Acting the Part: The Role of Gesture on Avatar Identity

Andrew Feng*[1], Gale Lucas†[1], Stacy Marsella‡[2], Evan Suma§[1], Chung-Cheng Chiu¶[1], Dan Casas‖[1], and Ari Shapiro**[1]

[1]Institute for Creative Technologies, University of Southern California
[2]Northeastern University

Figure 1: *We present a system whereby a human subject can be efficiently captured and simulated as a 3D avatar. We present a study where we ask subjects to judge which performance better resembles the person depicted in the avatar.*

Abstract

Recent advances in scanning technology have enabled the widespread capture of 3D character models based on human subjects. However, in order to generate a recognizable 3D avatar, the movement and behavior of the human subject should be captured and replicated as well. We present a method of generating a 3D model from a scan, as well as a method to incorporate a subjects style of gesturing into a 3D character. We present a study which shows that 3D characters that used the gestural style as their original human subjects were more recognizable as the original subject than those that don't.

CR Categories: I.3.2 [Computing Methodologies]: Computer Graphics—Methodology and Techniques; I.3.7 [Computing Methodologies]: Computer Graphics—Three-Dimensional Graphics and Realism; I.3.8 [Computing Methodologies]: Computer Graphics—Applications

*feng@ict.usc.edu
†lucas@ict.usc.edu
‡marsella@neu.edu
§suma@ict.usc.edu
¶chiu@ict.usc.edu
‖casas@ict.usc.edu
**shapiro@ict.usc.edu

Keywords: avatar, gesture, 3D, animation, simulation

1 Introduction

Recent advances in scanning technology have enabled the rapid creation of 3D characters from human subjects using image, video and depth sensing cameras. The fidelity of such static 3D models can vary based on the scanning equipment, lighting, method and the manual processes used to generate and refine the model. However, the efficiency of the capture process and the detail that can be acquired from such a process makes it appealling to 3D content producers. By contrast, traditional 3D construction of a static character model through manual means can take days, weeks, or longer in order to generate approximations or stylizations of a human subject. Such efficiency in scanning and 3D model creation opens the possibility of efficiently creating large numbers of 3D characters with appearances as varied as the population used as subjects for the capture. For example, a large crowd scene can be populated with a population of 3D characters generated from scanned data.

Another use of 3D characters models is to act as a representation of the human capture subject in a simulation, i.e as an avatar. In such cases, it is important for the human simulation user to be able to recognize the 3D characters in the simulation. As an example,

Motion in Games 2014, November 06 – 08, 2014, Los Angeles, California.

http://dx.doi.org/10.1145/2668064.2668102

a military training simulation might require a user to run practice drills with virtual squad members that look like the real squad members. Alternatively, a training simulation might require the presence of coworkers to be part of the 3D training environment. All of these uses require that a 3D character be a recognizable representation of the original human subject. There is also a growing body of research related to the the psychological effects of seeing yourself within a simulation [Bailenson 2012; Fox and Bailenson 2009; Fox and Bailenson 2010; Lee et al. 2010] and examining a person's preferences for gestural style on agents as it relates to their own gestural style [Luo et al. 2013].

Avatars can be identified as representations of their human counterparts through many modalities, such as through visual inspection of their static or moving 3D image, or through recognition of their voices. However, under prolonged exposure to a 3D avatar in a simulation, we examine whether the behavior and movement of an avatar also contributes to the recognition or the immersion into a simulation that includes such an avatar. Early work has shown that people can recognize others based on their movements [Cutting and Kozlowski 1977], are more sensitive to their friend's movements that to stranger's [Loula et al. 2005]. In this paper, we show a framework by which we can efficiently capture both the visual appearance of a human subject and the gestural style. We perform a study that shows that observers found the 3D avatar performance that used the same gestures as the original human subject were rated as 'more like' the original human subject by groups that knew that subject, than 3D avatars that performed using another human subject's gestures.

2 Related Work

In this paper, we describe a framework for the efficient capture and synthesis of a 3D avatar from a human subject. Our framework consists of stages of 3D model acquisition, 3D character rigging, motion capture and gesture synthesis.

2.1 Avatar Capture and Creation

Creating a virtual character from a particular subject is not a trivial task and usually requires extensive work from a 3D artist to model, rig, and animate the virtual character.

The first step of avatar creation requires reconstruction of a 3D model from either a set of images or depth range scans. With the availability of low-cost 3D cameras (Kinect and Primesense), many inexpensive solutions for 3D human shape acquisition have been proposed. The work by [Tong et al. 2012] employs three Kinect devices and a turntable. Multiple shots are taken and all frames are registered using the Embedded Deformation Model [Sumner et al. 2007]. The work done in [Zeng et al. 2013] utilizes two Kinect sensors in front of the self-turning subject. The subject stops at several key poses and the captured frame is used to update the online model.

More recently, solutions which utilize only a single 3D sensor have been proposed, and this allows for home-based scanning and applications. The work in [Wang et al. 2012] asks the subject to turn in front of a fixed 3D sensor and 4 key poses are uniformly sampled to perform shape reconstruction. To improve the resolution, KinectAvatar [Cui et al. 2013] considers color constraints among consecutive frames for super-resolution. More recently, the work in [Li et al. 2013] asks the subject to come closer and obtains a super-resolution scan at each of 8 key poses.

The second step is to create an animated virtual character from the scanned 3D human model. A 3D model needs to be rigged with a skeleton hierarchy and appropriate skinning weights. Traditionally, this process needs to be done manually and is time consuming even for an experienced animator. An automatic skinning method is proposed in [Baran and Popović 2007] to reduce the manual efforts of rigging a 3D model. The method produces reasonable results but requires a connected and watertight mesh to work. The method proposed by [Bharaj et al. 2011] complements the previous work by automatically skinning a multi-component mesh. It works by detecting the boundaries between disconnected components to find potential joints. Such a method is suitable for rigging the mechanical characters that consists of many components. Other rigging algorithms can include manual annotation to identify important structures such as wrists, knees and neck [Mix 2013].

In the last few years, video-based methods have enabled the capture and reconstruction of human motions as a sequence of 3D models [Starck and Hilton 2007]. Such methods, which are capable of reproducing surface and appearance details over time, have been used to synthesize animations by the combination of a set of mesh sequences [Casas et al. 2014]. This results in a novel motion that preserves both the captured appearance and actor style, without the need of a rigging step. However, current approaches only demonstrate successful results for basic locomotion motions such as walk, jog and jump. The complexity of the gesture motions needed in this work would still require the video-based 3D models to be rigged.

2.2 Gesture Synthesis

In this work, we seek to reproduce a human subject's gestural style onto its 3D avatar. This can be achieved through traditional motion capture and transfer means via retargeting [Gleicher 1998]. However, a general synthesis technique is needed in order to generate motions for unrecorded performances. For the work in this paper, it is also important that synthesized motions are faithful to the styles of the scanned subject. Previous works exists for synthesizing novel gesture motions from speech input. [Stone et al. 2004] used mocap segments that correspond to pre-recorded phrases and rearrange them to match the new sentences. [Levine et al. 2009] uses prosody-based features extracted from audio to train hidden Markov Models to generate appropriate gesture. Motion is realized using segmented motion captured gesture. [Levine et al. 2010] performed realtime generation of gestures including word spotting (you, me). Since the gesture selection is based on similarity instead of the semantics, the resulting gestures are not likely to match the semantic content of the speech.

The work in [Neff et al. 2008] annotates and profiles the gesture styles of a particular speaker from video recording. The gesture motions are then synthesized procedurally using heuristic and inverse kinematics from speech input. The method can produce highly correlated gestures to that particular speaker. However, due to the fact that the synthesis algorithm is procedural, the resulting motions usually are not as natural as the motions generated by example based methods.

More recently, the work in [Marsella et al. 2013] uses a rule-based method to infer suitable gesture labels from speech input. A corresponding motion are then synthesized by concatenating gesture segments from the database. Gesture hold and co-articulation are handled by blending adjacent segments and adding post-stroke holds. Although the method produces promising results, it requires significant setup of appropriate gesture motions for the gesture database.

3 3D Avatar Synthesis

We used the method proposed in [Shapiro et al. 2014] to obtain an articulated 3D character from the test subject. The process requires the subject turns in front of the Kinect, while remaining static for four key poses. For each key pose, a super-resolution scan is created from its corresponding view. The resulting four scans are then aligned and merged through both rigid and articulated alignment to register all scans. The final static geometry is then produced via Poisson mesh reconstruction. The texture information is also inferred from the four individual scans via Poisson texture blending. The body scanning capture takes approximately 4 minutes. Due to the lack of facial detail obtained from the body scan, a separate face scan using [Hernandez et al. 2012] is used to produce a 3D face model, which is then stitched onto the avatar body, replacing that area generated during the body scan (Figure 2). Since the body scan is also not able to distinguish individual fingers, 3D hands from a female 3D model where used to replace the hands of the avatar from the body scan. Note that our motion capture process does not preserve finger motion. The facial scan takes only a few minutes to scan and generate, but the changes to the geometry and textures of the 3D model require manual artist intervention.

Figure 2: *High fidelity face scan that replaces the face from the avatar body scan.*

The scanned character model requires proper rigging structure in order to produce movements. The method automatically builds and adapts a skeleton to the 3D scanned character. Thus, it can transfer the gesture motions onto the scanned character via motion retargeting. The auto-rigging method is a variation to the one proposed in [Baran and Popović 2007] by building a distance field from the mesh and using the approximate medial surface to extract the skeletal graph. The difference is that instead of requiring a watertight mesh, the method uses voxel representation to build distance fields. Thus the skeleton extraction process would be more robust on meshes with topological artifacts and also make the processing time independent of the mesh resolution.

Once the skinned avatar is created, the captured gesture motions can then be applied to the character via motion retargeting. We use the on-line retargeting method proposed in [Feng et al. 2014] to transfer the motion. The retargeting is done by converting the joint angles encoded in a motion from a source skeleton associated with the gesture motions to the target skeleton from the scanned avatar. To encode the discrepancy between two skeletons, each bone segment in target skeleton is rotated to match the global direction of that segment in source skeleton at default pose. Once the default pose is matched, we address the discrepancy between their local frames by adding suitable pre-rotation and post-rotation at each joint. By combining these alignment rotations and pre/post-rotations, a suitable new offset rotation can be found to compensate for the difference between two skeletons. Thus at run-time, this offset rotation can be used to convert joint angles from source skeleton onto target skeleton.

4 Gesture Motion Synthesis

We believe that the movements of a person, in a addition to his appearance, play an important part on the effectiveness and recognizability of his scanned avatar. Thus an important stage of our pipeline is to recreate the movement styles of the scanned subject. This requires a method that takes some example performance of a person, and then synthesize new motions that reproduce similar movement styles.

We used the method from [Chiu and Marsella 2014] to generate the novel gesture motions from a new speech utterance. Given a motion capture sequence of gesture animations, the method utilizes gaussian process latent variable models (GPLVMs) to learn a low-dimensional embedding that encodes the gesture motions. To generate new gesture motions from speech input, the method first maps the speech into gesture labels from a speech-annotation mapping learned using conditional random field (CRF). Then the synthesis process selects the corresponding gesture segments that match the gesture labels from low-dimensional space and concatenate them together to form a gesture motion for the speech. Finally, the discontinuity between gesture segments are interpolated by inferring a smooth trajectory inside the low dimensional space.

The advantage of the method is that it separates the gesture label inference from animation synthesis, and thus the two learning tasks can be solved individually with different datasets. This allows us to easily use the gesture motion from a particular individual to as input motions to specifically encode his gesture styles. Since the method does not require the input motion to be a structured set of gestures, the raw captured gesture motions can be readily used as input to build the GPLVM without further processing. Thus the synthesis method fits well with our rapid avatar capture pipeline to further obtain and inject the styles of movements onto the captured avatars.

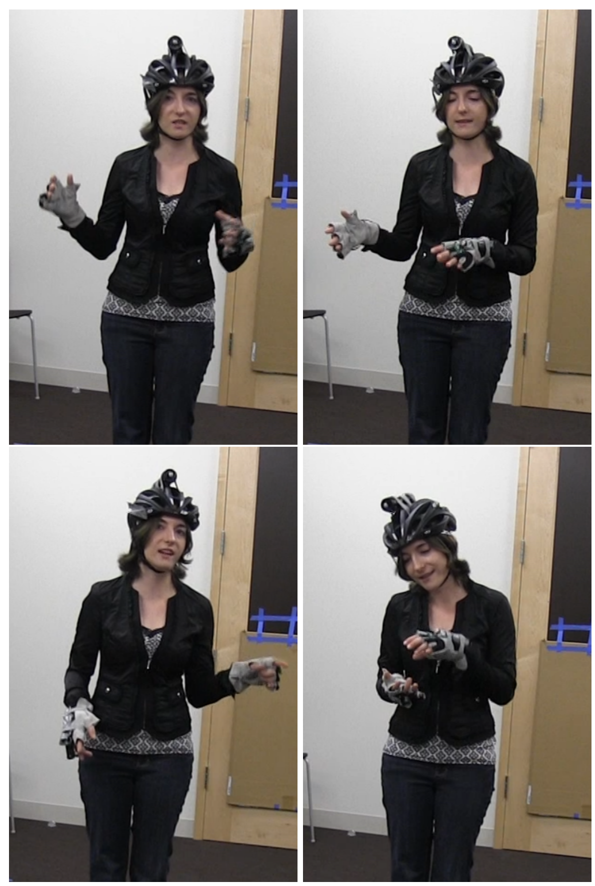

Figure 3: *Performance capture of the actress using markerless motion capture. Additional head and wrist sensors are added to supplement the*

To produce the input motion data of a specific individual for GPLVMs, we need to capture the gesture performance from that subject. A low cost motion capture system from iPiSoft [iPiSoft] is used to rapidly capture the gesture motions of our subjects. The system utilizes a single depth sensor such as Microsoft Kinect to capture a point cloud sequence and the character animation is inferred by fitting a skinned actor to the point clouds. Head and wrist rotations are also recorded by attaching inertial sensors on subjects' hands and head. Thus the resulting motions contain the full-body movements as well as head nods and wrist rotations. We ask the subjects to speak naturally in front of the capture system for about 5 minutes to record their gesture performance while speaking. Figure 3 shows some example snapshots of the actress gesturing during motion capture. The captured motion is then used as the input for gesture synthesis.

5 Evaluation

We conducted a test to examine whether participants identify a 3D avatar as being more like the actor who inspired the avatar if it gestures like that actor rather than a different actor.

5.1 Evaluation study

For this study, we modeled avatars after two different amateur actresses to test the possibility that people would identify their avatars as being more like them to the extent that that avatar's gestures were patterned after that actress's own gestures (rather than the other actress's gestures). All participants in the study knew one or both of the actresses. Specifically, 29 participants who knew amateur actress A and 41 participants who knew amateur actress B volunteered to complete the study for no compensation.

Participants in this study who knew actress A viewed two short videos (about a minute each) of actress A's avatar: one with gestures recorded from actress A, and another one with gestures recorded from actress B (Figure 4). After the first video, participants were asked how much is the avatar like this actress, to which they responded on a scale from 1 (not very much like this actress) to 7 (very much like this actress). After the second video, participants answered this same question again, and then were also asked to choose which avatar was more like the actress.

Likewise, those in this study who knew actress B viewed two videos of her avatar: one with gestures recorded from actress A, and another one with gestures recorded from actress B. They answered the same set of questions, but in reference to actress B.

5.2 Results

We conducted a Wilcoxon [Wilcoxon 1945] signed ranks test to determine whether participants reported that actress A's avatar was more like actress A when it gestured like actress A or like actress B. This analysis confirmed that participants who new actress A reported that her avatar seemed 'more like her' when it gestured like actress A (Median = 5.00, IQR = 4.00-6.00) than when it gestured like actress B (Median = 3.00, IQR = 1.00-5.00, Z = -3.49, p ¡ .001). A χ^2 goodness of fit test also revealed that, when given a forced-choice, these participants were also marginally more likely than chance to choose the avatar that had matching gestures as being like actress A than to choose the avatar with actress B's gestures (19 vs 9, χ^2 (1) = 3.57, p = .059).

Among participants who knew actress B, we find comparable results for her avatar. Participants reported that the avatar was 'more like her' when it gestured like actress B (Median = 5.00, IQR = 4.00-6.00) than when it gestured like actress A (Median = 3.00, IQR = 1.50-4.50, Z = -3.13, p = .002).

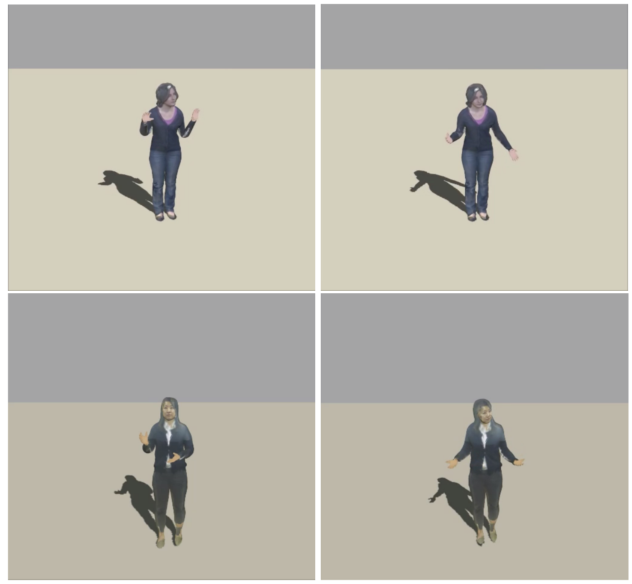

Figure 4: *3D avatars captured with gestures synthesized by our system. Each avatar performance is synthesized with the original subject's gestures and with the other's gestures.*

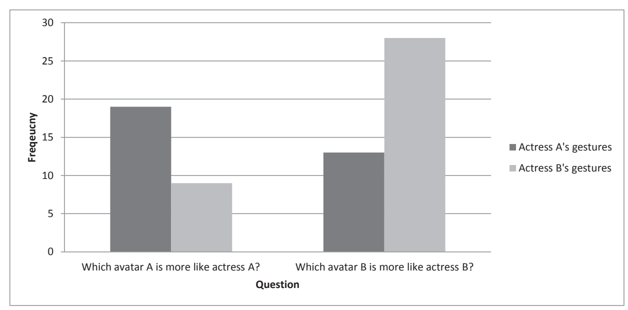

Figure 5: *Study results showing whether participants reported that actress A's avatar was more like actress A when it gestured like actress A or like actress B, and vice-versa.*

Further evidence also confirms that participants should be able to identify an avatar as being more like the actor who inspired the avatar if it gestures like the actor rather than a different actor.

6 Discussion

We present a framework for capturing and simulating a 3D avatar of a human subject using their own body gesturing style. There are several limitations to the fidelity and capability of 3D avatars, including a lack of facial movement or expression, as well as a lack of finger movement. A complete gesture style would include such elements. In addition, the 3D avatars are best suited for distance viewing, and not for close camera angles. We also expect that any gesture synthesis algorithm that preserved the style of the original gestures would be suitable for the application described in this work. Our study did not include audio; only the body movement was to be evaluated, and not the voice.

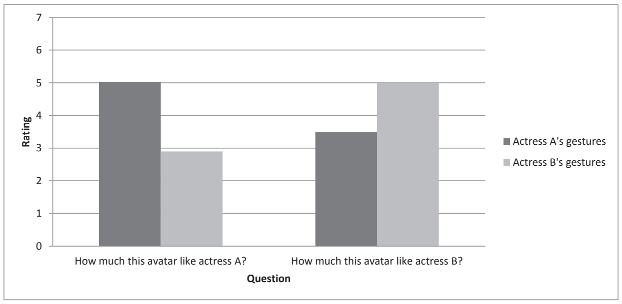

Figure 6: *Ratings showing how actress A's avatar was more like actress A when it gestured like actress A or like actress B, and vice-versa.*

Results of our evaluation confirmed that viewers identify an avatar as being more like the actor who inspired the avatar if it gestures like that actor rather than a different actor. Specifically, we found that when participants who knew one of our two amateur actresses viewed videos of her avatar, the gestures that avatar performed affected both ratings of the avatar and choice of which version of her avatar was more like her. First, an actress's avatar was rated as more like her when it gestured like that actress than when it gestured like the other actress. In the forced-choice, participants were also more likely to say that the avatar that had matching gestures was more that actress than to choose the very same avatar when it moved like the other actress.

In sum, these results suggest that an avatar can be made to seem more like the actor it was modeled after by adding movement to appearance. There may be important psychological implications of this for the user. For example, it is possible that we would engage with an avatar more like we would engage with the actor if the avatar appears more like the actor to us. This kind of transference, where we treat one agent like another agent if they seem alike, occurs as part of social perception [Andersen et al. 1995; Chen and Andersen 1999]. Research on transference suggests that, when a person appears more like another person, we engage with and treat that person as if they were the other person. For example, if a woman were to meet someone who seems like her mother, the woman would treat that new person like she treated her mother. Although the concept has psychoanalytical roots, researchers tend to view transference in social-cognitive terms; they would argue that this woman's behavior would reflect the activation and use of the representation of her mother when encountering this new person.

Recent research on transference more directly supports the possibility that gestures may help to activate and use of a representation of a given person during the process of social perception. For example, research on the intersection of face perception and transference found that facial feature resemblance can elicit transference [Kraus and Chen 2010]. When participants viewed an image of a stranger that was manipulated to look more like his or her significant other, they treated that stranger more like they would treat their interaction partner compared to those participants who saw unaltered images of a stranger's face. If gesture resemblance can also elicit transference (like facial-feature resemblance does), it is possible that avatars who utilize matching gestures may lead people to engage with the avatar more like they would the actor who inspired it.

Future research should explicitly test whether adding matching gestures lead people to engage with an avatar like they would treat the actor compared to avatars without matching gestures. Further psychological work should also verify that these kinds of social-cognitive mechanisms (i.e., activation of a person's representation) account for such engagement when avatars are considered in place of humans. Additionally, avatars that are modeled after actors that are trusted and liked by users, to the extent that the avatar can encourage users to treat them like the actor, such relational features (i.e., trust, liking) could be enhanced. Additional research should consider whether other forms of movement, such as locomotion, attentive gazes, idle posturing and weight shifts would also make the avatar seem more like the actor, potentially triggering users to engage with the avatar more like it were the actor himself.

6.1 Potential Impact

Our findings have a potential impact for immersive training. For example, some military groups uses VR for training teams and squads, the members of which know each other closely as they live and work together. In such simulations, the agents all wear similar uniforms, helmets and equipment that makes it difficult to identify individuals based on physical appearance, regardless of the personalization of the avatar in the virtual environment. However, if infusing these avatars with gestures captured from each individual makes them more uniquely recognizable, this has strong potential benefits for team coordination and training value. Beyond the scope of this work, we seek to understand if we can achieve more engagement with an avatar if it more closely resembles, in both appearance and behavior, the original actor.

This framework also gives us the opportunity to study how similarities in nonverbal behavior impact interpersonal relations - such as trust and liking. We can for example place person a's behaviors on an avatar of person b and see how that differentially impact attitudes of a for b.

There is a literature that people like and trust people that look like themselves [Bailenson et al. 2005], that people have built in cultural preferences for things like gaze patterns and interpersonal distance. There is less evidence for things like similar gestures. this raises the question whether similar gestures and postures might have an impact on factors like trust and liking. This could have an impact in education and training; if trust and liking could be improved through replication of gestural style, then training outcomes that depend on trust and liking could likewise be impacted.

References

ANDERSEN, S. M., GLASSMAN, N. S., CHEN, S., AND COLE, S. W. 1995. Transference in social perception: The role of chronic accessibility in significant-other representations. *Journal of Personality and Social Psychology 69*, 41–57.

BAILENSON, J., IYENGAR, S., AND YEE, N. 2005. Facial identity capture and presidential candidate preference. *Proceedings of the 55th Annual International Communication Association*.

BAILENSON, J. N. 2012. Doppelgangers-a new form of self? *PSYCHOLOGIST 25*, 1, 36–38.

BARAN, I., AND POPOVIĆ, J. 2007. Automatic rigging and animation of 3d characters. *ACM Trans. Graph. 26*, 3 (July).

BHARAJ, G., THORMÄHLEN, T., SEIDEL, H.-P., AND THEOBALT, C. 2011. Automatically rigging multi-component characters. *Comp. Graph. Forum (Proc. Eurographics 2012) 30*, 2.

CASAS, D., VOLINO, M., COLLOMOSSE, J., AND HILTON, A. 2014. 4D Video Textures for Interactive Character Appearance.

Computer Graphics Forum (Proc. EUROGRAPHICS 2014) 33, 2, 371–380.

CHEN, S., AND ANDERSEN, S. M. 1999. Relationships from the past in the present: Significant-other representations and transference in interpersonal life. *Advances in experimental social psychology 31*, 123–190.

CHIU, C.-C., AND MARSELLA, S. 2014. Gesture generation with low-dimensional embeddings. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, International Foundation for Autonomous Agents and Multiagent Systems, 781–788.

CUI, Y., CHANG, W., NÖLL, T., AND STRICKER, D. 2013. Kinectavatar: fully automatic body capture using a single kinect. In *Computer Vision-ACCV 2012 Workshops*, Springer, 133–147.

CUTTING, J. E., AND KOZLOWSKI, L. T. 1977. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the psychonomic society 9*, 5, 353–356.

FENG, A., HUANG, Y., XU, Y., AND SHAPIRO, A. 2014. Fast, automatic character animation pipelines. *Computer Animation and Virtual Worlds 25*, 1, 3–16.

FOX, J., AND BAILENSON, J. N. 2009. Virtual self-modeling: The effects of vicarious reinforcement and identification on exercise behaviors. *Media Psychology 12*, 1, 1–25.

FOX, J., AND BAILENSON, J. 2010. The use of doppelgängers to promote health behavior change. *CyberTherapy & Rehabil 3*, 2, 16–17.

GLEICHER, M. 1998. Retargetting motion to new characters. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, ACM, 33–42.

HERNANDEZ, M., CHOI, J., AND MEDIONI, G. 2012. Laser scan quality 3-d face modeling using a low-cost depth camera. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, 1995–1999.

IPISOFT. Motion Capture for the Masses, year = 2014, url = http://ipisoft.com/, urldate = 2014-09-15.

KRAUS, M. W., AND CHEN, S. 2010. Facial-feature resemblance elicits the transference effect. *Psychological Science 21*, 4, 518–522.

LEE, S., CARLSON, G., JONES, S., JOHNSON, A., LEIGH, J., AND RENAMBOT, L. 2010. Designing an expressive avatar of a real person. In *Intelligent Virtual Agents*, Springer, 64–76.

LEVINE, S., THEOBALT, C., AND KOLTUN, V. 2009. Real-time prosody-driven synthesis of body language. *ACM Trans. Graph. 28*, 5 (Dec.), 172:1–172:10.

LEVINE, S., KRÄHENBÜHL, P., THRUN, S., AND KOLTUN, V. 2010. Gesture controllers. In *ACM SIGGRAPH 2010 papers*, ACM, New York, NY, USA, SIGGRAPH '10, 124:1–124:11.

LI, H., VOUGA, E., GUDYM, A., LUO, L., BARRON, J. T., AND GUSEV, G. 2013. 3d self-portraits. *ACM Transactions on Graphics (Proceedings SIGGRAPH Asia 2013) 32*, 6 (November).

LOULA, F., PRASAD, S., HARBER, K., AND SHIFFRAR, M. 2005. Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Performance 31*, 1, 210.

LUO, P., NG-THOW-HING, V., AND NEFF, M. 2013. An examination of whether people prefer agents whose gestures mimic their own. In *Intelligent Virtual Agents*, Springer, 229–238.

MARSELLA, S., XU, Y., LHOMMET, M., FENG, A., SCHERER, S., AND SHAPIRO, A. 2013. Virtual character performance from speech. In *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, ACM, New York, NY, USA, SCA '13, 25–35.

2013. Mixamo auto-rigger. http://www.mixamo.com/c/auto-rigger.

NEFF, M., KIPP, M., ALBRECHT, I., AND SEIDEL, H.-P. 2008. Gesture modeling and animation based on a probabilistic recreation of speaker style. *ACM Transactions on Graphics 27*, 1, 5.

SHAPIRO, A., FENG, A., WANG, R., LI, H., BOLAS, M., MEDIONI, G., AND SUMA, E. 2014. Rapid avatar capture and simulation using commodity depth sensors. *Computer Animation and Virtual Worlds 25*, 3-4, 201–211.

STARCK, J., AND HILTON, A. 2007. Surface capture for performance-based animation. *Computer Graphics and Applications, IEEE 27*, 3, 21–31.

STONE, M., DECARLO, D., OH, I., RODRIGUEZ, C., STERE, A., LEES, A., AND BREGLER, C. 2004. Speaking with hands: creating animated conversational characters from recordings of human performance. *ACM Trans. Graph. 23*, 3 (Aug.), 506–513.

SUMNER, R. W., SCHMID, J., AND PAULY, M. 2007. Embedded deformation for shape manipulation. In *ACM Transactions on Graphics (TOG)*, vol. 26, ACM, 80.

TONG, J., ZHOU, J., LIU, L., PAN, Z., AND YAN, H. 2012. Scanning 3d full human bodies using kinects. *Visualization and Computer Graphics, IEEE Transactions on 18*, 4, 643–650.

WANG, R., CHOI, J., AND MEDIONI, G. 2012. Accurate full body scanning from a single fixed 3d camera. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, IEEE, 432–439.

WILCOXON, F. 1945. Individual comparisons by ranking methods. *Biometrics bulletin*, 80–83.

ZENG, M., ZHENG, J., CHENG, X., AND LIU, X. 2013. Templateless quasi-rigid shape modeling with implicit loop-closure. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, IEEE, 145–152.