# **3D Photorealistic Human Body Modelling and Reconstruction Techniques**

Subjects: Computer Science, Interdisciplinary Applications Contributor: Radostina Petkova, Ivaylo Bozhilov, Desislava Nikolova, Ivaylo Vladimirov, Agata Manolova

The continuous evolution of video technologies is now primarily focused on enhancing 3D video paradigms and consistently improving their quality, realism, and level of immersion. Both the research community and the industry work towards improving 3D content representation, compression, and transmission. Their collective efforts culminate in the striving for real-time transfer of volumetric data between distant locations, laying the foundation for holographic-type communication (HTC).

Keywords: human body modelling ; human body reconstruction ; holographic-type communication ; 3D avatars ; deepbased human body reconstruction

### 1. Introduction

Technological advancements have initiated the era of HTC. As explained in [1], HTC involves the transition from one person's actual location to another without the need for a physical traversal of the intervening space. However, the actual enablement of a truly immersive holographic experience necessitates the creation of convincing Mixed Reality (MR) environments, incorporating virtual elements and lifelike human avatars. The support of natural interactions between the virtual participants (the avatars), manipulated by real individuals, is one of the greatest aspects distinguishing future HTC from conventional voice and video-based modes of communication. Moreover, besides HTC, many other applications in the field of healthcare, such as remote consultations, surgical training, remote collaboration, remote rehabilitation, etc. [2][3] [4][5][6][7][8]; education, including remote collaborative learning, remote guest speakers, anatomy education, etc. [9][10][11][12] [13]; entertainment, such as interactive storytelling, Augmented Reality (AR)/Virtual Reality (VR) gaming, live concerts, etc. [14][15][16][17][18]; and e-commerce, in particular virtual try-on [19][20][21] will benefit from a visually authentic and interactive human appearance. Both academia and industry attempt to automate the detailed acquisition of 3D human pose and shape. The availability of sophisticated 3D acquisition equipment and powerful reconstruction algorithms has made realistic avatar generation possible. In fact, significant advancements are made in this field as digital avatars progressively acquire greater lifelike qualities, leading to increased trust among individuals. However, replicating real social interactions, including eye contact, body language, and conveying emotions through nonverbal cues (such as touch) and social signals (such as coexistence, closure, and intimacy), remains a formidable challenge, even with the current state of technological advancements. Therefore, the creation of lifelike human avatars that accurately represent both human appearance and behavior is a prominent subject within the realm of holographic experiences.

So far, a multitude of distinct methods for 3D human modelling and reconstruction have been developed, and tremendous efforts are still ongoing in this direction. However, the existing methods exhibit significant diversity in terms of whether they employ a parametric model, the chosen reconstruction approach, the dataset utilized, and, most notably, the input data type. Previous surveys have placed emphasis on variations in the parametric modelling of the 3D human body shape <sup>[22]</sup>, as well as on the types of reconstruction approaches, such as traditional, regression-based, or optimization-based methods, among others <sup>[23][24][25][26]</sup>.

## 2. Parametric Human Body Models

Building lifelike human avatars and the subsequent animation is one of the main challenges facing HTC. There is a need for creating accurate representations of HTC participants, which necessitates the detailed reconstruction of 3D digital human models. The generation of such models requires individual- or population-based anthropometric data. Anthropometric measurements are used to describe a person's physical appearance. They are estimates of the distances (both linear and curved) between anatomical landmarks or circumferences at specific human body regions of interest. Height (stature), weight (body mass), upright sitting height, triceps skinfold (upper arm girth), arm circumference (upper

arm girth), abdominal circumference (waist circumference), calf circumference, knee height, and elbow breadth are all common anthropometric measurements <sup>[27]</sup>. The anthropometric database must be extremely thorough in order to be credible for a specific group and to account for multivariate coherences.

Constructing an accurate human body model from various types of input data, such as single images, multi-view images, videos, or depth maps, is a great challenge. Existing methods for fitting a pose to the input data typically rely on parametric, yet statistical, human body models. Such an approach usually requires the indication of body joints, which is mostly carried out manually, but automatic and semi-automatic methods <sup>[28]</sup> also exist. Further, deep neural networks have been recently used to compute statistical models' parameters <sup>[29]</sup>. These types of modelling techniques have become an integral component in the recent methods for 3D human body reconstruction and animation. Here, the researchers present two of the most popular statistical body models and one that is a promising improvement of the second presented model.

#### 3. Human Body Datasets

Human bodies are flexible, moving in various ways and deforming their clothing and muscles. Another complicating issue, like the occlusion of different body parts during movement, may necessitate comprehensive scene modelling in addition to the peoples in the scenario. Such image understanding scenarios push, for example, the avatar body animation system's ability to use prior knowledge and structural correlations by constraining estimates of unseen body components using limited visible information. Insufficient data coverage is one of the most significant issues for trainable systems. So, many researchers have concentrated their efforts on creating publicly available datasets that can be used to build operational systems for realistic scenarios.

One of the largest motion capture datasets is the Human 3.6 M dataset <sup>[30]</sup>. It consists of 3.6 million fully visible human poses and corresponding images. All of them are captured by a high-speed motion capture system. The recording setup consists of 15 sensors (4 calibrated high-resolution progressive scan cameras that acquire video data at 50 Hz, 1 time-of-flight sensor, and 10 motion cameras), using hardware and software synchronization. This allows for accurate capture and synchronization. The dataset contains activities performed by 11 professional actors (6 male, 5 female) in 17 scenarios—taking photos, discussing, smoking, talking on the phone, etc. Also, accurate 3D joint positions and joint angles from high-speed motion capture systems are provided. Other useful additions are 3D laser scans of the actors, high-resolution videos, and accurate background subtraction.

The MPI-INF-3DHP <sup>[31]</sup> is a 3D human body pose estimation dataset. It consists of both constrained indoor and complex outdoor scenes. The dataset comprises eight actors (4 male, 4 female) enacting eight distinct activity sets, each lasting about a minute. With a diverse camera setup, 14 cameras in total, over 1.3 M frames have been obtained, with 500 k originating from cameras at chest height. The dataset provides genuine 3D annotations and a skeleton compatible with the "universal" skeleton of Human 3.6 M. To bridge the gap between studio and real-world conditions, chroma-key masks are available, facilitating extensive scene augmentation. The test set, enriched with various motions, camera viewpoints, clothing varieties, and outdoor settings, aims to challenge and benchmark pose estimation algorithms.

The Synthetic Humans for Real Tasks (SURREAL) dataset <sup>[32]</sup> contains 6.5 M frames of synthetic humans, organized into 67,582 continuous sequences. The SMPL <sup>[33]</sup> body model is employed to generate these synthetic bodies, with body deformations distinguished by pose and intrinsic shape. Created in 2017, it is the first large-scale person dataset to generate depth, body parts, optical flow, 2D/3D pose, surface, normal, and ground truth for Reed Green Blue (RGB) video input. The provided images are photorealistic renderings of people in different shapes, textures, viewpoints, and poses.

Dynamic Fine Alignment Using Scan Texture (DFAUST) <sup>[34]</sup> is considered a 4D dataset. It consists of high-resolution 3D scans of moving non-rigid objects, captured at 60 fps. A new mesh registration method is proposed. It uses both 3D geometry and texture information to register all scans in a sequence according to a common reference topology. The method makes use of texture constancy across short and long time intervals, as well as dealing with temporal offsets in shape and texture.

Microsoft Common Objects in Context (MS COCO)<sup>[35]</sup> is a large-scale object detection, segmentation, and captioning dataset. It consists of many other objects, but also humans and human photos. The dataset offers recognition in context, superpixel stuff segmentation, and 250,000 people with key points.

Leeds Sports Pose (LSP) <sup>[36]</sup> and its extended version—LSPe <sup>[37]</sup>— are human body joint detection datasets. The LSPe dataset contains 10,000 images gathered from Flickr searches for the tags 'parkour', 'gymnastics', and 'athletics' and uses

poses that are challenging to estimate. Each image has a corresponding annotation that might not be highly accurate because it is gathered from Amazon Mechanical Turk. Each image is annotated with up to 14 visible joint locations.

The Bodies Under Flowing Fashion (BUFF) dataset, as delineated by the authors of <sup>[38]</sup>, offers over 11,000 3D human body models engaged in complex movements. It is distinctive in its inclusion of videos featuring individuals in clothing paired with 3D models devoid of clothing textures. This dataset emerges from a multi-camera active stereo system, utilizing 22 pairs of stereo cameras, color cameras, speckle projectors, and white-light LED panels operating at varied frame rates. This system outputs 3D meshes averaging around 150 K vertices, capturing subjects in two distinct clothing styles. Of the initial six subjects, the data from one were withheld, resulting in a public release of 11,054 scans. To derive a semblance of "ground truth", the subjects were captured in minimal attire, with the dataset's accuracy showcased by the proximity of more than half of the scan points to the mean of the estimates. The BUFF dataset efficiently captures detailed aspects of human movement while also considering the impact of different clothing on a body's shape and motion.

The HumanEva datasets <sup>[39]</sup>, comprising HumanEva-I and HumanEva-II, offer a blend of video recordings and motion capture data from subjects performing predefined actions. HumanEva-I encompasses data from four subjects executing six distinct actions, each with synchronized video and motion capture, and one with only motion capture. This dataset leverages seven synchronized cameras, utilizing multi-view video data coupled with pose annotations. On the other hand, HumanEva-II focuses on two subjects, both of whom are also present in HumanEva-I, performing an extended "Combo" sequence, resulting in roughly 2500 synchronized frames. The data, collected under controlled indoor conditions, capture the intricacies of natural movement, albeit with challenges posed by illumination and grayscale imagery.

The UCLA Human–Human–Object Interaction (HHOI) dataset <sup>[40]</sup> is a novel RGB-D (Reed Green Blue—Depth) video collection detailing both human–human and human–object–human interactions, captured using a Microsoft Kinect v2 sensor. Comprising three human–human interactions—hand shakes, high-fives, and pull ups-—and two human–object–human interactions—-throw and catch and hand over a cup—the dataset features an average of 23.6 instances for each interaction. These instances are performed by eight actors, recorded from multiple angles, and spanning 2–7 s at a frame rate of 10–15 fps. While objects within the dataset are discerned using background subtraction on both RGB and depth images, the Microsoft Kinect v2's skeleton estimation is also utilized. The dataset, divided into four distinct folds for training and testing, ensures no overlap of actor combinations between the sets. The training algorithm demonstrates robust convergence within 100 iterations, operating on a standard 8-core 3.6 GHz and yielding an average synthesis speed of 5 fps using an unoptimized Matlab code.

To address prevalent challenges in viewpoint invariant pose estimation, a novel technical solution has been presented in <sup>[41]</sup>. It integrates local pose details into a learned, viewpoint-invariant feature space. This approach enhances the iterative error feedback model to incorporate higher-order temporal dependencies and adeptly manage occlusions via a multi-task learning methodology. Complementing this endeavor is the introduction of the Invariant Top View (ITOP) dataset, a comprehensive collection of 100 K depth images capturing 20 individuals across 15 diverse actions, encompassing a wide range of views, from front, top, to side, inclusive of occluded body segments. Each image in the ITOP dataset is meticulously labeled with precise 3D joint coordinates relative to the camera's perspective. With its unique blend of front/side and top views—the latter captured from ceiling-mounted cameras—the ITOP dataset stands as a significant resource for benchmarking and furthering advancements in viewpoint-independent pose estimation.

The 3D Human Body Model dataset established by the authors of <sup>[42]</sup> is a synthetic dataset that consists of 20,000 threedimensional models of human bodies in static poses and an equal gender distribution. It is generated with the STAR parametric model <sup>[43]</sup>. While generating the models, two primary considerations were maintained: the natural Range of Motion (ROM) for each joint and the prevention of self-intersections in the 3D mesh. Existing research on the human ROM was referenced to define the limitations of joint rotations. Despite adhering to ROM constraints, certain non-idealities sometimes result in self-intersections in areas like the pelvic region, knees, and elbows. To address this, each vertex of the mesh is associated with a specific bone group. Self-intersections between non-adjacent bone groups are considered forbidden, and an algorithm flags such meshes as invalid.

The 3D Poses in the Wild Dataset (3DPW) <sup>[44]</sup> offers a unique perspective by capturing scenarios in challenging outdoor environments. This extensive dataset encompasses more than 51,000 frames featuring seven different actors donning 18 distinct clothing styles. The data collection process involves the use of a handheld smartphone camera to record the actions of one or two actors. Notably, 3DPW enhances its utility by providing highly accurate mesh ground truth annotations. These annotations are generated by fitting the SMPL model to the raw ground truth markers.

The Max Planck Institute for Informatics (MPII) dataset serves to evaluate the accuracy of articulated human pose estimation. This dataset comprises approximately 25,000 images, featuring annotations for over 40,000 individuals, including their body joints. These images were systematically compiled, capturing a wide array of everyday human activities. In total, the dataset encompasses 410 different human activities, with each image labeled according to the specific activity depicted. The images are extracted from YouTube videos.

#### 4. Evaluation Metrics

Applying evaluation metrics is crucial to quantitatively assessing the reconstruction quality of generated models in the field of 3D human body reconstruction. Here, the researchers briefly introduce some of the most commonly utilized metrics for this purpose. These metrics provide a standardized and objective means of evaluating the accuracy and fidelity of reconstructed 3D human body models, ensuring a reliable assessment of the reconstruction process.

Mean Per-Joint Position Error (MPJPE)

The MPJPE <sup>[45]</sup> is a common metric that evaluates the performance of human pose estimation algorithms. It measures the mean distance in mm between the skeleton joints of the ground truth 3D pose and the joints from the estimated pose. The formulation is provided in the following Equation:

$${E_{MPJPE}} = rac{1}{{{N_S}}}\sum\limits_{i = 1}^{{N_S}} {{{{\left\| {{m_{{f,S}}^{\left( f 
ight)}}\left( i 
ight) - {m_{{g_{t},S}}^{\left( f 
ight)}}\left( i 
ight) } 
ight\|_2}}$$

where *NS* corresponds to the total number of skeleton joints, m(f)f,S(i) is a function that returns the coordinates of the *i*-th joint of skeleton *S* in frame *f*, and m(f)gt,S(i) is a function that refers to the *i*-th joint of the skeleton in the ground truth frame. A commonly used modification of the MPJPE is the Procrustes-aligned MPJPE (PA-MPJPE), which is calculated in a similar way with the difference that the reconstructed model and the ground truth one are previously aligned using the Procrustes algorithm.

Mean Average Vertex Error (MAVE)

The MAVE <sup>[46]</sup> is used to find the averaged distance between the vertices of the reconstructed 3D human model and the vertices of the ground truth data. It is defined by

$$E_{MAVE} = rac{1}{N}\sum_{i=1}^{N} \sqrt{ig\|artheta_i - \overline{artheta}_lig\|_2^2}$$

where *N* is the total number of vertices of the 3D model,  $\vartheta i$  is a vertex from the predicted 3D human body model, and  $\vartheta i^{-}$  is a vertex from the corresponding ground truth data.

Chamfer Distance

The symmetric point-to-point Chamfer distance measures the similarity between two point clouds P and Q. A common formulation is given in the following Equation:

$$d_{CD}\left(P,Q
ight) = rac{1}{|P|} \sum_{x \in P} \min_{y \in Q} \, ||x-y||_2^2 + rac{1}{|Q|} \sum_{y \in Q} \min_{x \in P} \, ||x-y||_2^2$$

Vertex-to-Surface Distance (VSD)

The VSD metric quantifies the average distance between the vertices of a point cloud and their corresponding points on the surface of a triangular mesh. These surface points can either be the vertices of the mesh or points that reside on its faces or edges. The authors of <sup>[47]</sup> incorporate this metric into their 3D model-fitting algorithm, employing a lifted optimization technique. Points on the surface are defined via surface coordinates, represented as  $u=\{p,v,w\}$ . Here,  $p\in\mathbb{N}$  denotes the triangle's index where the point is situated, and  $v\in[0,1]$ ,  $w\in[0,1-v]$  are the coordinates within the unit triangle. Therefore, the 3D coordinates of a point on the surface can be defined as shown in the following Equation, where v1,v2 and v2 are the vertices of the *p*-th triangle.

$$S\left(u
ight)=\left(1-v-w
ight)\mathbf{v}_{1}+v\mathbf{v}_{2}+w\mathbf{v}_{3}$$

The distance between a point from a point cloud and its correspondence on the surface of the mesh can be further calculated as described in the following Equation, where *D* is the number of points xi in the point cloud and  $U=\{ui\}Di=1$  are

the surface coordinates of those points.

$$E_{VDS}\left(U
ight)=rac{1}{D}\sum_{i=1}^{D} ert ert S\left(u_{i}
ight)-x_{i}ert ert^{2}$$

Within the confines of the algorithm described in  $\frac{[47]}{}$ , the mesh under consideration is parametric. This implies that its vertex coordinates are contingent on the parameter vector,  $\theta$ . Given that the algorithm adopts lifted optimizations, both  $\theta$  and the surface coordinates, *U*, are optimized concurrently.

#### 5. Taxonomy of Existing 3D Human Body Modelling and Reconstruction Techniques

**Figure 1** presents a taxonomy of existing 3D human body modeling and reconstruction techniques based on the input data. In essence, 3D human body modeling is a task in computer vision and computer graphics aimed at generating a 3D photorealistic representation of the human body. This task often involves, but is not limited to, processes such as data acquisition, 2D pose estimation through the detection of 2D body joints, camera calibration and triangulation in the case of multiple views, 3D pose estimation to derive the pose of the future 3D model, model shape optimization when using parametric body models, surface reconstruction and texturing for obtaining a detailed 3D representation of the model's surface, and post-processing and refinement to increase the quality of the reconstructed 3D model. However, it is challenging to compose a concrete framework of operations that applies universally across different input data types and reconstruction approaches. Nevertheless, several key steps, which are visualized in **Figure 2**, are commonly accomplished in most of the examined methods, including 3D pose estimation, coarse 3D shape estimation, 3D shape refinement, and texture recovery. The blocks surrounded with dash lines indicate that the utilization of the parametric body model and the appliance of 3D pose estimation is sometimes omitted by some of the examined algorithms, most of which are deep learning based. The figure also illustrates the different types of input data.



Figure 1. Taxonomy of 3D human body modelling and reconstruction techniques based on input data.



#### References

- 1. Manolova, A.; Tonchev, K.; Poulkov, V.; Dixir, S.; Lindgren, P. Context-aware holographic communication based on semantic knowledge extraction. Wirel. Pers. Commun. 2021, 120, 2307–2319.
- Haleem, A.; Javaid, M.; Khan, I.H. Holography applications toward medical field: An overview. Indian J. Radiol. Imaging 2020, 30, 354–361.
- 3. Jumreornvong, O.; Yang, E.; Race, J.; Appel, J. Telemedicine and medical education in the age of COVID-19. Acad. Med. 2020, 95, 1838–1843.
- 4. Nayak, S.; Patgiri, R. 6G communication technology: A vision on intelligent healthcare. Health Inform. Comput. Perspect. Healthc. 2021, 1–18.
- Ahmad, H.F.; Rafique, W.; Rasool, R.U.; Alhumam, A.; Anwar, Z.; Qadir, J. Leveraging 6G, extended reality, and IoT big data analytics for healthcare: A review. Comput. Sci. Rev. 2023, 48, 100558.
- 6. Ahad, A.; Tahir, M. Perspective—6G and IoT for Intelligent Healthcare: Challenges and Future Research Directions. ECS Sens. Plus 2023, 2, 011601.
- Bucioli, A.A.; Cyrino, G.F.; Lima, G.F.; Peres, I.C.; Cardoso, A.; Lamounier, E.A.; Neto, M.M.; Botelho, R.V. Holographic real time 3D heart visualization from coronary tomography for multi-place medical diagnostics. In Proceedings of the 2017 IEEE 15th International Conference on Dependable, Autonomic and Secure Computing, 15th International Conference on Pervasive Intelligence and Computing, 3rd International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), Orlando, FL, USA, 6–10 November 2017; pp. 239–244.
- Sirilak, S.; Muneesawang, P. A new procedure for advancing telemedicine using the HoloLens. IEEE Access 2018, 6, 60224–60233.
- 9. Choi, P.J.; Oskouian, R.J.; Tubbs, R.S.; Choi, P.J.K. Telesurgery: Past, present, and future. Cureus 2018, 10, e2716.
- Barkhaya, N.M.M.; Abd Halim, N.D. A review of application of 3D hologram in education: A meta-analysis. In Proceedings of the 2016 IEEE 8th International Conference on Engineering Education (ICEED), Kuala Lumpur, Malaysia, 7–8 December 2016; pp. 257–260.
- Ramachandiran, C.R.; Chong, M.M.; Subramanian, P. 3D hologram in futuristic classroom: A review. Period. Eng. Nat. Sci. 2019, 7, 580–586.
- 12. Ahmad, A.S.; Alomaier, A.T.; Elmahal, D.M.; Abdlfatah, R.F.; Ibrahim, D.M. EduGram: Education Development Based on Hologram Technology. Int. J. Online Biomed. Eng. 2021, 17, 32–49.
- 13. Yoo, H.; Jang, J.; Oh, H.; Park, I. The potentials and trends of holography in education: A scoping review. Comput. Educ. 2022, 186, 104533.
- 14. Hughes, A. Death is no longer a deal breaker: The hologram performer in live music. Future Live Music 2020, 114–128. Available online: https://books.google.bg/books?id=QB3LzQEACAAJ (accessed on 20 October 2023).
- Matthews, J.; Nairn, A. Holographic ABBA: Examining Fan Responses to ABBA's Virtual "Live" Concert. Pop. Music Soc. 2023, 1–22.
- 16. Rega, F.; Saxena, D. Free-roam virtual reality: A new avenue for gaming. In Advances in Augmented Reality and Virtual Reality; Springer: Berlin/Heidelberg, Germany, 2022; pp. 29–34.
- 17. Fanini, B.; Pagano, A.; Pietroni, E.; Ferdani, D.; Demetrescu, E.; Palombini, A. Augmented Reality for Cultural Heritage. In Springer Handbook of Augmented Reality; Springer: Berlin/Heidelberg, Germany, 2023; pp. 391–411.
- Banfi, F.; Pontisso, M.; Paolillo, F.R.; Roascio, S.; Spallino, C.; Stanga, C. Interactive and Immersive Digital Representation for Virtual Museum: VR and AR for Semantic Enrichment of Museo Nazionale Romano, Antiquarium di Lucrezia Romana and Antiquarium di Villa Dei Quintili. ISPRS Int. J. Geo Inf. 2023, 12, 28.
- 19. Meng, Y.; Mok, P.Y.; Jin, X. Interactive virtual try-on clothing design systems. Comput. Aided Des. 2010, 42, 310–321.
- Santesteban, I.; Otaduy, M.A.; Casas, D. Learning-based animation of clothing for virtual try-on. Proc. Comput. Graph. Forum 2019, 38, 355–366.
- 21. Zhao, F.; Xie, Z.; Kampffmeyer, M.; Dong, H.; Han, S.; Zheng, T.; Zhang, T.; Liang, X. M3d-vton: A monocular-to-3d virtual try-on network. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 13239–13249.

- 22. Cheng, Z.Q.; Chen, Y.; Martin, R.R.; Wu, T.; Song, Z. Parametric modeling of 3D human body shape—A survey. Comput. Graph. 2018, 71, 88–100.
- 23. Chen, L.; Peng, S.; Zhou, X. Towards efficient and photorealistic 3d human reconstruction: A brief survey. Vis. Inform. 2021, 5, 11–19.
- 24. Correia, H.A.; Brito, J.H. 3D reconstruction of human bodies from single-view and multi-view images: A systematic review. Comput. Methods Programs Biomed. 2023, 239, 107620.
- 25. Tian, Y.; Zhang, H.; Liu, Y.; Wang, L. Recovering 3D human mesh from monocular images: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 2023, 1–25.
- Sun, M.; Yang, D.; Kou, D.; Jiang, Y.; Shan, W.; Yan, Z.; Zhang, L. Human 3D avatar modeling with implicit neural representation: A brief survey. In Proceedings of the 2022 14th International Conference on Signal Processing Systems (ICSPS), Zhenjiang, China, 18–20 November 2022; pp. 818–827.
- 27. Christoff, N. Modeling of 3D Human Body for Photorealistic Avatar Generation: A Review. In Proceedings of the iCEST, Ohrid, North Macedonia, 27–29 June 2019.
- 28. Zhou, S.; Fu, H.; Liu, L.; Cohen-Or, D.; Han, X. Parametric reshaping of human bodies in images. ACM Trans. Graph. 2010, 29, 1–10.
- 29. Omran, M.; Lassner, C.; Pons-Moll, G.; Gehler, P.; Schiele, B. Neural body fitting: Unifying deep learning and model based human pose and shape estimation. In Proceedings of the 2018 International Conference on 3D Vision (3DV), Verona, Italy, 5–8 September 2018; pp. 484–494.
- 30. Ionescu, C.; Papava, D.; Olaru, V.; Sminchisescu, C. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. IEEE Trans. Pattern Anal. Mach. Intell. 2013, 36, 1325–1339.
- Mehta, D.; Rhodin, H.; Casas, D.; Fua, P.; Sotnychenko, O.; Xu, W.; Theobalt, C. Monocular 3d human pose estimation in the wild using improved cnn supervision. In Proceedings of the 2017 International Conference on 3D Vision (3DV), Qingdao, China, 10–12 October 2017; pp. 506–516.
- 32. Varol, G.; Romero, J.; Martin, X.; Mahmood, N.; Black, M.J.; Laptev, I.; Schmid, C. Learning from synthetic humans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 109–117.
- 33. Loper, M.; Mahmood, N.; Romero, J.; Pons-Moll, G.; Black, M.J. SMPL: A skinned multi-person linear model. ACM Trans. Graph. 2015, 34, 1–16.
- Bogo, F.; Romero, J.; Pons-Moll, G.; Black, M.J. Dynamic FAUST: Registering human bodies in motion. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6233– 6242.
- Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Part V 13. Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
- 36. Johnson, S.; Everingham, M. Clustered pose and nonlinear appearance models for human pose estimation. Proc. BMVC 2010, 2, 5.
- Johnson, S.; Everingham, M. Learning effective human pose estimation from inaccurate annotation. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 1465–1472.
- Zhang, C.; Pujades, S.; Black, M.J.; Pons-Moll, G. Detailed, accurate, human shape estimation from clothed 3D scan sequences. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4191–4200.
- 39. Sigal, L.; Balan, A.O.; Black, M.J. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. Int. J. Comput. Vis. 2010, 87, 4–27.
- 40. Shu, T.; Ryoo, M.S.; Zhu, S.C. Learning social affordance for human-robot interaction. arXiv 2016, arXiv:1604.03692.
- Haque, A.; Peng, B.; Luo, Z.; Alahi, A.; Yeung, S.; Fei-Fei, L. Towards viewpoint invariant 3d human pose estimation. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Part I 14. Springer: Berlin/Heidelberg, Germany, 2016; pp. 160–177.
- 42. Bozhilov, I.; Tonchev, K.; Manolova, A.; Petkova, R. 3d human body models compression and decompression algorithm based on graph convolutional networks for holographic communication. In Proceedings of the 2022 25th International Symposium on Wireless Personal Multimedia Communications (WPMC), Herning, Denmark, 30 October–2 November 2022; pp. 532–537.

- Osman, A.A.; Bolkart, T.; Black, M.J. Star: Sparse trained articulated human body regressor. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Part VI 16. Springer: Berlin/Heidelberg, Germany, 2020; pp. 598–613.
- 44. Von Marcard, T.; Henschel, R.; Black, M.J.; Rosenhahn, B.; Pons-Moll, G. Recovering accurate 3d human pose in the wild using imus and a moving camera. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 601–617.
- 45. Liu, L.; Wang, K.; Yang, J. 3D Human Body Shape and Pose Estimation from Depth Image. In Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Shenzhen, China, 14–17 October 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 410–421.
- 46. Shi, M.; Aberman, K.; Aristidou, A.; Komura, T.; Lischinski, D.; Cohen-Or, D.; Chen, B. Motionet: 3d human motion reconstruction from monocular video with skeleton consistency. ACM Trans. Graph. 2020, 40, 1–15.
- Shen, J.; Cashman, T.J.; Ye, Q.; Hutton, T.; Sharp, T.; Bogo, F.; Fitzgibbon, A.; Shotton, J. The phong surface: Efficient 3D model fitting using lifted optimization. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Part I 16. Springer: Berlin/Heidelberg, Germany, 2020; pp. 687–703.

Retrieved from https://encyclopedia.pub/entry/history/show/117137