CASH: Comparative Anatomical Study of Humans

Prof. D. R. Kasat

Associate Professor, Dept. of Computer Engg, Sarvajanik College of Engineering & Technology, Surat, India

Abstract – The universal truth that none of us can deny is that 'I am a human, and no human in the universe can be of indifference in anatomy to me'. In addition, none of us can disagree with the fact that the face is an intrinsic part of human body and is crucial for human identity. The technological developments in engineering field of computer vision had led to automation of many tasks that the human visual system can do. In this paper, we present a survey cum analytic study of how the recent vision-based body modeling techniques differ for human body and the intrinsic part of the human body called face.

Keywords – body, comparative study, face, human, modeling, review

I. INTRODUCTION

One can always be fascinated by the universal fact of the human body i.e. humans resemble humans anatomically. All the humans share a long list of fundamental similarities with each other. All healthy and normal human bodies consist of the legs, the torso, the arms, the neck, and the head. The face found on the anterior surface of the head is crucial for human identity. Recent developments in computer vision technology focus on application of model-based approaches to deal with images of complex and variable structures such as faces and the human body. This paper presents a comparative study of how the most recent vision-based modeling techniques differ for human body and the intrinsic part of the human body called face.

To my knowledge, this is the first study attempting to assess the differences in modeling human body and face based on prior knowledge of human body anatomy.

II. FACE MODELING

Face analysis based on a model with prior knowledge of anatomy is a general standard with numerous applications. The list include face recognition, expression recognition, lip-reading, head pose estimation, and eye gaze estimation. To obtain the plausible results in the application of interest, we can classify the prior knowledge of expected structure shapes, spatial relationships and their appearance to prefer Generative Models or Deformable Models [1]. Generative Models are preferred in case we need to formulate image analysis as a matching problem i.e. given any face image; particular structures can be located and tracked by adjusting the model's parameters to generate the desired face image without any artifact. Deformable Models are preferred when we need to deal with variability in faces in a general and specific manner both i.e. if we are dealing with human face then the deformable model is capable of generating any other general face of human class only in a plausible form. We thus conclude that *a face model is a compact representation to describe a wide variety of facial images.* Depending upon the application consider a set of varied parameters to form the compact representation. Fig 1 illustrates the position of head accounting for model fitting and the identification of facial features like eye corners, nose tip, lip corners etc for face detection and tracking.The construction methodology of a face model broadly categorizes the model as either 2D or 3D. The model constructed from a collection of training data, either 2D images or 3D range scans. In the next step, the face model fit to the input image(s) and the model parameters are worn in whatever the application is. The most well known face models are 2D Active Appearance Models (AAMs) [2] and 3D Morphable Models (3DMMs) [3]. Both the models include a combination of a linear shape model and a linear appearance (texture) model. Table 1 exhibits a comparative analysis of why 3D linear face models are more preferable over 2D linear face models.

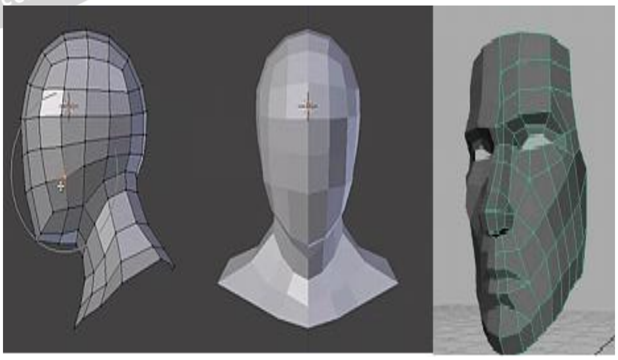


Figure1: Head Pose and face Model

IIA) COMPARISON OF RELATED EXISTING WORK ON FACE MODELING

In 2004, V. Zanella et al. [4] proposed method to perform the morphing of face images in frontal view with uniform illumination automatically, using a generic model of a face and evolution strategies to find the features in both face images. The authors used a model of 73 points based on a

 DOI : 10.18231/2454-9150.2018.1010

simple parameterized face model using information about the geometrical relationship among the elements of the face. In this work, automatic fitting of this model to source image and target image has been proposed using genetic operation. Once the model has been adjusted to

Table 1 : Comparative Analysis of 2D/3D linear face model

	2D Model	3D Model
Shape component	The shape component is 2D	The shape component is 3D
Model fitting	The fitting of 2D face model is more prone to local minima and so is less robust.	More robust and takes fewer iterations to converge
Representation power	May generate some "physically unrealizable" model instances and may require up to 6 times more parameters compared to 3D model to represent the same phenomenon as the 3D model.	Representation power is good because the parameterization is natural compared to 2D model and also require comparatively less number of parameters
Construction	Less dense compared to 3D model	More dense i.e. contains more number of triangles in a mesh compared to 2D model
	2D model computed from the 3D model is better able to model faces at high degrees of pose variation than the original 2D model was able to.	We can construct 3D models from 2D data but quality is poor compared to 3D models constructed from 3D range scans.

the images, it performed image deformation, or warping, by mapping each feature in the source image to its corresponding feature in the target image.The described method works only for frontal view face morphing; otherwise this face morphing technique tends to generate blurry intermediate frames when the two input faces differ significantly.

In 2012, F. Yang et al. [5] proposed a new fully automatic face morphing approach that dealt explicitly with large pose and expression variations. The system requires only the two end images as input and it first extract facial landmarks from each input image, then project them on a subspace learned from an external face shape dataset to recover the 3D face geometry. The authors factor this geometry into the pose and the expression of the input face. Then, it linearly interpolates

both the intrinsic and external parameters of the two input faces, and generates a series of interpolated 3D face models. The results shared by the authors claim that it takes about 10 minutes to create eight intermediate frames for face regions about 200 by 200 pixels, on an Intel CPU of 2.40 GHz. However, the system depends on ASM for localizing facial components and thus can fail under large viewpoint variations or occlusions. In addition the system use optical flow to capture subtle geometric changes and might generate poor results when there are large differences in facial appearance, skin tones or illumination between two inputs.

Certain methods also allowed for automatic face replacement of people in single image [7,8]. For example, in 2004 the method by V. Blanz et al. [7] fits a Morphable model to faces in both the source and target images and renders the source face with the parameters assessed from the target image. Finally, it replaces the target face with source face in the target image. Morphable model[3] is built from a statistical analysis of human faces, obtained from a large database of 3D scans, which can be morphed by adjusting parameters. It can estimate the 3D shape of a human face, its orientation in the space, and illumination conditions in the scene. Thus the reconstructed face extracted from 2D image can be manipulated in 3D.

In 2008, D. Bitouk et al. [7] described another system for automatic face swapping using a large database of faces. Though it is hard for user's to find a candidate face to match the target face in appearance and pose from their images, the system allowed de-identification automatically by selecting candidate face images from a large face library that is similar to the target face in appearance and pose. Lastly, it replaces the target candidate with selected candidate from the library image using image based method.

In 2013, S. Gao et al. [9] introduced simple 3D face model, which is known as Morphable guidelines. They proposed a system which allows morphing specific part of face like nose, mouth, cheek etc. in single image. This Morphable guideline is a 3D model structured similar to the ball and plane method. This model consists of simple curves like a circle, line etc. which is controlled by the 3D Vertices. Individual shape can be changed by changing the parameter. In this paper they have applied this model to reshape human face parameter such like nose, mouth, cheek etc. But model fitting process to the human face in image is manual. In 2015, D. Kasat et al [8] applied the Morphable guidelines of S. Gao et al [9] with an improvement of making the model fitting automatic in real time for 3D face morphing.

In 2009, Y. T. Cheng et al. [10] proposed the system that replaces the target subject face in the target video with the source subject face, under similar pose, expression, and

 DOI : 10.18231/2454-9150.2018.1010

illumination. This approach is based on 3D morph-able model [3] and an expression model database to deal with expressions and the input information of the source subject face is reduced to one to two images. The system takes a target video and one source image as input, and the output is the video with the target subject face replaced with the source subject face. Given the source image, it reconstructs the 3D model of the source subject face using 3D Morphable model [4]. The 3D face synthesizer derives a Morphable face to fit the input image, and map the texture from the image to the derived 3D face model. A face alignment algorithm is applied to the target video to detect the detailed facial features and outlines of the target subject face. A pose estimator exploits the face alignment results to estimate the head pose parameters of the target subject face. Here method employs a 3D face expression database to clone the expressions to the source face model. To fit the expressions to the target video, Y. T. Cheng et al. [10] proposed an algorithm to extract the expression parameters. In some videos, directly rendering the source subject face model onto the target frame results in illumination inconsistency. A relighting algorithm relights the rendered source subject face for illumination consistency. Finally, it seamlessly composites the rendered source model with the target frame using Poisson equation. The output is a video with the target face replaced by the source face, with similar pose, expression, and lighting.

In 2011, K. Dale, et al. [11] proposed the method which allows replacing facial performances in video. It also provides face replacement in target video from source video. The system tracks both the faces in source and target video using multilinear model [1]. Using this tracked 3D geometry, source face is warped to target face in every frame of video. It is sometimes important that the timing of the facial performance matches exactly in the source and target video; this is done by retiming algorithm. After tracking and retiming, it blends the source performance in the target video to produce the final result. The authors computed optimal seam through the video volume that maintains temporal consistency in the final composite.

II B) Review of Face Modeling

These automated face warping/morphing techniques each have their own respective weaknesses as depicted in Table 2. Face tracking AAM packages, AdaBoost algorithm are weak when trying to match a wide range of facing angles, and morphable models are overly complex, time consuming and inefficient for nonprofessional use and minor changes. Therefore, these packages are not suitable for real time face warping/morphing methods. Microsoft face tracking SDK packages for Kinect depth sensor provide efficient face tracking in real time. Table 2 exhibits a quick performance analysis with the pros and cons of some major face morphing techniques.

III. Human Body Modeling

The increasing demands of innovative applications in games, animations, entertainment industry, virtual reality etc have lead to a continuous and accelerated growth of interest in human body modeling. The challenge is to produce a Human Body Model that is able to follow the movements of a real person faithfully.

III A) 3D model for Human Body and Pose

Modeling a human body first implies the version of an articulated structure, in order to represent the branching parts of human body. Secondly, the definition derives a mathematical model used to administrate the activities of that articulated structure. Several computational methods are developed in the literature to obtain 3D models of the human body exhibiting the articulated representations and mathematical formalisms related to both the structure and movements of a human body.

Generally, a human body model can be represented as a binary tree of non-mechanical body parts, interconnected to one another by joints starting with the torso as root body part. For the purpose of representing these non mechanical body parts several authors in the literature have used either of Links, sticks , polyhedron, generalized cylinders or super quadrics . A joint interconnects two rigid body parts by means of rotational motions about the axes. The number of independent rotation parameters defines the degrees of freedom (DOF) associated with a given joint. Fig. 2(a) presents an illustration of an articulated structure and 2(b) presents the human body in some varied poses of the many possible.

The geometric representation is then integrated with a mathematical model of human motion, to make the human body model look alive. When we wish to apply the model to a human body in real time/offline video, we need to track the human body in each frame to refit the model, so that an application such as human body tracking systems may be improved. There is a wide variety of ways to model mathematically articulated from a kinematics and dynamics point of view. In addition, the human body comes in all shapes and sizes, from ballet dancers to sumo wrestlers. Many attempts were found in the literature to measure and categorize the scope of human body variation. Graphics applications often require a complete surface model for rendering and animation. Obtaining a complete model of a particular person is often difficult or impossible. Even when the person is constrained to remain motionless inside of a Cyber ware full body scanner, incomplete surface data is obtain due to occlusions. When the task is to obtain a 3D sequence of the person in motion, the situation can be even more difficult. The SCAPE (Shape Completion and Animation of People) model [14], which represents both articulated and non-rigid deformations of the human body, proved to be a data-

 DOI : 10.18231/2454-9150.2018.1010

driven method required for building a human shape model that spans variation in both subject shape and pose. SCAPE consists of two components: P*ose deformation model and Shape deformation model.* The *pose deformation model* captures how the body shape of a person varies as a function of their pose. For example, this can model the bulging of a bicep or calf muscle as the elbow or knee joint varies. The second component is a *shape deformation model, which* captures the variability in body shape across people using a low-dimensional linear representation.

Table 2 : Performance Analysis of Different Automatic face morphing Methods

Face morphing Techniques	Ref. Works	Pros	Cons
Face morphing in images	An Approach to Automatic Morphing of Face Images in Frontal View[5]	Comparatively less time consuming	Works only if face images in frontal view with uniform illumination
	Face morphing using 3D aware appearance optimization [6]	Deals explicitly with large pose and expression variations in face images	Time consuming
	Exchanging Faces In Images [7]	It can be used to replace face; Also works for non-frontal view	Does not give pleasant result if large illumination difference
	Face-Swapping: Automatically Replacing Faces In Photographs[8]	It works even large illumination difference in both images	Does not work in case of any Occlusion
	Morphable Guidelines For The Human Head[9]	Allow to change face parameter like nose, cheek, jaw, chin, mouth and eyes etc.	Model fitting process is manual
Face morphing in video	3D Model Based Face Replacement in Video[10]	Here source is reduced to single image; It takes care of facial expression and pose of target face	Time consuming process because of 3d model based method; The tolerance to pose variance is still limited by the robustness of face alignment algorithm
	Video Face Replacement [11]	It gives plausible results;This approach is fully automatic with less user interference; It takes care about facial expression	Tracking is based on optical flow, which requires that the lighting change slowly over face; Tracking often degrades beyond the range of poses outside 45^0 from frontal; Lighting must also be similar between source and target.
Face morphing in real time video	Real-time view morphing of video streams [20]	Morphing effect in real time; It can be used to create virtual environment	Require higher configuration of hardware
	Entertaining video warping [12]	It morphs face area only instead of complete frame	It provides limited warping function like face bulging, face shrinking
	Real time face morphing [8]	Fully automated system for morphing specific face features in real time	It requires 3D data for its morphable model, which makes it non-compatible with the smart phone containing 2D camera.

The authors could form these two models from examples by capturing a rich and natural range of body shapes, and provide a more detailed 3D triangulated mesh model of the human body than previous models used in video-based pose estimation. The model has many advantages over previous deformable body models used in computer vision. In particular, since it is learned from a database of human shapes it captures the correlations between the sizes and shapes of different body parts. It also captures a wide range of human forms and shape deformations due to pose. Modeling how the shape varies with pose reduces problems of other approaches associated with modeling the body shape at the joints between parts. Jain et al. [16] employed a variant of the SCAPE [14] model to represent the pose and the body proportions of an actor in 3D. One of the major difficulties with the SCAPE model is that they rely on different means for encoding shape and pose. Pose (and the effects thereof, e.g., muscle bulging) are stored with the help of an underlying skeleton while body

 DOI : 10.18231/2454-9150.2018.1010

shape is encoded using variation methods or envelop skinning. During animation, the outcomes of the two methods have to be combined in an additional step. Hasler et al. [17]

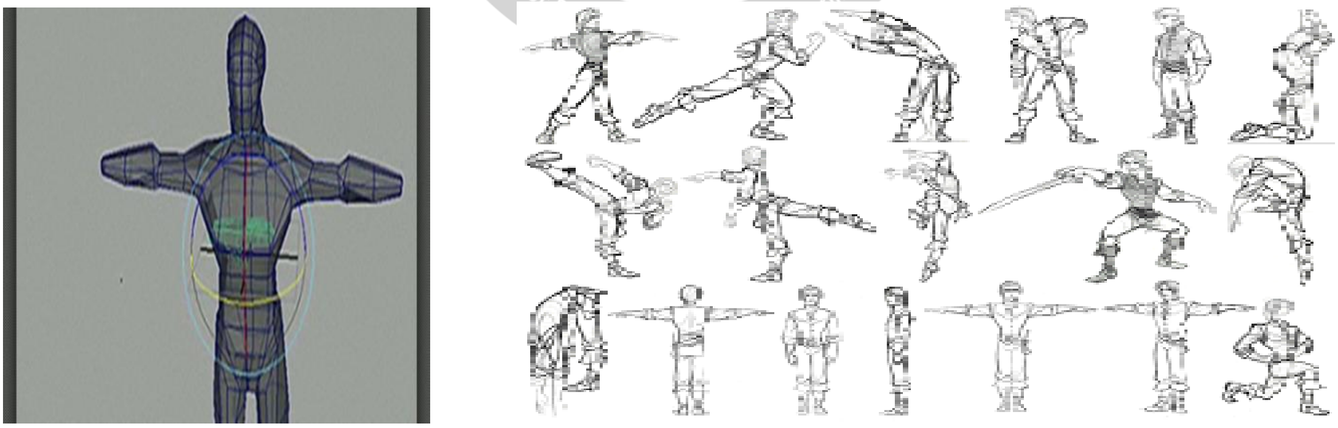
present a different statistical body model by simultaneously modeling both body shapes, pose from a database of 3D laser scans consisting of over 100 people of varying age and gender, and in various poses. The system of Ritcher et al. [13], i.e. a system for real-time deformation of the shape and appearance of people who are standing in front of a depth+RGB camera, such as the Microsoft Kinect relies on a mathematical model for human body shape. The variations across several men and women are modeled and the model is built on top of the model developed by Hasler et al. [17].

III B) Review of Human Body Modeling

Human body modeling, since its inception has remained an application-oriented research field, i.e., the need had dictated the kind of systems that has been developed. Due to widespread interest, in the computer vision applications like games, virtual reality , augmented reality etc. there has been an abundance of work on human body modeling during the previous years. The major constraint involved is the computational complexity required to produce realistic models with natural behaviors. This requirement of automatic generation of a realistic and fully configurable human body model is still an open problem with issues pertaining to topics on unconstrained image segmentation, limitations in tracking, development of models including prior knowledge, modeling of multiple person environments, real-time performance. In addition, the major class of users of such applications is young generation, thus giving rise to a new milestone in human body modeling. The new milestone is speed i.e. the reduction of processing time, which is highly dependent on two factors: on the one hand, computational complexity and, on the other hand, current technology. On the contrary, improvements in the current technology have become commonplace (e.g., reduction in acquisition and processing times, increase in the memory size). Therefore, computationally demanding algorithms are expected to have a good performance with the next technologies. The latter gives rise to a promising future for human body modeling applications and extension non-rigid object modeling, in general. The area of human body modeling is growing considerably fast. Therefore, it is expected that most of the current drawbacks will be solved efficiently through the next years.

IV. CONCLUSION

After comparative performance analysis of recent research trends and models of human face and body one can easily draw the conclusion that most recent vision-based body modeling techniques differ for human body and the intrinsic part of the human body called face. Human faces are highly non-rigid and are able to perform large 3D shape deformations under expression and pose variations. Moreover, human perception is sensitive to even a small amount of artifacts in faces. A comparative analysis of important parameters considered in the design and selection of model based on prior knowledge of human body anatomy allows us to understand the differences in modeling human body and face due to their anatomy



(a) (b)

Figure 2: a) Articulated structure of Human Body Model b) Different body Postures

- **Nature of Model preferred:** The nature of model that is preferred for different face applications analyzed is statistical/parametric model based on land-marking techniques or feature extraction whereas for body modeling the articulated model is more important depending upon the anatomy knowledge. The number of joints and the degree of freedom per joint depends upon the articulated structure under consideration. The articulated model is of no use for face modeling, as there are no joints in a human face. For a face model, the nature of

 DOI : 10.18231/2454-9150.2018.1010

ISSN : 2454-9150 Vol-04, Issue-07, Oct 2018

model can be 2D or 3D but for a body model, only 3D models are preferred.

- **Pose estimation and tracking:** The pose of a face is estimated through the position of its rigid part called head and for the tracking purpose the position of landmark, points/features on the face are detected and tracked. The pose of a human body is estimated by governing the human model fitted through a kinematic chain of linear motion equations for the joints of the articulated structure.

- **Real time performance:** Many of the recent face models work efficiently in real time with the help of depth sensors or multi-view cameras but body models are still striving hard to produce plausible performance in real time.

REFERENCES

[1] T.F. Cootes and C.J.Taylor, "Statistical Models of Appearance for Computer Vision ",Imaging Science and Biomedical Engineering,University of Manchester U.K. March 8, 2004

[2] T.Cootes, G. Edwards, and C.Taylor, "Active appearance models", IEEE Trans. Pattern Analysis Machine Intelligence, vol. 23,no. 6, pp. 681- 685,June 2001.

[3] Blanz and T.Vetter, "A Morphable Model for the Synthesis of 3D Faces", in Proc. 26th ACM Ann. Conf. Comp. Graph.& Int. Tech., pp. 187-194, Aug. 8-13, 1999.

[4] Zanella and O. Fuentes, "An Approach to Automatic Morphing of Face Images in Frontal View", Springer LNCS, MICAI vol. 2972, pp. 679-687, Jan 2004.

[5] F. Yang, E. Shechtman, J. Wang, L. Bourdev and D. Metaxas, "Face morphing using 3D-aware appearance optimization", in Proc. Graphics Interface, pp. 93-99, 2012.

[6] V. Blanz, K. Scherbaum, T. Vetter, and H. Seidel, "Exchanging Faces In Images", Computer Graphics Forum, vol.23, no. 3, pp. 669-676, Sept. 2004.

[7] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. Nayar, "Face Swapping: Automatically Replacing Faces in Photographs", ACM Trans. on Graphics, vol. 27, no. 3, pp.39,Aug 2008.

[8] D. Kasat , S. Jain and V. Thakare , "Real Time Face Morphing " in Proc. IEEE Int Conf CINE, Jan 12-13,2015

[9] S. Gao, C. Werner and A. Gooch, "Morphable Guidelines For The Human Head", in Proc. of the Symposium On Computational Aesthetics, pp. 21-28, July 19-20, 2013.

[10] Y. Cheng, V. Tzeng, Y. Liang, C.Wang, B.Chen, Y.Chuang and M. Ouhyoung, "3DMorphable Model Based Face Replacement in Video", in Proc. ACM SIGGRAPH, Aug. 3-7, 2009.

[11] K. Dale, K. Sunkavalli, M. Johnson, D. Vlasic, W. Matusik, and H. Pfister, "Video Face Replacement", ACM Trans. Graphics, vol. 30, no. 6, pp. 130:1-130:9,2011

[12] Y.Yang,Y. Zhu, C. Liu, C. Song and Q. Peng, "Entertaining video warping", in Proc.,11th IEEE Int. Conf. CAD/Graphics, pp. 174-177, Aug. 2009.

[13] M. Richter, K. Varanasi, N. Hasler and C. Theobalt, "Real-time Reshaping Of Humans", in Proc. IEEE Second Int. Conf. 3DIMPVT,pp. 340-347, 13-15 Oct. 2012

[14] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers and J. Davis, "Scape: shape completion and animation of people", ACM Trans. Graph., vol. 24, no. 3, pp. 408–416, July 2005.

[15] S. Zhou, H. Fu, L. Liu, D. Cohen-Or, and X. Han, "Parametric reshaping of human bodies in images", in Proc.ACM SIGGRAPH, pp. 126:1–126:10, July 25-29, 2010.

A. Jain, T. Thormahlen, H.Seidel, and C. Theobalt, "Movie reshape: Tracking and reshaping of humans in videos", ACM Trans. Graph., vol. 29, no. 6, pp. 149-157, Dec. 2010.

[16] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H. Seidel, "A statistical model of human pose and body shape", in Proc. EUROGRAPHICS ,vol. 28,no. 2, Mar. 2009.

[17] S. Schaefer, T. Mcphail and J. Warren, "Image deformation using moving least squares", ACM Trans. Graphics, vol. 25, no. 3, pp. 533–540, 2006.

[18] Y. Zhu and S.Gortler "3D deformation using moving least squares", Harvard Univ., Dept. of Comp. Sci., Technical Report: TR-10-07, May 2007.

[19] K. Timm, "Real-time view morphing of video streams",PhD Dissertation, Dept. of Comp. Sci., University of Illinois at Chicago,GE Medical Systems, 2003.

[20] Kinect for Windows SDK – MSDN Docs http://msdn.microsoft.com/en-us/library/hh855347.aspx, April 2013.

[21] Linford and P. Blanchard, "Aesthetic imaging system", US Patent 5 825 941, 1998.

[22] D. Kasat, S. Jain and V. Thakare, "A Survey of Face Morphing Techniques", IJCA RTINFOSEC pp.14-18, Feb 2014.

 DOI : 10.18231/2454-9150.2018.1010