Vision-based Hand Posture Estimation using a New Hand Model Made of Simple Components

Shahrzad Saremi, Seyedali Mirjalili, Andrew Lewis

School of Information and Communication Technology, Griffith University, Brisbane, 4111 QLD, Australia

**seyedali.mirjalili@griffithuni.edu.au*

Abstract

Hand posture estimation is a building block in hand gesture detection systems. One of the most popular techniques in this field is to generate a 3D hand model and find its optimal structure using an optimisation algorithm. The main advantages of such methods are flexibility to model complex hand postures, ability to better address occlusions, and estimating both discrete and contentious poses. However, rendering and finding optimal structure for a 3D hand model are computationally expensive. Popular models in the literature are made of mesh or simple components. This work first analyzes a number of hand models in the literature. Then, a new hand model with simple components is proposed. The performance of the proposed model is benchmarked on 50 poses extracted from four standard datasets. The results show that the proposed 3D model shows superior results compared to the best models in the literature.

Keywords: Hand posture estimation, Particle Swarm Optimisation, PSO, Vision-based hand gesture detection

1. Introduction

There are different approaches for modeling hands in computers: motion-based models, skeleton models, volumetric models, geometric models, kinematic models, etc [1]. All these methods have their own advantages and disadvantages. Finding a suitable model of hand for an application is one challenge. Another challenge is to register and update the shape of the hand model in real time. In other words, the main difficulty is to find optimal values for the structural parameters of a model to match with the shape, orientation, and position of a hand in a 2D or 3D space. This problem can be considered as an optimisation problem.

The field of hand posture estimation has been very popular over the course of last decade and in the literature a significant number of studies have been undertaken. For instance, Ouhaddi and Horian employed mathematical optimisation algorithms such as Levenberg-Marquardt, downhill simplex, and Powell optimisation to find the optimal shape for a 3D hand model [2]. Hand-poses have been estimated using a 3D model fitting between the hand model and the voxel model [3] as well. Despite the success of these deterministic methods, they tend to stagnate in local solutions since the problem of hand posture estimation involves a large number of local solutions.

In 2011, a stochastic algorithm called Particle Swarm Optimisation (PSO) [4] was employed to minimise the discrepancy between the predicted and the actual (observed) features of a hand model [5]. A 3D hand model was constructed using other 3D primitive shapes and then optimised by PSO in this approach. In addition, PSO was employed to find the optimal shape for the model [6]. The same authors had adapted and modified an evolutionary algorithm called tailored Evolu-

tionary Quasi-random Search for hand articulation tracking in another work too [7]. They again found the optimal shape for a model of hand with 27 parameters (54 for two hands) and proved that stochastic optimisation techniques can be very efficient in this field. In 2015, the parameters of a 3D hand model have been optimised by the PSO algorithm per frame in conjunction with a robust fitting framework [8]. The framework estimates the best 3D hand model parameters within the history of frames obtained by PSO.

All the above works have proved that the problem of finding optimal parameters for a hand model is very challenging and stochastic optimisation techniques are essential in this field. They can be divided in three classes: model-based (generative), appearance-based (discriminative), and hybrid. A model-based approach generates a 3D model and finds its optimal structural parameters to estimate hand postures. However, appearance-based methods use images and repositories to search for a hand pose. Hybrid algorithms use both model-based and appearance-based methods sequentially or in parallel.

This paper focuses on model-based methods since they benefit from flexibility to model complex hand poses, ability to better address occlusions, and estimating both discrete and contentious poses. We target the main drawback of such methods, high computational cost of rendering and finding optimal structure for a 3D hand model, and proposed a new hand model with simple components. The rest of the work is organised as follows. Section 2 reviews the literature and identifies the current gaps. The new hand model is proposed in Section 3. Experimental results on 50 poses extracted from four standard datasets are discussed in Section 4. Section 5 concludes the work and suggest avenues for future research.

Table 1: Different models used in the literature in generative methods

Generative methods			
Few simple components (geometry-based)		Many simple components (mesh-based)

Advantages	Drawbacks	Advantages	Drawbacks
1. Rigidity during motion 2. Computationally cheap since the components do not change	1. Not very accurate for different hand sizes 2. Highly depends on initial solution 3. Difficulty in recovering from a wrong hand posture estimation or when there is a failure in hand tracking.	1. Accurate hand approximation 2. Suitable for different hand sizes.	1. Flexible during motion. 2. Creating polygons are computationally more expensive than simple components 3. Depends of initial solutions although it is less than geometry-based models
Drawbacks of both methods			
1. Expensive objective function 2. Sensitive to initial pose from the previous frame.			

2. Related work

Model-based fitting (hypothesize-and-test) methods include two main classes: models with limited simple components versus models made of meshes. The well-regarded 3D hand models in the literature are presented in Table 1. This table also discusses pros of cons of each class.

It may be seen that a 3D model of hand can be constructed with simple shapes (*e.g.* spheres [9]) or mesh [17]. In such models, simple 3D primitive shapes are mostly used: sphere, cylinder, cubes, etc. A skeleton is also used to define the relative position of each component. The skeleton defines the kinematic movement of all components. The advances of such models is that the shape of hand remains rigid when there is a movement, especially for abrupt movements. The computational cost of changing the shape of model and rendering is relatively cheap as well. However, these models are not very accurate for different hand sizes and highly depend on an initial pose. The initial pose here does not refer the first pose. It refers to the first pose that the optimisation starts with, which might be random or procedurally generated by an algorithm. As a consequence, hand recovery is also difficult and require a very good hand pose initializer.

In the mesh-based models, a mesh made of usually triangles is employed to model hand and estimate hand poses. Obviously, the main advantage is the accuracy and smoothness. Also, such models are suitable for estimating different hand sizes since they are able to shrink or grow. However, there are some drawbacks when using mesh-based methods. Firstly, such models are very flexible and deform during sudden motions. Secondly,

creating polygons to form the mesh and rendering are computationally more expensive than 3D hand models made of simple components. Finally, these methods depend on initial poses although less than hand models with simple components.

Any of these methods have their own pros and cons summarized in Table 2. The rigidity and computational cost of creating models with simple components make them more practical. In hand gesture recognition systems, feature selection and classification phases are computationally expensive. Therefore, if we can perform the hand modeling faster, it would be fruitful. The main drawbacks of both techniques (computation cost of objective function and sensitivity to initial pose as shown in Table 1) can be alleviated with more research in this area.

In each of the classes mentioned above, there is a lot of work in the literature using different models and algorithms. Despite the differences, the most common part is the use of kinematic model to simulate the movement of joints. The most popular models are 26 and 27 Degree Of Freedom (DOF) models. Fig. 1 illustrates an example of hand kinematic model with 26 DOF. Note the node in the palm have three DOF for rotation and three for translation in a 3D space.

The current works are mostly on stretching of the morphable mesh-shaped models to calibrate the hand before estimating its optimal shape. In 2006, Rhee *et al.* [18] first used skin color to find the location of joints and then mesh to fill silhouette (segmented image of hand to make it black and white). In other works, the vertices of hand models are changed to find a good calibration [10, 13]. Another attempt in personalizing hand pose using mesh-based models was done by Ballan *et al.* [12]. The authors showed that a robust hand pose esti-

Table 2: Pros and cons of the hand models in the literature

Few simple components (geometry-based)	Many simple components (mesh-based)

Rigid	Very accurate
Cheap	Suitable for different hand sizes
Sensitive to hand size	Flexible during abrupt motions
Sensitive to initial pose	Computationally expensive mesh
Difficulty in recovering from a wrong hand posture estimation or when there is a failure in hand tracking	Sensitive to initial pose
Expensive cost function	Expensive cost function

Figure 1: An example of hand kinematic model with 26 DOF

mation can be obtained with an accurate personalized 3D hand model. It was also showed that a personalized hand model for both hand improve the overall performance of a hand pose estimation system. However, the main issues were the need for multiple cameras and manual rigging of both hands.

Generally speaking, in mesh-based approaches, an optimisation method is used to first find the optimal shape for hand and then estimate its posture. One might say here that starching a mesh to customize the shape of human hand is suitable and we do not need to use simple shapes anymore since it is very accurate. According to Makris *et al.* in 2015 [19], however, shape and pose estimation are different phases and by nature different problems. They experimentally proved that separation of these two phases results in a more accurate hand pose estimation. In other words, they proved that separation of shape and pose estimation and treating them as two separate problems are beneficial.

Taken together, finding an efficient hand model in the field of hand posture estimation is still an open research question. Despite the merits of hand models with simple components, there

are a few models in the literature. Most of them suffer from several drawbacks: weak collision detection model, low flexibility, difficulty in extracting a well-distributed point cloud, and difficulty to use for different hand sizes.

This work proposes a new hybrid hand model using simple components to tackle some of the aforementioned problems. A better hand model to reduce the impacts of occlusion, speed of changing poses, and different hand sizes on the performance of hand gesture detection is a valuable long-term benefit of this work.

3. Proposed hand model with simple components

It was mentioned in the preceding section that hand models with simple shapes are rigid and computationally cheap to render. Among several models in the literature, two of the most successful ones have been proposed in [6, 10]. The main components of these two models are illustrated in Fig. 2. It may be seen that the model proposed by Oikonomidis *et al.* consists of cylinders, spheres, elliptical cylinders, and ellipsoids. However, the model proposed by Qian *et al.* only uses spheres. Despite the merits of these two hand models, each of them suffer from a number of drawbacks. The main problem of the first model is the palm. It includes one elliptical cylinder, which is not flexible and hard to get a uniformly distributed set of points to be used in the point cloud. Qian's model has a flexible palm due to the use of 16 spheres of different sizes. However, there is a lot of spheres for the fingers with small gaps between them or on their borders. A uniform set of solutions on each circle may results in many close points in point cloud. The thumb in Qian's model is better due to the flexibility as well.

To alleviate the drawbacks of both models, a hybrid model is proposed as can be seen in Fig. 3. This figure shows that the palm includes circles only, but the fingers are built with cylinders and spheres. The base of thumb is similar to that of Qian but there is a modification in which there are only three spheres. The main motivation of using this limited set of shapes is to

Figure 2: Current hand models with simple component (top) Oikonomidis *et al.* [6], (bottom) Qian *et al.* [10]

Figure 3: The components of the proposed hand model

Figure 4: Visual comparison of hand models (left) Oikonomidis *et al.* [6], (middle) Qian *et al.* [10], (right) the proposed hand model

benefit from the advantages of both hand models. To better compare the hand models, Fig. 5 is given.

Since hand is highly articulated, we have to consider a way to prevent collisions between fingers together and with the palm. In the literature several spheres are considered to prevent collision in the 3D hand model. The collision models of hand in literature and the proposed hand model are illustrated in Fig. 5.

This figure shows that in the Antonis's model, there are 25 spheres to detect collisions. One of the objections to this model is the big gaps between the collision spheres in the palm which might result in a finger passing through the palm. To alleviate this drawback, this work proposes a better collision detection system. It may be seen in the left part of Fig. 5 that there are 20 spheres in the palm, which make it almost impossible for a finger to pass through the palm. For every posture, the distances between all the spheres in different fingers and palm nodes are calculated. If the distance is less than the sum of radii between two spheres, there is a collision. Each hand model returns the number of collided spheres in the model and will be used in the penalty function for optimisation algorithms.

The proposed hand model has 26 variables as shown in Fig.

Figure 5: Collision spheres in the proposed hand model

6. It may be seen that the first three variables define the position of hand in the 3D space. Note that these three values define the position of wrist joint and the rest on joints will be drawn relative to the wrist. The second three values show rotation on hand along each axis: roll, pitch, and yaw. The remaining 20 variables are the rotation of all joints with respect to their DOF.

Obviously, the parameters for translation and rotation significantly change the objective value and define the global feature of the hand. The remaining 20 variables define the local features. In the next section, the effectiveness of the proposed hand model is benchmarked on 50 case studies.

Figure 6: The variable vector and its components in defining the final position, orientation, and shape of 3D model

4. Benchmarking the proposed hand model

In order to benchmark the performance of the proposed hand model, 50 hand postures are utilised from four standard datasets in the literature as shown in the appendix (see Fig. 9). PSO is used to find the optimal shape for all hand models including the proposed one. For every hand posture, PSO is run 30 times using 10 particles and 100 iterations to draw a reliable conclusion. The objective function is to minimise the discrepancy between the point cloud obtained from the 3D hand model and the depth image of hand. After all, the results are presented in Table 3 – 6. The results are arranged in four tables due to the large number of postures. Two tables show the average performance and two tables include the results of the Wilcoxon ranksum test to prove the significance of the results.

Although the distance between two point clouds can be minimised even further with increasing the number of particles and iterations, the main purpose of this experiment is to see which hand model gives a better point cloud and is more helpful for the optimiser under the same conditions.

The tables show that the proposed hand model is able to provide better results compared to the current 3D hand models in the literature of 3D models with simple components. The statistical results prove that this superiority is occasionally significant. The reason is due to the better distribution of point in the point cloud generated from the proposed 3D hand model. In other words, the proposed hand model aggregates the best components of the current hands models (palm from the first hand model and fingers from the second hand model).

The reason of the worse results obtained by the first hand model proposed by Oikonomidis *et al.* is mainly due to the palm. The palm is a deformed cylinder which give a non-rigid and convex shapes on both side of hand. However, an actual palm is flat or concave. This cannot be achieved by the Oikonomidis *et al.*. On the other hand, the proposed and Qian's model have a better palm created with many spheres. This gives the model more flexibility and adaptability to different hand poses with diverse palm orientation and shapes. The shape of palm is the only component of hand that changes shape and this can be addressed by the Qian's and proposed models. The reason of worse results of the Qian's hand model compared to the proposed one is the fingers. Fingers do not need to be very flexible and cylinders are able to give a uniformly distributed set of points along the finger. Also, in the proposed hand palm there very little gaps between the spheres while the Oikonomidis *et*

Table 3: Mean and standard deviation of PSO+EPD when estimating proposed hand model over 30 runs, Qian's hand model, and Oikonomidis' hand model for Pose1 – Pose25

Case study	Proposed model		Qian's model		Oikonomidis model	
	Mean	STD	Mean	STD	Mean	STD
Pose1	8.2877	1.1887	9.6005	2.6650	10.5298	3.8466
Pose2	10.5782	2.2068	12.4835	2.4916	11.2748	2.9439
Pose3	9.7457	1.1487	9.4281	1.7907	10.9107	1.2049
Pose4	11.7525	1.8358	10.8206	2.3967	10.4482	1.5459
Pose5	10.9421	2.1834	9.8700	1.5868	11.3083	2.5220
Pose6	10.9442	1.4105	11.2226	1.3763	11.0611	1.7483
Pose7	9.5182	1.3709	10.3092	0.9724	10.4425	0.6932
Pose8	11.8169	1.3809	10.7200	1.0364	11.2220	1.9576
Pose9	10.0103	0.8994	10.1154	0.7723	9.7722	1.3077
Pose10	11.1620	1.5940	10.9950	1.7753	11.0374	1.2205
Pose11	8.5142	1.1749	9.2201	2.1389	9.0601	2.0271
Pose12	10.2844	1.6015	10.3230	0.8498	11.1699	3.5334
Pose13	10.1371	2.7956	9.6429	1.9888	10.7071	2.7877
Pose14	10.4923	3.3984	9.6058	2.9267	9.4696	1.8079
Pose15	9.2062	0.6800	9.6717	1.7446	9.9169	1.4818
Pose16	9.3213	1.1225	10.2975	1.5554	10.4308	1.3698
Pose17	10.1493	1.5231	10.1190	1.8545	9.6179	1.3799
Pose18	8.9936	1.7106	8.6307	0.8217	8.6794	0.6325
Pose19	7.5254	1.2768	7.7426	1.2664	7.2883	0.6723
Pose20	9.8146	1.0960	10.6096	2.3593	9.0903	1.0925
Pose21	8.6265	2.1246	8.2086	0.9406	8.2280	0.7845
Pose22	10.4864	1.8352	9.7911	1.6613	9.9264	2.2595
Pose23	8.5314	1.0761	9.6469	0.6123	9.9935	1.3873
Pose24	11.3369	0.9491	11.8063	1.0963	11.6274	1.2524
Pose25	10.1665	1.3801	9.8496	0.9050	10.5600	1.3357

Table 4: Mean and standard deviation of PSO+EPD when estimating proposed hand model over 30 runs, Qian's hand model, and Oikonomidis' hand model for Pose26 – Pose50

Case study	Proposed model		Qian's model		Oikonomidis model	
	Mean	STD	Mean	STD	Mean	STD
Pose26	7.4210	1.1275	8.1726	1.0302	8.4713	2.6736
Pose27	9.2959	1.8895	9.2923	0.8422	8.9581	0.9557
Pose28	11.0374	1.3329	10.1123	2.2749	9.4699	1.2087
Pose29	9.4083	0.9223	10.2393	1.5004	10.3738	1.1404
Pose30	9.8812	1.6878	9.8850	1.6592	8.6063	1.5608
Pose31	7.9240	1.5711	7.4660	0.7813	7.0028	0.7537
Pose32	8.3063	0.8703	8.2791	0.9933	8.9494	1.1156
Pose33	7.0367	0.5279	7.6762	0.8475	7.9341	1.6169
Pose34	10.4931	1.1092	11.0054	0.7353	10.5266	1.1178
Pose35	8.8473	1.5723	9.3617	1.5933	7.7291	0.8199
Pose36	10.5204	1.7387	10.8585	1.0550	10.2719	1.0894
Pose37	9.7579	0.9252	10.2141	1.0229	9.9435	1.4166
Pose38	9.7322	1.1786	9.8835	1.0468	10.5781	1.2712
Pose39	8.1375	1.1064	9.2186	1.3278	8.4705	1.4098
Pose40	8.1664	1.7748	10.3368	2.6938	10.7061	3.0760
Pose41	10.0902	1.1028	10.7282	1.4308	10.5817	1.2962
Pose42	10.4752	1.7755	10.6269	0.4663	10.1712	0.8214
Pose43	10.8881	1.6034	11.1536	1.1463	11.3683	1.2986
Pose44	8.1568	0.9867	8.9428	1.9549	7.7998	1.2323
Pose45	8.3096	1.2152	8.7789	1.7618	8.5527	1.6878
Pose46	9.6638	1.6965	10.4524	1.0987	9.7581	1.5265
Pose47	9.0038	1.5823	9.1192	0.8329	8.6785	0.6801
Pose48	10.7837	1.1268	10.9083	1.2475	11.5591	2.0793
Pose49	9.6674	1.1506	10.4162	1.3804	10.0176	0.8952
Pose50	11.2525	0.6526	11.5231	1.3746	11.3974	1.4390

Table 5: P-values of Wilcoxon ranksum test PSO+EPD when estimating proposed hand model over 30 runs, Qian's hand model, and Oikonomidis' hand model for Pose1 – Pose25

Case study	Proposed model	Qian's model	Oikonomidis' model
Pose1	N/A	0.2413	0.1859
Pose2	N/A	0.0757	0.7337
Pose3	0.1859	N/A	0.0211
Pose4	0.1041	0.9097	N/A
Pose5	0.3447	N/A	0.3075
Pose6	N/A	0.6232	0.7913
Pose7	N/A	0.1859	0.0757
Pose8	0.0539	N/A	0.8501
Pose9	0.8501	0.9698	N/A
Pose10	0.8501	N/A	0.9698
Pose11	N/A	0.8501	0.9097
Pose12	N/A	0.9097	0.9097
Pose13	0.9698	N/A	0.4274
Pose14	0.7337	0.7337	N/A
Pose15	N/A	0.5205	0.3447
Pose16	N/A	0.1041	0.0539
Pose17	0.3447	0.7913	N/A
Pose18	0.6776	N/A	0.8501
Pose19	0.7337	0.3847	N/A
Pose20	0.1620	0.0757	N/A
Pose21	0.7913	N/A	0.7913
Pose22	0.3847	N/A	0.4727
Pose23	N/A	0.0376	0.0173
Pose24	N/A	0.2730	N/A
Pose25	0.5205	N/A	0.1405

Table 6: P-values of Wilcoxon ranksum test PSO+EPD when estimating proposed hand model over 30 runs, Qian's hand model, and Oikonomidis' hand model for Pose26 – Pose50

Case study	Proposed model	Qian's model	Oikonomidis' model
Pose26	N/A	0.1405	0.5708
Pose27	0.7913	0.3847	N/A
Pose28	0.0312	0.7337	N/A
Pose29	N/A	0.1859	0.0757
Pose30	0.0257	0.0376	N/A
Pose31	0.3075	0.1620	N/A
Pose32	0.9097	N/A	0.2123
Pose33	N/A	0.1620	0.2413
Pose34	N/A	0.2123	0.6776
Pose35	0.0890	0.0073	N/A
Pose36	0.6776	0.2730	N/A
Pose37	N/A	0.2413	0.5708
Pose38	N/A	0.7913	0.1212
Pose39	N/A	0.0757	0.6232
Pose40	N/A	0.1620	0.0452
Pose41	N/A	0.3075	0.3447
Pose42	0.7913	0.2730	N/A
Pose43	N/A	0.8501	0.7913
Pose44	0.3847	0.1620	N/A
Pose45	N/A	0.4274	0.7913
Pose46	N/A	0.1859	0.7913
Pose47	0.4274	0.3447	N/A
Pose48	N/A	0.9698	0.6232
Pose49	N/A	0.3447	0.7913
Pose50	N/A	0.6776	N/A

al. hand model has several gaps. This will be an issue when using silhouette of hand in calculating the objective function. After all, these results show that the proposed hand model can be very efficient and is worth of consideration when estimating hand postures.

The p-values obtained from the Wilcoxon ranksum test are shown in Table 5 and 6. These tables show that the best 3D hand model for poses are occasionally significant. This is due to the similarity of the 3D hand models — they are all made of simple components. However, slight superiority of the propped 3D hand model is evident.

To see how the PSO algorithm improves the 3D hand models in every postures, the convergence curves are depicted in Fig. 7 and 8. It may be seen that in the majority of postures, PSO better finds an optimal posture for the proposed 3D hand model. The comparative results of all hand models can be observed in a number of postures as well: Pose9, Pose10, Pose11, Pose18, Pose19, Pose36, Pose37, Pose49, and Pose50. The second best hand model is the Qian's model.

Taken together, the results showed that the current hand models with simple components can be improved due to several drawbacks. The proposed hand model benefits from hybridizing the best components of the current hand models. Due to the use of simple components, the computational cost is low. Also, the structure of the proposed hand model allows better adaptability to different hand sizes. Therefore, we recommend it to be used in different hand gesture detection systems for estimating hand pose and/or hand shape.

5. Conclusion

This work first identified the current gap in the literature of 3D hand model when using model-based techniques. I was discussed that there should be more hand model with simple components to alleviate the drawbacks of the current 3D hand models: non-uniform point cloud and inaccurate collision detection. A hybrid 3D hand model was then proposed as an alternative to the current hand models. The proposed 3D model was tested on 50 postures extracted from four standard data sets. The PSO algorithm was employed to find the optimal shape. A comparative study was conducted with two hand models in the literature and the results were collected qualitatively and quantitatively. The results and finding proved the reliability and efficiency of the proposed hand model.

For future works, it is recommended to test size adaptation of the proposed hand model in hand shape estimation.

References

- [1] Siddharth S Rautaray and Anupam Agrawal. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, 43(1):1–54, 2015.
- [2] Hocine Ouahdi and Patrick Horain. 3d hand gesture tracking by model registration. In *Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging*, pages 70–73, 1999.
- [3] Etsuko Ueda, Yoshio Matsumoto, Masakazu Imai, and Tsukasa Ogasawara. A hand-pose estimation for vision-based human interfaces. *Industrial Electronics, IEEE Transactions on*, 50(4):676–684, 2003.

Figure 7: Average convergence of PSO+EPD when estimating proposed hand model over 30 runs, Qian's hand model, and Oikonomidis' hand model for Pose1 – Pose25

Figure 8: Average convergence of PSO+EPD when estimating proposed hand model over 30 runs, Qian's hand model, and Oikonomidis' hand model for Pose26 – Pose50

- [4] James Kennedy. Particle swarm optimization. In *Encyclopedia of Machine Learning*, pages 760–766. Springer, 2010.
- [5] Iasonas Oikonomidis, Nikolaos Kyriazis, and Antonis A Argyros. Markerless and efficient 26-dof hand pose recovery. In *Computer Vision–ACCV 2010*, pages 744–757. Springer, 2011.
- [6] Iason Oikonomidis, Nikolaos Kyriazis, and Antonis A Argyros. Efficient model-based 3d tracking of hand articulations using kinect. In *BMVC*, volume 1, page 3, 2011.
- [7] Iason Oikonomidis, Manolis Lourakis, Antonis Argyros, et al. Evolutionary quasi-random search for hand articulations tracking. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 3422–3429. IEEE, 2014.
- [8] Alexandros Makris and Antonis Argyros. Model-based 3d hand tracking with on-line shape adaptation. In Mark W. Jones Xianghua Xie and Gary K. L. Tam, editors, *Proceedings of the British Machine Vision Conference (BMVC)*, pages 77.1–77.12. BMVA Press, September 2015.
- [9] Rui Wang, Kun Zhou, John Snyder, Xinguo Liu, Hujun Bao, Qunsheng Peng, and Baining Guo. Variational sphere set approximation for solid objects. *The Visual Computer*, 22(9-11):612–621, 2006.
- [10] Chen Qian, Xiao Sun, Yichen Wei, Xiaou Tang, and Jian Sun. Realtime and robust hand tracking from depth. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1106–1113, 2014.
- [11] Andrea Tagliasacchi, Matthias Schröder, Anastasia Tkach, Sofien Bouaziz, Mario Botsch, and Mark Pauly. Robust articulated-icp for real-time hand tracking. In *Computer Graphics Forum*, volume 34, pages 101–114. Wiley Online Library, 2015.
- [12] Luca Ballan, Aparna Taneja, Jürgen Gall, Luc Van Gool, and Marc Pollefeys. Motion capture of hands in action using discriminative salient points. In *European Conference on Computer Vision*, pages 640–653. Springer, 2012.
- [13] Toby Sharp, Cem Keskin, Duncan Robertson, Jonathan Taylor, Jamie Shotton, David Kim, Christoph Rhemann, Ido Leichter, Alon Vinnikov, Yichen Wei, et al. Accurate, robust, and flexible real-time hand tracking. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3633–3642. ACM, 2015.
- [14] Stan Melax, Leonid Keselman, and Sterling Orsten. Dynamics based 3d skeletal hand tracking. In *Proceedings of Graphics Interface 2013*, pages 63–70. Canadian Information Processing Society, 2013.
- [15] Srinath Sridhar, Antti Oulasvirta, and Christian Theobalt. Interactive markerless articulated hand motion tracking using rgb and depth data. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2456–2463, 2013.
- [16] Philip Krejov, Andrew Gilbert, and Richard Bowden. Combining discriminative and model based approaches for hand pose estimation. In *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, volume 1, pages 1–7. IEEE, 2015.
- [17] Jonathan Taylor, Richard Stebbing, Varun Ramakrishna, Cem Keskin, Jamie Shotton, Shahram Izadi, Aaron Hertzmann, and Andrew Fitzgibbon. User-specific hand modeling from monocular depth sequences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 644–651, 2014.
- [18] Taehyun Rhee, Ulrich Neumann, and John P Lewis. Human hand modeling from surface anatomy. In *Proceedings of the 2006 symposium on Interactive 3D graphics and games*, pages 27–34. ACM, 2006.
- [19] Alexandros Makris and A Argyros. Model-based 3d hand tracking with on-line hand shape adaptation. *Proc. BMVC*, pages, pages 77–1, 2015.
- [20] Jonathan Tompson, Murphy Stein, Yann Lecun, and Ken Perlin.
- [21] Srinath Sridhar, Antti Oulasvirta, and Christian Theobalt. Interactive markerless articulated hand motion tracking using rgb and depth data. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2013.
- [22] Anastasia Tkach, Mark Pauly, and Andrea Tagliasacchi. Sphere-meshes for real-time hand modeling and tracking. *ACM Transactions on Graphics (TOG)*, 35(6):222, 2016.

Figure 9: Case studies: 50 postures extracted from four datasets [20, 21, 22]

6. Appendix

Dataset used: