Learning Stylized Character Expressions from Humans

Deepali Aneja[1] Alex Colburn[2] Gary Faigin[3] Linda Shapiro[1] Barbara Mones[1]
[1]University of Washington, Seattle [2]Zillow Group [3]Gage Academy of Art
http://grail.cs.washington.edu/projects/deepexpr/

1. Introduction

We present **DeepExpr**, a novel expression transfer system from humans to multiple stylized characters via deep learning. We developed : 1) a data-driven perceptual model of facial expressions, 2) a novel stylized character data set with cardinal expression annotations : FERG (Facial Expression Research Group) - DB, and 3) a mechanism to accurately retrieve plausible character expressions from human expression queries. We evaluated our method on a set of retrieval tasks on our collected stylized character dataset of expressions. We have also shown that the ranking order predicted by the proposed features is highly correlated with the ranking order provided by a facial expression expert and Mechanical Turk (MT) experiments.

2. Related Work

Historically, the Facial Action Coding System (FACS) [2] is often used for Facial Expression Recognition (FER) and character animation systems. We have shown FACS-based systems do not consistently generate expressions that are recognized by humans as the intended expression. Modern methods apply CNNs to the FER problem and are quite effective in recognition tasks [4]. CNN features fused with geometric features for customized expression recognition [6] and Deep Belief Networks have also been utilized to solve the FER, but have mixed results when used to generate expressions.

3. Methodology

To learn deep CNN models that generalize well across a wide range of expressions, we need sufficient training data to avoid over-fitting of the model. For human facial expression data collection, we combined publicly available annotated facial expression databases. We also created a novel database (**FERG-DB**) of labeled facial expressions for six stylized characters. An animator created the key poses for each expression, and they were labeled via MT to populate the database initially. The details of the database collection and data pre-processing are given in [1].

We have extended **FERG-DB** by adding two new characters as shown in Fig. 1 with almost 10,000 images for each new character. All the images are labeled for each of six cardinal expressions: joy, sadness, anger, surprise, fear, disgust, and neutral.

Figure 1. Two new characters added to FERG-DB with surprise expression (left) and anger expression (right).

3.1. Data Processing and Training

For our human dataset, we register faces to an average frontal face via an affine transformation. Geometric measurements between the points are also taken to produce geometric features for refinement of expression retrieval results. Once the faces are cropped and registered, the images are re-sized for analysis.

With approximately 70,000 images of labeled samples of human faces and 50,000 images for stylized character faces, we trained two Convolutional Neural Networks to recognize the expressions of humans and stylized characters independently. Then we utilize a transfer learning technique to learn the mapping from humans to characters by creating a shared embedding feature space.

To create the shared feature space, we fine-tuned the CNN pre-trained on the human dataset with the character dataset for every character by continuing the backpropagation step. The last fully connected layer of the human trained model was fine-tuned, and earlier layers were kept fixed to avoid overfitting. We decreased the overall learning rate while increasing the learning rate on the newly initialized next-to-last layer. This embedding also allows human expression-based image retrieval and character expression-based image retrieval.

3.2. Distance Metrics

In order to retrieve the stylized character with the closest expression match to the human expression, we formulate a distance metric with terms for expression clarity and geo-

Figure 2. Result from our combined approach - DeepExpr and Geometric features. The leftmost image is the query image and all six characters are shown portraying the top match of the same joy expression.

metric distance. The closest expression match minimizes the distance function in eq. 1 :

$$_d = \ j\text{JS Distance}j + \ j\text{Geometric Distance}j \quad (1)$$

where the clarity term, JS Distance, is the Jensen-Shannon divergence [3] distance between next to last layer feature vectors of *human* and *character*, and Geometric distance is the L^2 norm distance between geometric features of *human* and *character*. Our implementation uses JS Distance as a retrieval parameter and then geometric distance as a sorting parameter to refine the retrieved results with and as relative weight parameters.

4. Results and Discussion

DeepExpr features combined with geometric features produce significant performance enhancement in retrieving the stylized character facial expressions based on human facial expressions. The top results for a human joy expression on six stylized characters are shown in Fig. 2.

We measured the retrieval performance of our method by calculating the average normalized rank of relevant results [5]. The best score of 0 indicates that all the relevant database images are retrieved before all other images in the database. A score that is greater than 0 denotes that some false positives are retrieved before all relevant images. Table 1 shows that DeepExptr consistently achieves lower scores than geometry alone, and Figure 3 illustrates with examples.

The Spearman and Kendall correlation coefficients of our combined approach ranking with the expert ranking and MT test ranking for 30 validation experiments show high correlation score of more than 0.8 for 93% of the experiments.

Expression	Geometry	DeepExpr
Anger	0.384	0.213
Disgust	0.386	0.171
Fear	0.419	0.228
Joy	0.276	0.106
Neutral	0.429	0.314
Sad	0.271	0.149
Surprise	0.322	0.125

Table 1. Average retrieval score for each expression across all characters using only geometry and DeepExpr features.

Our system demonstrates a perceptual model of facial expressions that provides insight into facial expressions displayed by stylized characters. The model can also be incorporated into the animation pipeline to help animators and artists to better understand expressions, communicate how to create expressions to others, and transfer expressions from humans to stylized characters.

Figure 3. Best match results from our approach compared to only geometry-based retrieval for disgust (top) and fear (bottom).

References

[1] D. Aneja, A. Colburn, G. Faigin, L. Shapiro, and B. Mones. Modeling stylized character expressions via deep learning. In *Asian Conference on Computer Vision (ACCV 2016)*. Springer, 2016. 1

[2] P. Ekman and W. Friesen. Facial action coding system: A technique for the measurement of facial movement. *Consulting Psychologists Press*, 1978. 1

[3] J. Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 1991. 2

[4] A. Mollahosseini, D. Chan, and M. H. Mahoor. Going deeper in facial expression recognition using deep neural networks. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, pages 1–10. IEEE, 2016. 1

[5] H. Müller, S. Marchand-Maillet, and T. Pun. The truth about corel-evaluation in image retrieval. In *Image and Video Retrieval*, pages 38–49. Springer, 2002. 2

[6] X. Yu, J. Yang, L. Luo, W. Li, J. Brandt, and D. Metaxas. Customized expression recognition for performance-driven cutout character animation. In *Winter Conference on Computer Vision*, 2016. 1