Personality-Driven Gaze Animation with Conditional Generative Adversarial Networks

Funda Durupinar
funda.durupinarbabur@umb.edu
University of Massachusetts Boston
Boston, Massachusetts

ABSTRACT

We present a conditional generative adversarial learning approach to synthesize the gaze behavior of a given personality. Training is done using an existing data set that comprises eye-tracking data and personality traits of 42 participants performing an everyday task. Given the values of Big-Five personality traits (openness, conscientiousness, extroversion, agreeableness, and neuroticism), our model generates time series data consisting of gaze target, blinking times, and pupil dimensions. We use the generated data to synthesize the gaze motion of virtual agents on a game engine.

CCS CONCEPTS

• **Computing Methodologies**; • **Machine Learning**; • **Machine Learning Algorithms**; • **Computer Graphics**; • **Animation**;

KEYWORDS

gaze animation, generative adversarial networks, convolutional neural networks

ACM Reference Format:
Funda Durupinar. 2020. Personality-Driven Gaze Animation with Conditional Generative Adversarial Networks. In *Proceedings of MIG'20: ACM SIGGRAPH Conference on Motion, Interaction, and Games (MIG '20).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Expressive eye movements are essential for believable virtual character animation. They effectively communicate attention in addition to conveying information about the emotional and mental states of the individual. Personality is among the factors that control and explain the various manners of gaze behavior. Studies show correlations between different aspects of personality and gaze parameters such as gaze shifts and blink rates [1, 4, 8].

In this work, we propose a data-driven, generative approach to synthesize gaze behaviors for different personalities. We use data acquired from individuals in an everyday setting as opposed to data from actors playing a given role based on known personality-gaze correlations [9]. This helps capture the small details of gaze cues not yet conceptualized but reflecting certain personality traits. We employ the Big-Five (openness, conscientiousness, extroversion, agreeableness, and neuroticism) model of personality [2]. We train a generative adversarial network (GAN) conditioned on personality using an existing personality-annotated dataset [4]; and animate the eye movements of a virtual model with the generated data. Although deep learning has been used to synthesize gaze movement [5], the applications are limited to eye and body pose coordination for target following. To our knowledge, this work is the first to apply deep learning to generate eye movement data based on personality expression.

MIG '20,
,

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/1122445.1122456

2 METHOD

2.1 Gaze Parameter Synthesis

To synthesize gaze parameters, we build a GAN conditioned on personality values [3]. For training, we use the dataset provided by Hoppe et al. [4], which consists of binocular eye movement data of 42 participants, each with an average of 12.51 minutes of tracking information and personality scores for five factors binned into three groups of low, medium and high. The dataset includes participants' Big-Five values, time-series data for gaze coordinates, blinking times, and pupil dimensions. The data was acquired by head-mounted eye trackers while participants walked around the campus and purchased an item of their choice from a campus shop. The availability of personality information and the everyday nature of the performed tasks make this dataset a good fit for our goals.

The GAN is composed of two competing networks: a discriminator (D) and a generator (G), trained simultaneously. G learns a distribution p_g over data x while D is trained to discriminate between the real data and synthetic data $G(z)$, where z is input noise drawn from a random normal distribution. Conditional GAN [6] extends the model on given classes, allowing direct data generation given class labels. In our model, we integer-encode the personality values into class labels. The model handles conditional labels y with distribution p_l to minimize $\mathbb{E}_{x\sim p_r, y\sim p_l}[log(D(x,y))] + \mathbb{E}_{z\sim p_g, y\sim p_l}[1 - log(D(G(z,y),y)]$. For G and D, use the DCGAN architecture [7], but with 1D convolutional neural networks(CNNs) for time-series data generation and discrimination. CNNs allow time-invariant feature extraction and can be easily extended to multivariate time-series data.

Because the blinking information is discrete (0 or 1 depending on blink detection at each timestep), it requires an additional step between the generator and the discriminator to provide continuous gradients. For this, we pre-train an autoencoder to encode and decode the binary data, and send the output of the generator to the decoder first, then send the output of the decoder to the discriminator. The model architecture is shown in Figure 1.

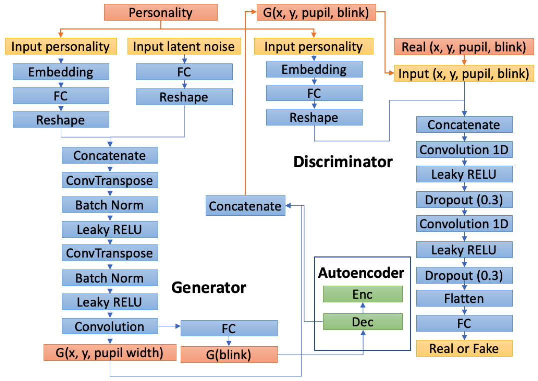

Figure 1: Label-preserving, conditional GAN architecture.

Discriminator input is a vector comprising the x and y coordinates of gaze positions, pupil diameters, and binary blinking data. We organize the data by sliding windows of size 300 corresponding to 5 seconds of data sampled at 60 Hz. For each of these 300-frame windows, we perform a strict quality test and discard the windows that include at least one row with x or y coordinates beyond the $[0, 1]$ range or pupil dimensions equal to zero. This leaves only the valid data points. Before feeding the continuous data into the GAN, we normalize it into the $[-1, 1]$ range. Personality comprises the conditional class label that specifies each participant. It is introduced to the network as a 50-dimensional embedding vector that encodes 243 possible values (3^5 for each bin and personality dimension). Since the dataset is limited, only 24 of 243 possible classes are represented in the training set. This conditions the discriminator on the *seen* classes, but allows the generator to predict sequences for *unseen* classes. We also train the model for each personality dimension separately, where the personality dimension has three labels representing the low, medium, and high values per personality. We train mini-batches of size 64, using Adam optimizer with a learning rate of 0.0001 both for D and G.

2.2 Evaluation

To evaluate the GAN model quantitatively, we train a deep 1-D CNN classifier on real data, synthesize a large number of data points and predict the probability of them belonging to each personality bin (class). Inception score is a metric to summarize these predictions [10]. Table 1 shows the average scores of 1000 iterations for the test data for real and synthetic values. When we train the classifier on all the five dimensions, the inception score is low because the representation in the training data is limited. We also compute the inception scores when each dimension is introduced as a condition separately. Considering that there are three classes per personality dimension, the closer the inception score to 3, the better the results.

Table 1: Inception scores for synthetic and real data

Data	O	C	E	A	N	All dims
Synthetic	2.38	2.25	2.56	2.41	2.56	6.23
Real test	2.87	2.63	2.88	2.78	2.89	15.62

2.3 Gaze Animation

We animate the eye movement, pupil dilation, and blinking on a 3D humanoid model with blend shapes for face. The eye tracker glasses that were used to capture gaze data (SMI) have 60° horizontal and 46° vertical field of view angles. Using these angles, we convert the x and y values, which are in the range $[0, 1]$ corresponding to the device space coordinates, to the world space. The target in the world space is the look-at direction of the eyes. For convenience, we take the middle point of the left and right eyes as the eye position. In addition to rotating the eyeballs to align with the look-at vector, we update the weights of the eyelid blendshapes so that they move naturally when the eyes move up and down. We update the pupil dimension by applying forces to the vertices on the pupil perimeter towards or out of the center of the pupil. Figure 2 shows a model with different personalities and gaze behaviors.

Figure 2: Animated gaze of an extrovert (left) vs. an introvert (right) model.

3 CONCLUSION

Our method is a preliminary step in synthesizing and animating time-series gaze data. The next step will be to create a personality-annotated gaze dataset during a conversation and to use our generative approach on this set. The dataset will include similar features, in addition to information about the conversation target as well as head and torso pose information. We believe that the social nature of the task will help capture more salient features of personality expression. We will evaluate the realism of the animations through user studies.

REFERENCES

[1] Shlomo Berkovsky, Ronnie Taib, Irena Koprinska, Eileen Wang, Yucheng Zeng, Jingjie Li, and Sabina Kleitman. 2019. Detecting Personality Traits Using Eye-Tracking Data. In *Human Factors in Computing Systems (CHI '19)*. ACM, 1–12.
[2] Lewis R. Goldberg. 1990. An Alternative "Description of Personality": The Big-Five Factor Structure. *Journal of Personality and Social Psychology* 59 (1990), 1216–1229.
[3] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Eds.). Curran Associates, Inc., 2672–2680.
[4] Sabrina Hoppe, Tobias Loetscher, Stephanie A. Morey, and Andreas Bulling. 2018. Eye Movements During Everyday Behavior Predict Personality Traits. *Frontiers in Human Neuroscience* 12 (2018), 105.
[5] Alex Klein, Zerrin Yumak, Arjen Beij, and A. Frank van der Stappen. 2019. Data-Driven Gaze Animation Using Recurrent Neural Networks. In *Motion, Interaction and Games (MIG '19)*. Association for Computing Machinery, Article 4, 11 pages.
[6] M. Mirza and Simon Osindero. 2014. Conditional Generative Adversarial Nets. *ArXiv* abs/1411.1784 (2014).
[7] A. Radford, Luke Metz, and Soumith Chintala. 2016. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *CoRR* abs/1511.06434 (2016).
[8] John F. Rauthmann, Christian Seubert, Pierre Sachse, and Marco Furtner. 2012. Eyes as windows to the soul: Gazing behavior is related to personality. *Journal of Research in Personality* 46 (2012), 147–156.
[9] Kerstin Ruhland, Katja Zibrek, and Rachel McDonnell. 2015. Perception of personality through eye gaze of realistic and cartoon models. In *ACM SIGGRAPH Symposium on Applied Perception (SAP'15)*.
[10] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. 2016. Improved Techniques for Training GANs. arXiv:1606.03498 [cs.LG]