Emotion Transfer for Hand Animation

Ana-Sabina Irimia
Northumbria University

Jacky C. P. Chan
Caritas Institute of Higher Education

Kamlesh Mistry
Northumbria University

Wei Wei
Xi'an University of Technology

Edmond S. L. Ho*
Northumbria University

ABSTRACT

We propose a new data-driven framework for synthesizing hand motion at different emotion levels. Specifically, we first capture high-quality hand motion using VR gloves. The hand motion data is then annotated with the emotion type and a latent space is constructed from the motions to facilitate the motion synthesis process. By interpolating the latent representation of the hand motion, new hand animation with different levels of emotion strength can be generated. Experimental results show that our framework can produce smooth and consistent hand motions at an interactive rate.

CCS CONCEPTS

• **Computing methodologies → Animation**; *Machine learning*.

KEYWORDS

hand animation, emotion, motion capture, style transfer

ACM Reference Format:
Ana-Sabina Irimia, Jacky C. P. Chan, Kamlesh Mistry, Wei Wei, and Edmond S. L. Ho. 2019. Emotion Transfer for Hand Animation. In *MIG '19: ACM SIGGRAPH conference on Motion, Interaction and Games, Oct 28–30, 2019, Newcastle upon Tyne, UK*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

This paper introduces a new method of style transfer for hand animation. The first objective of our method is to create new motions by adding synthesized styles to a base motion. Our method first extracts the style from motions and applies it to a base motion to create new animations for the character. The second objective of the project is to create motions for the character that present four emotions: anger, sadness, fear, and joy. The character can express those emotions only by its hand movement, as some characters may not have a face or voice to express their emotions. Using only their fingers and palm, making it a challenging task for an existing method to express the emotions.

Liu et al. proposed an interactive physics-based motion synthesis technique for manipulating a 3D hand model [6]. The interactive physics-based simulation algorithm is capable of creating seemingly mundane hand movements that are hard to achieve with keyframe animation or motion capture. An optimization-based approach to hand manipulation of grasping pose is introduced by Liu et al. [7]. The process of animating the hand is an automatic method that starts by inserting the initial input, the grasping pose, and the partial trajectory of the object, thus resulting in a physically plausible hand animation. Ye et al. presented a randomized sampling algorithm that can synthesize detailed and physically plausible hand manipulation given input full-body human motion and interacting object [9]. Yunfei Bai and C. Karen Liu proposed a solution to the problem of manipulating the orientation of a polygonal object using both palm and fingers of a robotic hand [1]. The aforementioned methods can acquire hand motion for performing a specific task in only a single style or a random style. In this paper, we would like to generate hand motion with styles specified by user.

The contributions in this work can be summarized as follows:

- We captured and annotated a new hand motion dataset with 7 motion types in 5 different emotions.
- We proposed an efficient hand motion synthesis framework which can be used for synthesizing new hand motion with different emotion strength levels at an interactive rate.

2 METHODOLOGY

2.1 Motion Capture

To the best of our knowledge, there is no publicly available hand motion dataset with a wide range of hand motions as well as different emotion stats and styles. We decided to use the Senso VR glove (https://senso.me/) to capture hand motion in this research. Each frame in the hand motion is represented by a vector P_j

$$P_j = [p_{0,x}, p_{0,y}, p_{0,z}, \ldots, p_{n-1,x}, p_{n-1,y}, p_{n-1,z}] \quad (1)$$

where j is the frame index, n is the total number of joints and $n = 23$ in the hand model we used, and p is the joint angle of the corresponding joint and rotation axis, respectively. We removed the hand translations in this research to avoid the artifacts caused by the incorrect hand translations tracked by the gloves. We captured 7 different types of hand motion, including *Crawling, Griping, Pat, Impatient, Hand on Mouse, Pointing and Pushing*, and 5 different emotions are associated with each motion type (i.e. *, neutral, angry, happy, sad, and scary*). In total, we captured 35 motion sequences.

2.2 Standardizing Hand Motions

Standardizing raw data is an important step before we statistically model the hand motion for the motion synthesis tasks. In particular, motion sequences are usually having a different duration. To facilitate the data standardization process, the subject performed each motion type in exactly 2 cycles. In this work, we take a simple

*Corresponding author. email: e.ho@northumbria.ac.uk

MIG '19, Oct 28–30, 2019, Newcastle upon Tyne, UK

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

approach by extracting the same number of keyframes from each captured hand motion. We empirically found that good performance can be achieved by extracting 9 keyframes in our experiments. As a result, each motion sequence is represented by a vector M in the joint angle space:

$$M = [Pk_0, ..., Pk_{k-1}] \tag{2}$$

where k is the total number of keyframes and $k = 9$, Pk_i is the i-th keyframe and Pk is having the same representation as in P (Eq. 1).

2.3 Style Transfer for Hand Motion

To statistically model the hand motion, we first project the hand motion from the joint angle space into the latent space. Advantages in editing motion in the latent space, such as reducing the redundancy in the data, emotion transfer, preserving the naturalness and quality, have been demonstrated in previous research [2–5, 8]. Motivated by the aforementioned encouraging results, we project the joint angles of the hand pose in each frame into the latent space using PCA. A matrix, AP, which contains all hand poses is prepared

$$AP = [Pk_{0,0}, Pk_{0,1}, ..., Pk_{m-1,k-1}]^T \tag{3}$$

where m is the total number of hand motion, $P_{i,j}$ refers to the joint angles (Eq. 1) in the j-th frame in the m-th motion. The resultant dimension of AP is a $(m \times k)$-by-(27×3) matrix.

Next, AP is projected into the latent representation lP by PCA. We empirically found that using the first **20** principal components to represent each hand pose yields reasonable results. Thus, the dimension of the lP matrix is $(m \times k)$-by-20. Each motion sequence is then represented by a vector lM which concatenate the latent representation of each keyframe.

2.3.1 Interpolating Emotion Strength. One of the potential application of the proposed system is to create new hand motion by controlling the *emotion strength*. Here, we propose a simple approach to linearly interpolate the emotion strength in the latent space. For each motion type, we can interpolate between the *neutral* motion and another *styled* motion (i.e. motion with emotion) to generate new motion lM_{new} using the equation below

$$lM_{new} = S \times (lM_{type-a,emo} - lM_{type-a,ne}) + lM_{type-a,ne} \tag{4}$$

where S is the emotion strength, $lM_{type-a,emo}$ and $lM_{type-a,ne}$ refer to the same hand motions type (i.e. $type - a$ in this example) with emotion (i.e. angry, happy, sad or scary) and neutral styles, respectively. Finally, the latent motion representation will be back projected to the joint angle representation in order to animate the rigged 3D hand model in the final animation.

3 RESULTS

Here, we show some of the results obtained in our experiments. Due to the limited space, we show the interpolation results on 2 motion sets. We interpolate the emotion at 3 different levels: *neutral (0%)*, 50% and *full (100%)* using Eq. 4. The first set is the interpolation of *sad* emotion on the *Pointing* motion set (Figure 1). It can be seen that the hand motions become more soft and less exaggerated when the *sadness* level increases. The second set is the interpolation of *happy* emotion on the *Hand on Mouse* motion set (Figure 2). It can be seen that the hand motions become more soft and less exaggerated when

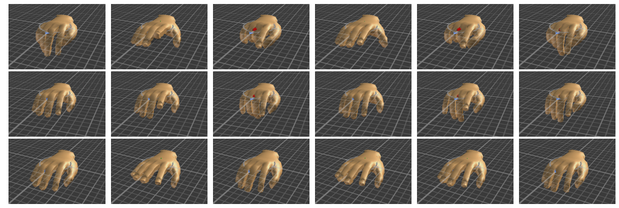

Figure 1: Interpolating the *sad* emotion on the *Pointing* motion with different levels of strength: *neutral (0%)* (1st row), 50% (2nd row) and *sad (100%)* (3rd row).

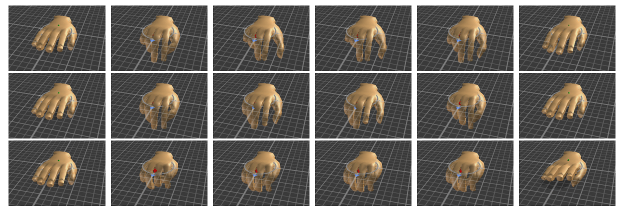

Figure 2: Interpolating the *happy* emotion on the *Hand on Mouse* motion with different levels of strength: *neutral (0%)* (1st row), 50% (2nd row) and *sad (100%)* (3rd row).

the *happiness* level increases. More synthesized hand animations can be found in the video demo submitted with this article.

4 CONCLUSION AND DISCUSSIONS

In this paper, we propose a new framework for synthesizing hand motion with different levels of emotion strength. The new motion synthesis framework is efficient and enables motion synthesis at an interactive rate. A more sophisticated approached that is based on learning the relationship between low-level motion features and high-level emotion strength using Ranking Support Vector Machine [3, 5] can be integrated into our framework in the future.

REFERENCES

[1] Y. Bai and C. K. Liu. 2014. Dexterous manipulation using both palm and fingers. In *Proceedings of ICRA 2014*. 1560–1565. https://doi.org/10.1109/ICRA.2014.6907059
[2] Jinxiang Chai and Jessica K. Hodgins. 2005. Performance Animation from Low-dimensional Control Signals. In *ACM SIGGRAPH 2005 Papers (SIGGRAPH '05)*. ACM, New York, NY, USA, 686–696. https://doi.org/10.1145/1186822.1073248
[3] Jacky C. P. Chan, Hubert P. H. Shum, He Wang, Li Yi, Wei Wei, and Edmond S. L. Ho. 2019. A generic framework for editing and synthesizing multimodal data with relative emotion strength. *Computer Animation and Virtual Worlds* 0, 0 (2019), e1871. https://doi.org/10.1002/cav.1871 e1871 cav.1871.
[4] Edmond S. L. Ho, Hubert P. H. Shum, Yiu-ming Cheung, and P. C. Yuen. 2013. Topology Aware Data-Driven Inverse Kinematics. *Computer Graphics Forum* 32, 7 (2013), 61–70. https://doi.org/10.1111/cgf.12212
[5] Edmond S. L. Ho, Hubert P. H. Shum, H Wang, and L Yi. 2017. Synthesizing Motion with Relative Emotion Strength. In *ACM SIGGRAPH ASIA Workshop: Data-Driven Animation Techniques (D2AT)*. http://eprints.whiterose.ac.uk/121250/
[6] C. Karen Liu. 2008. Synthesis of Interactive Hand Manipulation. In *Proceedings of SCA '08 (SCA '08)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 163–171. http://dl.acm.org/citation.cfm?id=1632592.1632616
[7] C. Karen Liu. 2009. Dextrous Manipulation from a Grasping Pose. In *ACM SIGGRAPH 2009 Papers (SIGGRAPH '09)*. ACM, New York, NY, USA, Article 59, 6 pages.
[8] H. P. H. Shum, E. S. L. Ho, Y. Jiang, and S. Takagi. 2013. Real-Time Posture Reconstruction for Microsoft Kinect. *IEEE Transactions on Cybernetics* 43, 5 (Oct 2013), 1357–1369. https://doi.org/10.1109/TCYB.2013.2275945
[9] Yuting Ye and C. Karen Liu. 2012. Synthesis of Detailed Hand Manipulations Using Contact Sampling. *ACM Trans. Graph.* 31, 4, Article 41 (July 2012), 10 pages. https://doi.org/10.1145/2185520.2185537