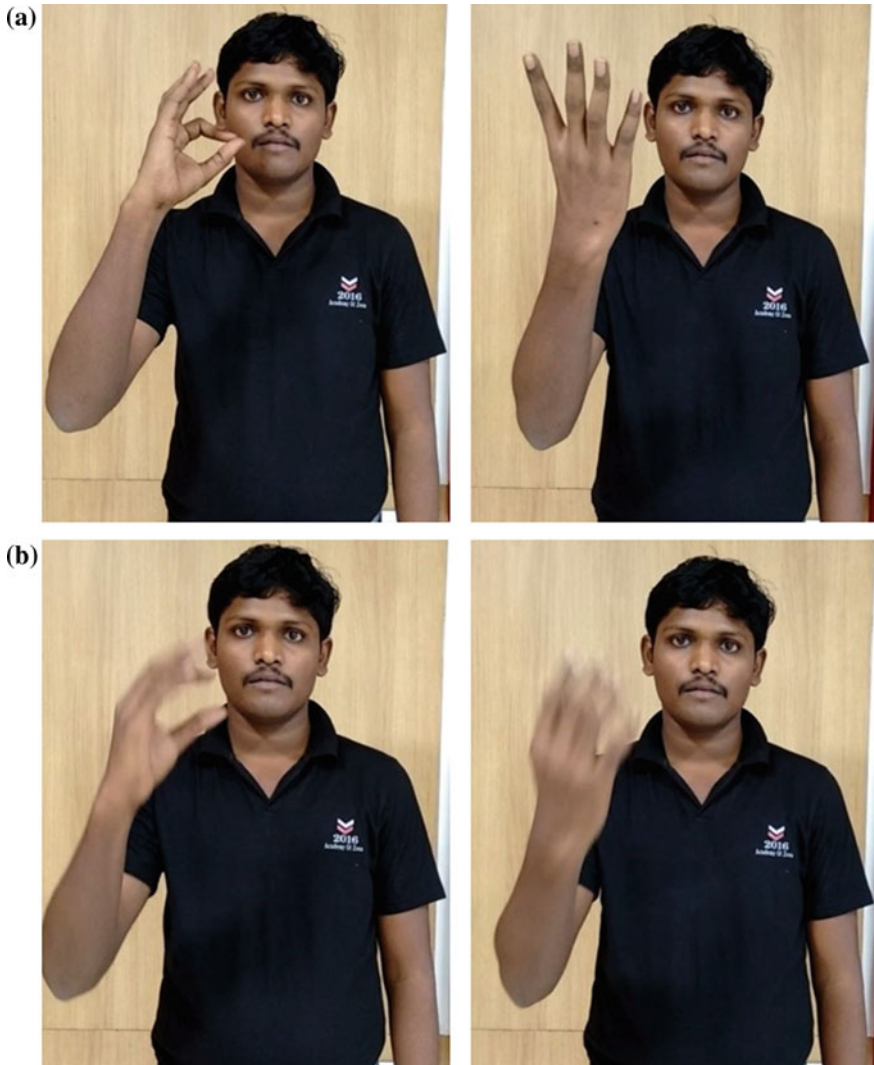


Fig. 5 a 2D signs with missing finger information, b 2D sign with blur

3 Results and Discussions

To validate our approach, we captured the Indian sign ‘Good Morning’ in 3D using motion capture technology and successfully obtained the sign in all the direction which shows the information that is missed in 2D. Figure 6 shows the sign in

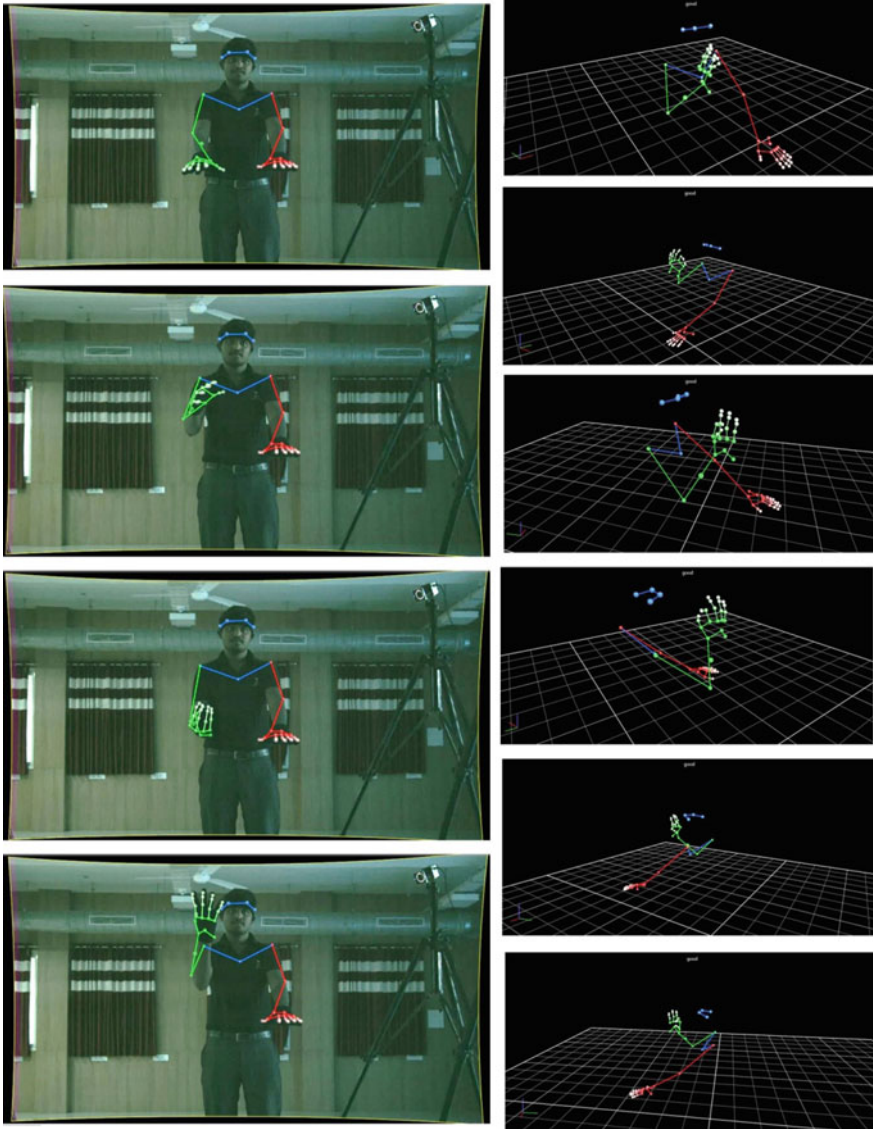


Fig. 6 3D image data captured for some signs from Indian sign language. The meaning of the sign is ‘good morning’

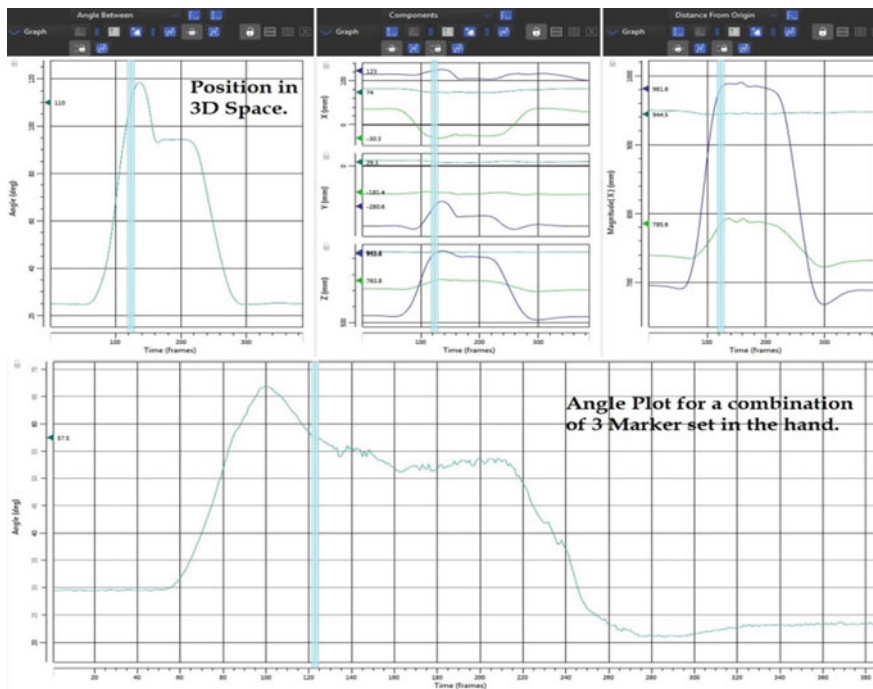


Fig. 7 3D data plots of signs during motion capture

different angle orientations. In frame 3 and 4 the finger information is missed. But the 3D frame will provide the missing information because it captured in all orientations. Angle plots and marker position plots were obtained as shown in Fig. 7, which helps in detailed study and accurate classification of sign.

4 Conclusion

The results show the advantage of 3D motion capture in sign language recognition. It captures the data on a Cartesian coordinate which provides the information in all orientations. In 2D some finger's information is missing and blurring is also affecting the recognition. Motion capture is immune to blur, lighting, color change and self- and external occlusions. A 3D hand model was designed for marker placements to capture the signs meaningfully. At this juncture, we can conclude that 3D motion capture is best suitable for Indian sign language recognition. We further working to develop best algorithms to process and classify the signs of Indian sign language.

Declaration: The images used in the work are of private image and due permission has been taken. Authors of the paper bear all responsibilities if any issues arise due to this. Publisher will not be responsible for same.

References

1. S. Kwak, B. Han, J. Han, Scenario-based video event recognition by constraint flow, in: Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Colorado Springs, 2011, pp. 3345–3352, <http://dx.doi.org/10.1109/CVPR.2011.5995435>.
2. U. Gaur, Y. Zhu, B. Song, A. Roy-Chowdhury, A string of feature graphs model for recognition of complex activities in natural videos, in: Proceedings of International Conference on Computer Vision (ICCV), IEEE, Barcelona, Spain, 2011, pp. 2595–2602, <http://dx.doi.org/10.1109/ICCV.2011.6126548>.
3. S. Park, J. Aggarwal, Recognition of two-person interactions using a hierarchical Bayesian network, in: First ACM SIGMM International Workshop on Video surveillance, ACM, Berkeley, California, 2003, pp. 65–76, <http://dx.doi.org/10.1145/982452.982461>.
4. I. Junejo, E. Dexter, I. Laptev, P. Pérez, View-independent action recognition from temporal self-similarities, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (1) (2011) 172–185, <http://dx.doi.org/10.1109/TPAMI.2010.68>.
5. Z. Duric, W. Gray, R. Heishman, F. Li, A. Rosenfeld, M. Schoelles, C. Schunn, H. Wechsler, Integrating perceptual and cognitive modeling for adaptive and intelligent human–computer interaction, *Proc. IEEE* 90 (2002) 1272–1289, <http://dx.doi.org/10.1109/JPROC.2002.801449>.
6. Y.-J. Chang, S.-F. Chen, J.-D. Huang, A Kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities, *Res. Dev. Disabil.* 32 (6) (2011) 2566–2570, <http://dx.doi.org/10.1016/j.ridd.2011.07.002>.
7. A. Thangali, J.P. Nash, S. Sclaroff, C. Neidle, Exploiting phonological constraints for handshape inference in ASL video, in: Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Colorado Springs, 2011, pp. 521–528, <http://dx.doi.org/10.1109/CVPR.2011.5995718>.
8. A. Thangali Varadaraju, Exploiting phonological constraints for handshape recognition in sign language video (Ph.D. thesis), Boston University, MA, USA, 2013.
9. H. Cooper, R. Bowden, Large lexicon detection of sign language, in: Proceedings of International Workshop on Human–Computer Interaction (HCI), Springer, Berlin, Heidelberg, Beijing, P.R. China, 2007, pp. 88–97.
10. J.M. Rehg, G.D. Abowd, A. Rozga, M. Romero, M.A. Clements, S. Sclaroff, I. Essa, O.Y. Ousley, Y. Li, C. Kim, et al., Decoding children’s social behavior, in: Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Portland, Oregon, 2013, pp. 3414–3421, <http://dx.doi.org/10.1109/CVPR.2013.438>.
11. L. Lo Presti, S. Sclaroff, A. Rozga, Joint alignment and modeling of correlated behavior streams, in: Proceedings of International Conference on Computer Vision-Workshops (ICCVW), Sydney, Australia, 2013, pp. 730–737, <http://dx.doi.org/10.1109/ICCVW.2013.100>.
12. H. Moon, R. Sharma, N. Jung, Method and system for measuring shopper response to products based on behavior and facial expression, US Patent 8,219,438, July 10, 2012 (<http://www.google.com/patents/US8219438>).
13. G. Johansson, Visual perception of biological motion and a model for its analysis, *Percept. Psychophys.* 14 (2) (1973) 201–211.
14. Cerveri, P., De Momi, E., Lopomo, N. et al., Finger kinematic modelling and real-time hand motion estimation, *Ann Biomed Eng* (2007) 35: 1989. doi:[10.1007/s10439-007-9364-0](https://doi.org/10.1007/s10439-007-9364-0).

15. N. Miyata, M. Kouchi, T. Kurihara, and M. Mochimaru, "Modelling of human hand link structure from optical motion capture data," in Proc. Int. Conf. Intelligent Robots Systems, Sendai, Japan, 2004, pp. 2129–2135.
16. I. Carpinella, P. Mazzoleni, M. Rabuffetti, R. Thorsen, and M. Ferrarin, "Experimental protocol for the kinematic analysis of the hand: Definition and repeatability," *Gait Posture*, vol. 23, pp. 445–454, 2006.
17. G. Wu, F. C. T. van der Helm, H. E. J. Veeger, M. Makhsous, P. van Roy, C. Anglin, J. Nagels, A. R. Karduna, K. McQuade, X. Wang, F. W. Werner, and B. Buchholz, "ISB recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion—Part II: Shoulder, elbow, wrist and hand," *J. Biomech.*, vol. 38, pp. 981–992, 2005.