A Simple Approach to Animating Virtual Characters by Facial Expressions Reenactment

白泽琛，姚乃明，刘璐，陈辉*，王宏安

Poster: 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (IEEE VR **CCF-A**)

联系人：陈辉　　联系电话：18611384680　邮箱：chenhui@iscas.ac.cn

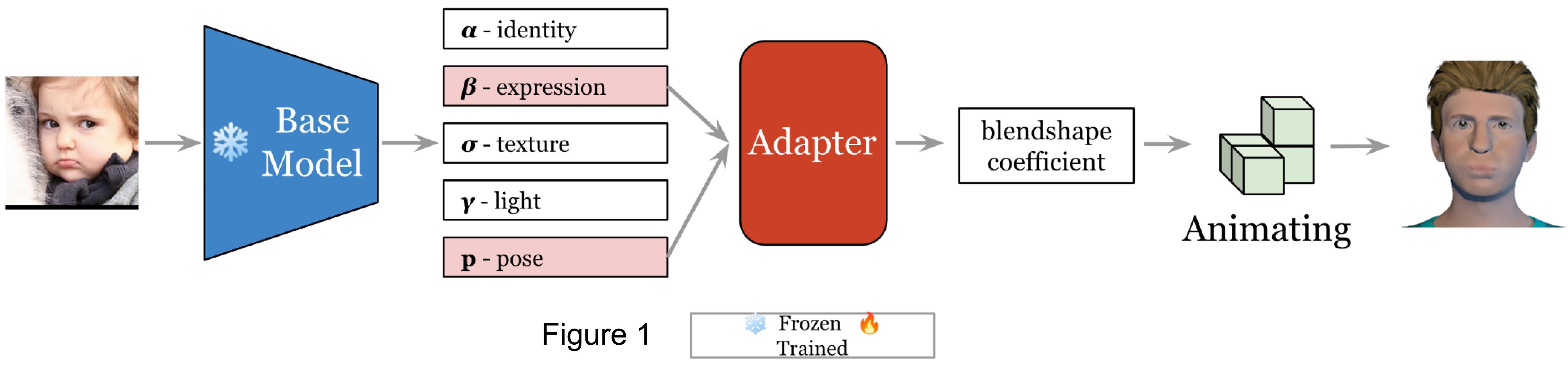

Figure 1

Introduction

Animating virtual characters is one of the core problems in virtual reality. Facial animations are able to intuitively express emotion and attitudes of virtual characters. However, creating facial animations is a non-trivial task. It depends on either expensive motion capture devices or human designers' time and effort to tune the animation parameters.

In this work, we propose a learning-based approach to animate virtual characters by facial expression reenactment from abundant image data. This approach is simple yet effective, and generalizable to various 3D characters. Preliminary evaluation results demonstrate its effectiveness and its potential to accelerate the development of VR applications.

Figure 2

Results

Qualitative Examples

In Fig. 2, we show examples with various virtual character gender and appearance. As can be observed, the reenacted facial expressions are capable of vividly replicating the source facial expressions, validating the effectiveness of our approach. Even though there are some subtle muscle movements that are not reproduced perfectly, the estimated coefficients still provide us with a fair reasonable initial state.

Comparison with human designers

We compare the proposed approach and human designers' performance in terms of satisfaction score and time usage . The result is shown in Tab.1. Though the human designer achieves a higher satisfaction score than the proposed approach, the margin is limited. For the inference time to generate or adjust blendshape coefficients, the proposed approach costs only 0.41s for running the whole testing pipeline, while the human designer need over 236s to tune the blend shape coefficient values into the desired state. The huge gap shows the effectiveness of our approach to reducing designers' workload without losing much fidelity.

Methods

As illustrated in Fig.1, the approach contains two models. We first pre-train a base model that estimates generic 3D facial parameters from the given image. Based on the frozen base model, we train a lightweight adapter model to adapt the generic parameters into the desired blendshape coefficients of the target virtual character.

The linear 3D Morphable Model (3DMM) is employed as the generic 3D face representation, in which the face shape $\mathbf{S}$ and texture

$$\mathbf{S} = \bar{\mathbf{S}} + \alpha \mathbf{B}_{id} + \beta \mathbf{B}_{exp}\ ,\quad \mathbf{T} = \bar{\mathbf{T}} + \sigma \mathbf{B}_{tex},$$

where $\bar{\mathbf{S}}$ and $\bar{\mathbf{T}}$ are mean face shape and texture respectively. $\mathbf{B}_{id}$, $\mathbf{B}_{exp}$, and $\mathbf{B}_{tex}$ denote PCA bases of identity, expression, and texture respectively. Apart from shape and texture, we also introduce illu-mination model and camera model to define the light coefficient γ and pose coefficient p respectively. All the coefficients mentioned above are concatenated into a single vector $\mathbf{v} = (\alpha,\beta,\sigma,\gamma,p)$. The model takes human face images as input and predicts the 3DMM coefficients vectors.

For facial expression reenactment, the adapter model takes the expression coefficient β and pose coefficient p as input and predicts blendshape coefficients $\hat{y}$. The adapter is implemented as a lightweight Multi-Layer Perceptron (MLP) neural network. Specifically, after the last layer, there is a *Clamp* operator truncating the output value into the range 0 ~ 1, which enforces the model to output valid values of coefficients.

Generally, the pipeline can be formulated as:

$$\alpha,\beta,\sigma,\gamma,\mathrm{p} = \textbf{BaseModel}(x),$$

$$\hat{\mathrm{y}} = \textbf{Adapter}(\beta,\mathrm{p}),$$

where x denotes the input face image.

Table 1: Comparison with human designers.

	Satisfaction Score	Inference Time
Algorithm	6.36 ± 1.80	0.41 ± 0.06s
Designer	6.92 ± 1.91	236.12 ± 48.61s

Conclusion

In this work, we propose a simple yet effective approach to ease the burden of creating facial animations of virtual characters for VR applications. The task is formulated as a facial expressions reenactment problem and is achieved by estimating blendshape coefficients based on image data. This approach not only generalizes well within the character family with the same blendshape topology but is also easy to be adapted to other customized characters. Evaluation results have shown promising performance on this task and verified the effectiveness of our approach. We hope this work will inspire research on automatically animating virtual characters.