EG UK Computer Graphics & Visual Computing (2020), pp. 1–8
P. D. Ritsos and K. Xu (Editors)

Emotion Transfer for 3D Hand Motion using StarGAN

Jacky C. P. Chan[1] and Ana-Sabina Irimia[2] and Edmond S. L. Ho[†2]

[1]Department of Computer Science, Hong Kong Baptist University, Hong Kong
[2]Department of Computer and Information Sciences, Northumbria University, United Kingdom

Abstract

In this paper, we propose a new data-driven framework for 3D hand motion emotion transfer. Specifically, we first capture high-quality hand motion using VR gloves. The hand motion data is then annotated with the emotion type and converted to images to facilitate the motion synthesis process and the new dataset will be available to the public. To the best of our knowledge, this is the first public dataset with annotated hand motions. We further formulate the emotion transfer for 3D hand motion as an Image-to-Image translation problem, and it is done by adapting the StarGAN framework. Our new framework is able to synthesize new motions, given target emotion type and an unseen input motion. Experimental results show that our framework can produce high quality and consistent hand motions.

Keywords: hand animation, emotion, motion capture, generative adversarial network, style transfer

1. Introduction

Hand motion plays a vital role in computer animation since the subtle hand gestures can express a lot of different meanings and are useful for understanding the personality of a person [WTWN16]. A classic example would be the character *Thing T. Thing* of the *"The Addams Family"* which is a hand, and it can 'act' and express a lot of different emotions solely by the fingers and hand movements. While the computer animation community focuses on new methodologies for synthesizing full-body motions, researchers have been proposing frameworks [JHS12, YL12] for synthesizing hand and finger movements based on the given full-body motion to improve the expressiveness of the animation.

However, synthesizing hand motion is not a trivial task. Capturing hand motion using an optical motion capture system is not easy as the fingers are in proximity, and the labeling of the markers can be mixed up easily. As a result, most of the previous hand motion synthesis frameworks are based on physics-based motion generation models [Liu08, AK12, Liu09, YL12, BL14]. Recently, more effective hand motion capturing approaches are proposed. Alexanderson et al. introduce a new system for a passive optical motion capture system that can better obtain correct markers labels of fingers in real-time [AOB16]. Han et al. [HLW*18] improves the difficulties in marker labeling for the optical MOCAP system using convolutional neural networks. While hand motion can be synthesized or captured using the aforementioned approaches, those motions are always challenging to be reused because of the difficulties in transferring the styles to improve the expressiveness in different scenes.

This paper introduces a new method of emotion transfer for hand animation. The task is to change a given hand animation to another particular style, such as changing the emotion type from neutral to angry (Figure 4 to 7). Our objective is to create motions for the character to 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 fingers and palm, making it a challenging task for an existing method to express different emotion types. Inspired by the success in style transfer on images using generative adversarial network (GAN) based frameworks (such as CycleGAN [ZPIE17], DualGAN [YZTG17], and StarGAN [CCK*18]), we propose formulating the hand motion emotion transfer as an Image-to-Image translation problem. We first propose an image-based representation for the hand motion to facilitate the learning process using StarGAN in the later stage. 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. The dataset will be available to the public. To the best of our knowledge, this is the first public dataset with annotated hand motions.
- We proposed a new image-based representation for the captured hand motion to facilitate the learning process of the emotion transfer using the StarGAN framework.
- We provide qualitative results on hand motion emotion transfer using the proposed framework, also showing its validity by comparing them to captured motions.

† Corresponding author, email: e.ho@northumbria.ac.uk

2. Related Work

2.1. Hand animation

Examples of hand animation can be easily found in various applications such as movies, games, and animations. However, capturing hand motion using existing motion capture systems is not a trivial task. There was no commercial or academic real-time vision-based hand motion capture solution until a recent work presented by Han et al. [HLW*18]. As a result, most of the previous work focused on synthesizing hand motions based on physics-based models. Liu et al. [Liu08] presented an interactive physics-based motion synthesis technique that can manipulate a 3D hand model to create seemingly mundane hand movements that are hard to achieve with keyframe animation or motion capture. Andrews and Kry [AK12] proposed a hand motion synthesis framework for object manipulation. The method divides a manipulation task into 3 phases (approach, actuate, and release), and each phase is associated with a control policy for generating physics-based hand motion.

Liu et al. [Liu09] introduced an optimization-based approach to hand manipulation of grasping pose. By providing the grasping pose and the partial trajectory of the object, a physically plausible hand animation will be created. Ye et al. [YL12] presented a randomized sampling algorithm that can synthesize detailed and physically plausible hand manipulation based on the input full-body human motion, and the object interacted with the subject. Bai and Liu [BL14] presented a solution to the problem of manipulating the orientation of a polygonal object using both the palm and fingers of a robotic hand. Their method considers the physical properties such as collisions, gravitational, and contact forces.

For data-driven approaches, a PCA-based framework is proposed in [WJZ13] for generating detailed hand animation from a set of sparse markers. Jörg et al [JHS12] proposed a data-driven framework for synthesizing the finger movements for an input full-body motion. The methods employ a simple approach for searching for an appropriate finger motion from the pre-recorded motion database based on the input wrist and body motion. The aforementioned methods can acquire hand motion for performing a specific task in only a single style or a random style. Irimia et al. [ICM*19] proposed a framework for generating hand motion from interpolation. Having hand motions captured with different types of emotions, the hand poses are collected and projected to the latent space using PCA. New motion can be created by interpolating the hand poses using the latent representation. In contrast, our framework enables emotion transfer between different hand motions.

2.2. Style transfer for motion

Motion style transfer is a technique used to convert the style of a motion to another style, thus creating new motions without losing the primitive content of the original one. An early work by Unuma et al. [UAT95] proposed using Fourier principles to create an interactive and real-time control of locomotion with emotion, as well as to include cartoon-ish exaggerations and expressions. Amaya et al. [ABC96] introduced a model that could emulate emotional animation by signal processing technique. The emotional transform is based on the speed and spatial amplitude of the movements. Hsu et al. [HPP05] presented a solution for translating the style of a human motion by comparing the difference of the behaviour of the aligned input and output motions using a linear time-invariant model. Shapiro et al. introduced a novel method of interactive motion data editing based on motion decomposition, which separates the style and expressiveness from the main motion [SCF06]. The method uses Independent Component Analysis (ICA) to separate the style from the motion data.

Holden et al. presented a new approach to transferring motion style by using a convolution neural network [HHKK17]. It takes a single motion as an example that can represent the style. Different from other methods, this method does not need to align the motion with the style clips since the style is calculated by the average of the frames from a motion's Gram matrix. One problem of style transfer is the high computational cost. Smith et al. [SCNW19] recently proposed an efficient style transfer framework based on a compact neural network architecture which consists of 3-layered network structure for Pose Network, Foot Contact Network, and Timing Network. Lee and Popović [LP10] proposed an inverse reinforcement learning based approach to learn a motion controller with behavior styles from a small set of motion samples. The behavior style can then be transferred to different unseen environments.

Until now, the only research of style transfer was for a full body character, thus making this paper the first emotion transfer research for the human hand. While we share a similar interest with the pilot study [ICM*19] on synthesizing hand motion with emotion, the previous work is technically interpolating emotion strength instead of emotion transfer.

3. Methodology

In this section, the proposed emotion transfer framework will be presented, and the overview is shown in Figure 1. Firstly, we introduce a new hand motion database captured using Senso VR Glove (Section 3.1). The new database contains hand motions with various emotional states and styles. Next, the captured motions are standardized (Section 3.2) as a pre-processing step for the learning process. The motion data will then be transformed into an RGB image representation for learning the emotion transfer model using StarGAN. The StarGAN model learns how to generate a new image given a target domain label and the input image (Section 3.3). Finally, the synthesized new image will be converted to the joint angle space for generating the final 3D animation (Section 3.4). The details of each step will be explained in the following subsections.

3.1. Capturing hand motion

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 status and styles. While some hand motions with different emotion types are captured in a pilot study [ICM*19], the data is not available to the public. To facilitate this research and stimulate future work in the related areas, we decided to capture and construct a new hand motion database. Several kinds of motion capture (MOCAP) systems can be used for capturing hand motions, such as optical MOCAP, depth sensor-based system (e.g. Microsoft Kinect), Leap Motion (`https://www.leapmotion.com/`),

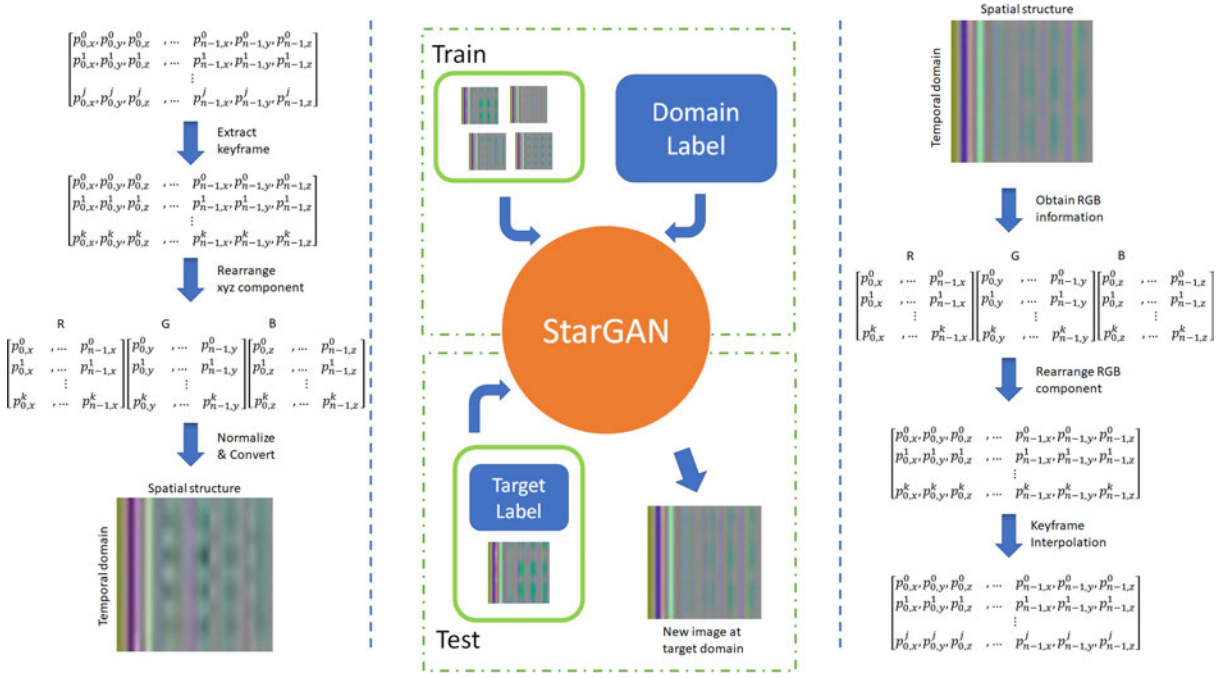

Figure 1: *The overview of the proposed emotion transfer framework. (left) Convert motion data to image format. (middle) StarGAN learn how to generate realistic fake image given a sample and target domain label. (right) Obtain new motion data from the generated image*

and VR gloves, etc. Given that there are a lot of finger movements in a relatively small space, self-occlusion occurs very frequently. In contrast, motions captured using a motion capture glove will have a better resolution, with fewer glitches, twists, and turns. As a result, we decided to use VR gloves to capture hand motion over the vision-based systems. In particular, the Senso VR glove (see Figure 2, `https://senso.me/`) is used in this research.

Senso VR glove is a light-weighted and wireless VR device. There are 7 Inertial Measurement Unit (IMU) sensors on each glove to track the hand and finger movements. The motion data will then be transferred to a computer for processing through Bluetooth wireless connection. There are 5 small vibration motors attached to the fingertips to generate haptic feedback. However, only the motion capturing functionality is used here. We used the Senso VR Glove to capture the motion of the right hand when constructing the database.

Each hand motion at each frame is represented by a vector P_j

$$P_j = [p^0_{j,x}, p^0_{j,y}, p^0_{j,z}, ..., p^{n-1}_{j,x}, p^{n-1}_{j,y}, p^{n-1}_{j,z}] \quad (1)$$

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. Since the hand translations cannot be accurately captured using the Senso VR Glove alone, we removed the hand translations in this research to avoid the artifacts caused by the incorrect hand translations.

We captured 7 different types of hand motion, including *Crawling*, *Gripping*, *Patting*, *Impatient*, *Hand on Mouse*, *Pointing*, and *Pushing*. There are 5 different types of emotions associated with each motion type and their characteristics are listed on Table 1. In total, 35 motion sequences were captured and the new hand motion database will be available to the public to stimulate the research in the community.

(a) *Senso VR glove* **(b)** *IMU sensors and vibration motors inside the glove*

Figure 2: *Senso VR Glove (images are reproduced from* `https://senso.me/`*)*

3.2. Standardizing hand motion

The captured hand motion data varies significantly in the joint angles representation both spatially and temporally. To facilitate the learning process in the later stage, data standardization (or normalization) is used. While some advanced techniques such as Recurrent Neural Network (RNN) can be used to model data sequences with variations in length, such a method requires a significant

Emotion	Characteristics
Angry	exaggerated, fast, large range of motion
Happy	energetic, large range of motion
Neutral	normal, styleless motion
Sad	sign of tiredness, small range of motion
Fearful	asynchronous finger movements, small range of motion

Table 1: *The 5 types of emotions used in this research and their characteristics.*

amount of data to train the model, which is not feasible with the dataset we have collected.

To facilitate the data standardization process, the subject performed each motion type in exactly 2 cycles. By this, the motion sequences can be temporally standardized by simple rescaling or finding the correspondence between frames by Dynamic Time Warping (DTW) [BC94]. However, hand motion contains a lot of temporal redundancy as the joint angles change smoothly over time. To reduce the dimensionality of the data to be handled by the deep learning framework, we extracted the same number of keyframes from each captured hand motion to facilitate the learning process. This will also make sure that all the images converted from the motions (to be explained in Section 3.3.1) are of the same size and retain the original information. 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).

3.3. Emotion transfer for hand motion

In this section, the process of transferring emotion types between hand motions will be explained. Inspired by the encouraging results in transferring styles between images from 2 different domains using Image-to-Image translation frameworks such as CycleGAN [ZPIE17] and DualGAN [YZTG17], we propose to formulate the emotion transfer problem as an Image-to-Image translation problem. There are several advantages to this approach. Firstly, a wide range of high-performance frameworks (such as [ZPIE17, YZTG17, CCK*18]) are available. Secondly, the effectiveness of encoding human motion (with joint angles representation) using convolutional neural networks (CNNs) has been demonstrated in [HSK16]. Thirdly, the generative adversarial network (GAN) based framework requires relatively fewer training samples than other deep learning architectures, which is more appropriate for this research with limited hand motion data.

In the rest of this section, we will first explain how to convert a hand motion into an image-based representation in Section 3.3.1. Next, the StarGAN [CCK*18] framework adopted in this work will be explained in Section 3.3.2. Finally, new hand motion will be reconstructed from the synthesized image (Section 3.4).

3.3.1. Representing hand motion as an image

In order to use Image-to-Image domain translation framework for hand motion emotion transfer, we represent a hand motion sequence as an image. The xyz-coordinate components of motion are arranged chronologically as the RGB components of the image. Each frame of a motion is represented as a row of an image while each joint of a motion is represented as a column of an image. Hence, each keyframed hand motion M will be arranged as a 3-channel $(27 \times k)$ matrix. The values in the 3-channel matrix are then re-scaled into the range of [0, 255], which is the common range of RGB value, as follows:

$$v_{i,c}^m = round(255 \times \frac{(p_{i,c}^m - p_{min})}{(p_{max} - p_{min})}) \tag{3}$$

where m is the joint index, i is the keyframe index, $c \in \{x, y, z\}$ represents the channel index, $V_{i,c}^m$ is the normalized pixel value, p_{max} and p_{min} are the maximum and minimum values among all the joint angles existed in the dataset. Noted that the images are saved in Bitmap format to avoid the data loss during compression. Examples of the image representation of the neutral motions are illustrated in Figure 3. It can be seen that the different motions are represented by different image patterns which will be useful for extracting the discriminative patterns in the learning process.

3.3.2. Emotion transfer as Image-to-Image domain translation

One of the potential applications of the proposed system is to create new hand motion by controlling the *emotion labels*. To support the translation between multi-domain and considering the robustness and scalability, StarGAN [CCK*18] is adapted to translate motion from one emotion to another emotion while preserving the basic information of the input motion. Comparing to typical GANs with cycle consistency losses such as CycleGAN [ZPIE17] and DualGAN [YZTG17] for style transfer, StarGAN [CCK*18] can perform image-to-image translations for multiple domains using only a single model which is suitable for transferring different types of emotions.

The StarGAN [CCK*18] model consists of two modules, a discriminator D and a generator G. The discriminator learns to distinguish between real and fake samples and performs domain classification. The generator learns to generate fake samples f that looks real r, given images with original domain labels and target domain labels. By leverages the adversarial loss $\mathcal{L}_{adv}$, domain classification loss $\mathcal{L}_{cls}$ and reconstruction loss $\mathcal{L}_{rec}$ during training:

$$\mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_{cls}\mathcal{L}_{cls}^r \tag{4}$$

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{cls}\mathcal{L}_{cls}^f + \lambda_{rec}\mathcal{L}_{rec} \tag{5}$$

where λ_{cls} and λ_{rec} are the hyper-parameters for controlling whether domain classification or reconstruction is more important. It makes the generated images as realistic as possible, at the same time, can be classified as target domain and preserve the content of its input images. So given an input motion with original domain labels and target emotion domain, StarGAN can generated a realistic motion performing with target emotion. Readers are refereed to [CCK*18] for the technical details.

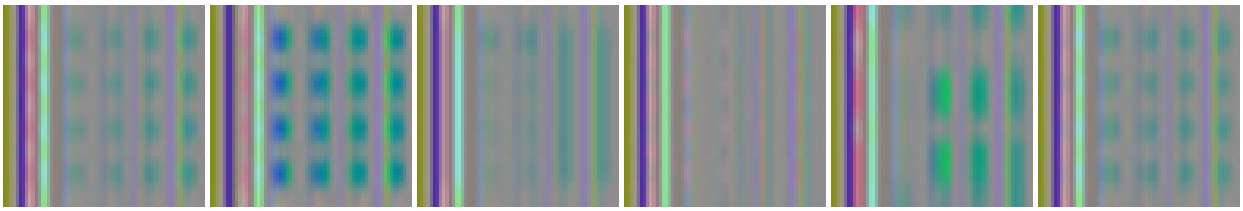

Figure 3: *Examples of the image representation of neutral motions. From left to right: Crawling, Gripping, Impatient, Patting, Pointing and Pushing.*

3.4. Reconstructing Hand Motion from Generated Images

Having synthesized new images from StarGAN, new hand motions are reconstructed from the images by re-scaling the RGB values:

$$p^m_{i,c} = (\frac{v^m_{i,c}}{255} \times (p_{max} - p_{min})) + p_{min} \quad (6)$$

where $v^m_{i,j}$ is the pixel value for the c-th channel of the m-th row at i-th column on the synthesized image, p_{max} and p_{min} are same values as in Equation 3. Next, we rearrange the pixel values to convert the image representation back to the keyframed motion M and finally the new motion is produced by interpolation on those keyframes.

4. Experimental Results

In this section, we evaluate the effectiveness of the proposed framework. We first evaluate the emotion transfer performance by training the StarGAN framework using the hand motion captured using the Senso Glove (Section 3.1). To ensure the framework learns the underlying style instead of 'memorizing' the motion, we employ a leave-one-out cross-validation approach to split the data into training and testing sets. The readers are referred to the video demo for more results.

4.1. Evaluation on Emotion Transfer

Here, we show some of the results obtained in our experiments. Due to the limited space, we visualize the results (Figure 4 to 7) on 4 motion sets including *crawling*, *patting*, *impatient* and *pushing*.

From the results, consistent experimental results are obtained. Specifically, input motions become more exaggerated by transferring to *angry*. The range of motion increases and the motion becomes faster. This is highlighted by the movement of the thumb in the video demo. By transferring to the *happy* emotion, the motion becomes more energetic with a larger range of motion when compared with the input (*neutral*) motion. The speed of the motion is getting higher as well, although the motion is less exaggerated than those transferred to *angry*. With the *sad* emotion, the synthesized motions show the sign of tiredness, which results in slower movement. Finally, the *fearful* emotion brings the asynchronous finger movements to the neutral motion as those characteristics can be found in the captured data. In summary, the consistent observation of the synthesized motions highlighted the effectiveness of our framework.

Figure 4: *Screenshots (one frame per row) of transferring the input (neutral) crawling motion to different types of emotions. Columns from left to right: angry, happy, input (neutral), sad, fearful.*

4.2. Comparing Emotion Transferred Motions with Captured Data

Next, we compare the synthesized motions with the captured data. Recall that leave-one-out cross-validation is used in defining the training and testing data sets. As a result, the motion type of the input (i.e. testing) motion is not included in the training data. It is possible that the action of the synthetic motion looks slightly different from the captured motion. Having said that, an effective emotion transfer framework should be able to transfer the characteristics of the corresponding emotion to the new motion.

The results are illustrated in Figure 8 to 12. It can be seen that the motions synthesized by the proposed framework have the characteristics of the corresponding emotion. For example, the motions are exaggerated in the *angry* crawling and pushing motions in Figure 8 and 11, respectively. On the other hand, the *sad* crawling motion shows the sign of tiredness. The *fearful* emotion can again transfer the asynchronous finger movements to the pushing motion,

Figure 5: *Screenshots (one frame per row) of transferring the input (neutral) impatient motion to different types of emotions. Columns from left to right: angry, happy, input (neutral), sad, fearful.*

Figure 6: *Screenshots (one frame per row) of transferring the input (neutral) pushing motion to different types of emotions. Columns from left to right: angry, happy, input (neutral), sad, fearful.*

Figure 7: *Screenshots (one frame per row) of transferring the input (neutral) patting motion to different types of emotions. Columns from left to right: angry, happy, input (neutral), sad, fearful.*

as illustrated in Figure 9. Finally, a larger range of motion can be seen in the *happy* impatient motion (Figure 10).

5. Conclusion and Discussions

In this paper, we propose a new framework for synthesizing hand motion by emotion transfer. In addition to the new framework, we further contribute to the community by making the hand motion dataset with emotion annotation available to the public to stimulate the research in the related areas. Experimental results show that our method can 1) generate different styles of motions according to the emotion type, and 2) the characteristics in each emotion type can be transferred to new motions.

Currently, the framework is evaluated based on the visual quality of the synthesized hand motions. In the future, we are interested in conducting a user-study to evaluate how users perceive the emotion from the synthesized motions. We will evaluate if the emotion demonstrated in our captured hand motions align with human judgment and perception. It is also an interesting future direction to quantitatively evaluate the results by comparing the differences between the synthesized and ground-truth motion numerically. To achieve this goal, more hand motions have to be captured. In this work, we have difficulties in capturing the global translation and rotation in high quality. As a result, the global transformation is discarded, which limits the expression of the emotion. One of the possible solutions is to capture the hand motions using state-of-the-art MOCAP solutions such as [HLW*18]. Finally, synthesizing motion with different emotion strength [CSW*19] can also be incorporated in the emotion transfer framework in the future.

Figure 8: *Screenshots (one frame per row) of the comparison between the input (neutral, left), synthesized (transferred to angry, middle) and captured (angry, right) crawling motions.*

Figure 9: *Screenshots (one frame per row) of the comparison between the input (neutral, left), synthesized (transferred to fearful, middle) and captured (fearful, right) pushing motions.*

Acknowledgements

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

Figure 10: *Screenshots (one frame per row) of the comparison between the input (neutral, left), synthesized (transferred to happy, middle) and captured (happy, right) impatient motions.*

Figure 11: *Screenshots (one frame per row) of the comparison between the input (neutral, left), synthesized (transferred to angry, middle) and captured (angry, right) pushing motions.*

References

[ABC96] AMAYA K., BRUDERLIN A., CALVERT T.: Emotion from motion. In *Proceedings of the Conference on Graphics Interface '96* (Toronto, Ont., Canada, Canada, 1996), GI '96, Canadian Information Processing Society, pp. 222–229. URL: `http://dl.acm.org/citation.cfm?id=241020.241079`.

[AK12] ANDREWS S., KRY P. G.: Policies for goal directed multi-finger

Figure 12: *Screenshots (one frame per row) of the comparison between the input (neutral, left), synthesized (transferred to sad, middle) and captured (sad, right) crawling motions.*

manipulation. In *VRIPHYS* (2012).

[AOB16] ALEXANDERSON S., O'SULLIVAN C., BESKOW J.: Robust online motion capture labeling of finger markers. In *Proceedings of the 9th International Conference on Motion in Games* (2016), ACM, pp. 7–13.

[BC94] BERNDT D. J., CLIFFORD J.: Using dynamic time warping to find patterns in time series. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining* (1994), AAAIWS'94, AAAI Press, pp. 359–370. URL: http://dl.acm.org/citation.cfm?id=3000850.3000887.

[BL14] BAI Y., LIU C. K.: Dexterous manipulation using both palm and fingers. In *2014 IEEE International Conference on Robotics and Automation (ICRA)* (May 2014), pp. 1560–1565. doi:10.1109/ICRA.2014.6907059.

[CCK*18] CHOI Y., CHOI M., KIM M., HA J., KIM S., CHOO J.: Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), pp. 8789–8797.

[CSW*19] CHAN J. C. P., SHUM H. P. H., WANG H., YI L., WEI W., HO E. S. L.: A generic framework for editing and synthesizing multimodal data with relative emotion strength. *Computer Animation and Virtual Worlds 30*, 6 (2019), e1871. e1871 cav.1871. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cav.1871, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/cav.1871, doi:10.1002/cav.1871.

[HHKK17] HOLDEN D., HABIBIE I., KUSAJIMA I., KOMURA T.: Fast neural style transfer for motion data. *IEEE Computer Graphics and Applications 37*, 4 (2017), 42–49. doi:10.1109/MCG.2017.3271464.

[HLW*18] HAN S., LIU B., WANG R., YE Y., TWIGG C. D., KIN K.: Online optical marker-based hand tracking with deep labels. *ACM Trans. Graph. 37*, 4 (July 2018). URL: https://doi.org/10.1145/3197517.3201399, doi:10.1145/3197517.3201399.

[HPP05] HSU E., PULLI K., POPOVIĆ J.: Style translation for human motion. In *ACM SIGGRAPH 2005 Papers* (New York, NY, USA, 2005), SIGGRAPH '05, ACM, pp. 1082–1089. URL: http://doi.acm.org/10.1145/1186822.1073315, doi:10.1145/1186822.1073315.

[HSK16] HOLDEN D., SAITO J., KOMURA T.: A deep learning framework for character motion synthesis and editing. *ACM Trans. Graph. 35*, 4 (July 2016). URL: https://doi.org/10.1145/2897824.2925975, doi:10.1145/2897824.2925975.

[ICM*19] IRIMIA A.-S., CHAN J. C. P., MISTRY K., WEI W., HO E. S. L.: Emotion transfer for hand animation. In *Motion, Interaction and Games* (New York, NY, USA, 2019), MIG '19, ACM, pp. 41:1–41:2. URL: http://doi.acm.org/10.1145/3359566.3364692, doi:10.1145/3359566.3364692.

[JHS12] JÖRG S., HODGINS J., SAFONOVA A.: Data-driven finger motion synthesis for gesturing characters. *ACM Trans. Graph. 31*, 6 (Nov. 2012). URL: https://doi.org/10.1145/2366145.2366208, doi:10.1145/2366145.2366208.

[Liu08] LIU C. K.: Synthesis of interactive hand manipulation. In *Proceedings of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (Aire-la-Ville, Switzerland, Switzerland, 2008), SCA '08, Eurographics Association, pp. 163–171. URL: http://dl.acm.org/citation.cfm?id=1632592.1632616.

[Liu09] LIU C. K.: Dextrous manipulation from a grasping pose. In *ACM SIGGRAPH 2009 Papers* (New York, NY, USA, 2009), SIGGRAPH '09, ACM, pp. 59:1–59:6. URL: http://doi.acm.org/10.1145/1576246.1531365, doi:10.1145/1576246.1531365.

[LP10] LEE S. J., POPOVIĆ Z.: Learning behavior styles with inverse reinforcement learning. *ACM Trans. Graph. 29*, 4 (July 2010). URL: https://doi.org/10.1145/1778765.1778859, doi:10.1145/1778765.1778859.

[SCF06] SHAPIRO A., CAO Y., FALOUTSOS P.: Style components. In *Proceedings of Graphics Interface 2006* (Toronto, Ont., Canada, Canada, 2006), GI '06, Canadian Information Processing Society, pp. 33–39. URL: http://dl.acm.org/citation.cfm?id=1143079.1143086.

[SCNW19] SMITH H. J., CAO C., NEFF M., WANG Y.: Efficient neural networks for real-time motion style transfer. *Proc. ACM Comput. Graph. Interact. Tech. 2*, 2 (July 2019). URL: https://doi.org/10.1145/3340254, doi:10.1145/3340254.

[UAT95] UNUMA M., ANJYO K., TAKEUCHI R.: Fourier principles for emotion-based human figure animation. In *Proceedings of the 22Nd Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 1995), SIGGRAPH '95, ACM, pp. 91–96. URL: http://doi.acm.org/10.1145/218380.218419, doi:10.1145/218380.218419.

[WJZ13] WHEATLAND N., JÖRG S., ZORDAN V.: Automatic hand-over animation using principle component analysis. In *Proceedings of Motion on Games* (New York, NY, USA, 2013), MIG '13, Association for Computing Machinery, p. 197–202. URL: https://doi.org/10.1145/2522628.2522656, doi:10.1145/2522628.2522656.

[WTWN16] WANG Y., TREE J. E. F., WALKER M., NEFF M.: Assessing the impact of hand motion on virtual character personality. *ACM Trans. Appl. Percept. 13*, 2 (Mar. 2016). URL: https://doi.org/10.1145/2874357, doi:10.1145/2874357.

[YL12] YE Y., LIU C. K.: Synthesis of detailed hand manipulations using contact sampling. *ACM Trans. Graph. 31*, 4 (July 2012), 41:1–41:10. URL: http://doi.acm.org/10.1145/2185520.2185537, doi:10.1145/2185520.2185537.

[YZTG17] YI Z., ZHANG H., TAN P., GONG M.: Dualgan: Unsupervised dual learning for image-to-image translation. In *2017 IEEE International Conference on Computer Vision (ICCV)* (2017), pp. 2868–2876.

[ZPIE17] ZHU J., PARK T., ISOLA P., EFROS A. A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)* (2017), pp. 2242–2251.