Local Anatomically-Constrained Facial Performance Retargeting

PRASHANTH CHANDRAN, ETH Zurich, Switzerland and DisneyResearch|Studios, Switzerland
LOÏC CICCONE, DisneyResearch|Studios, Switzerland
MARKUS GROSS, ETH Zurich, Switzerland and DisneyResearch|Studios, Switzerland
DEREK BRADLEY, DisneyResearch|Studios, Switzerland

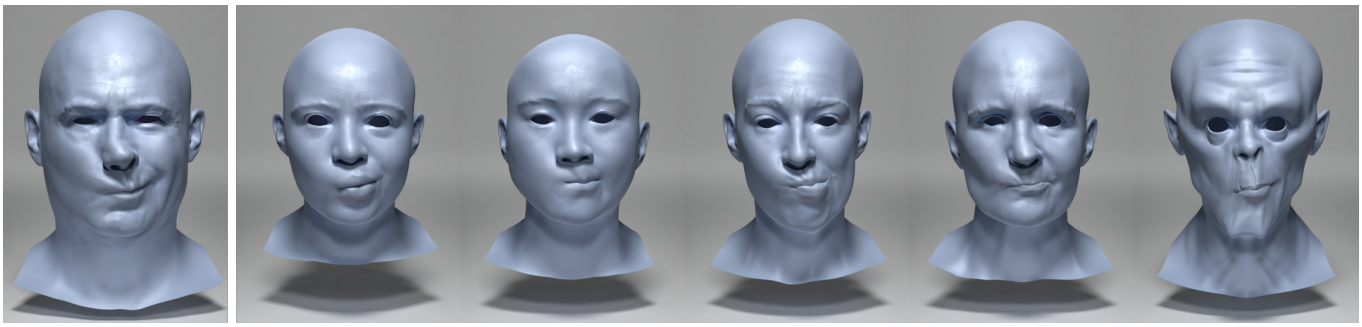

Fig. 1. We present a local, anatomically-constrained method for facial performance retargeting that is ideally suited for the complex problem of human-to-human facial animation transfer. Here we show one frame of retargeting the source character (left) to five different target characters (right). While the method targets human performances, it naturally also extends to fantasy characters (far right).

Generating realistic facial animation for CG characters and digital doubles is one of the hardest tasks in animation. A typical production workflow involves capturing the performance of a real actor using mo-cap technology, and transferring the captured motion to the target digital character. This process, known as *retargeting*, has been used for over a decade, and typically relies on either large blendshape rigs that are expensive to create, or direct deformation transfer algorithms that operate on individual geometric elements and are prone to artifacts. We present a new method for high-fidelity offline facial performance retargeting that is neither expensive nor artifact-prone. Our two step method first transfers local expression details to the target, and is followed by a global face surface prediction that uses anatomical constraints in order to stay in the feasible shape space of the target character. Our method also offers artists with familiar blendshape controls to perform fine adjustments to the retargeted animation. As such, our method is ideally suited for the complex task of human-to-human 3D facial performance retargeting, where the quality bar is extremely high in order to avoid the uncanny valley, while also being applicable for more common human-to-creature settings. We demonstrate the superior performance of our method over traditional deformation transfer algorithms, while achieving a quality comparable to current blendshape-based techniques used in production while requiring significantly fewer input shapes at setup time. A detailed user study corroborates the realistic and artifact free animations generated by our method in comparison to existing techniques.

CCS Concepts: • **Computing methodologies** → **Motion capture**; **Motion processing**.

Additional Key Words and Phrases: Facial Performance Retargeting, Facial Animation, Expression Transfer.

ACM Reference Format:
Prashanth Chandran, Loïc Ciccone, Markus Gross, and Derek Bradley. 2022. Local Anatomically-Constrained Facial Performance Retargeting. *ACM Trans. Graph.* 41, 4, Article 168 (July 2022), 14 pages. https://doi.org/10.1145/3528223.3530114

Authors' addresses: Prashanth Chandran, ETH Zurich, Switzerland and DisneyResearch|Studios, Switzerland, prashanth.chandran@disneyresearch.com; Loïc Ciccone, DisneyResearch|Studios, Switzerland, loic.ciccone@disneyresearch.com; Markus Gross, ETH Zurich, Switzerland and DisneyResearch|Studios, Switzerland, gross@disneyresearch.com; Derek Bradley, DisneyResearch|Studios, Switzerland, derek.bradley@disneyresearch.com.

0730-0301/2022/7-ART168 $15.00
https://doi.org/10.1145/3528223.3530114

1 INTRODUCTION

Character facial animation is a key aspect of many computer graphics applications. Creating realistic 3D facial animation is a difficult task, as even the slightest inaccuracies can make the result look uncanny. One practical way to obtain realism is to capture the performance of a real actor using motion-capture technology, and then transfer the resulting digital performance to a target 3D character. Such an approach is commonly referred to as *performance retargeting*, and is the primary method of generating facial animation for high-end visual effects in film and entertainment. While the motion capture side of the problem has seen tremendous technical advances over the past two decades, the retargeting side has advanced at a much slower pace. In practice, there are two methods commonly used in studio productions (and both are nearly two decades old). The first is based on blendshape animation [Lewis et al. 2014], where corresponding blendshape rigs are created for the source and target characters and retargeting becomes the simple task of copying the blendweights from the source performance to the target rig. This method requires very many facial shapes in the rigs, which must be in parity, incurring a large up-front cost. The benefit, however, is

that the source and target characters can be arbitrarily dissimilar (e.g. retargeting a young female face to an elderly man with wrinkles). The alternative method is to use deformation transfer [Sumner and Popović 2004], which requires only 1 (usually neutral expression) shape of the target character, and attempts to retarget triangle deformations from the source performance directly. While incurring a much smaller setup time, this method can be prone to geometric artifacts and is generally more applicable when the source and target are very similar, as surface details (e.g. wrinkles) from the source character would be copied to the target.

More recent 3D facial retargeting methods do exist, but are nearly all designed for retargeting human performances to cartoon or fantasy characters. A growing problem in the visual effects industry is the creation of photorealistic digital humans, where the precision and realism of facial animation is of highest importance. Here, if the actor corresponding to the digital human is available to perform, the solution lies primarily in the motion-capture domain. However, oftentimes the target human character is not physically available to perform, for example if they have passed away, or if the target character is a younger or older version of the actor[1]. In such scenarios, to create realistic facial animation we must resort again to the retargeting problem. Here, the quality bar is much higher than for retargeting to cartoon characters. We present a high-quality facial performance retargeting solution that is ideally suited for this realistic human-to-human retargeting scenario, while also demonstrating results for more traditional human-to-creature retargeting.

Our method considers the problem locally, by first retargeting small patches of the face surface individually. This offers a high degree of flexibility, allowing us to operate with only a small number of input shapes for establishing correspondence (i.e. complete facial rigs are unnecessary). In a second step, to retain global consistency, we fit a subject-specific anatomical face model, originally designed for monocular face tracking [Wu et al. 2016], by extending it to support performance retargeting.

Our method retains the benefits of both of the common approaches for facial retargeting (blendshapes and deformation transfer), without exhibiting either of their drawbacks. For example, with our approach the source and target characters can be arbitrarily dissimilar and our method will not directly copy fine expression details from one to the other. Furthermore, we can accomplish this with only a fraction of the number of input shapes that a typical blendshape retargeting approach would require (for example, approx. 20 versus hundreds of shapes). Our method also exhibits fewer geometric artifacts than deformation transfer, which we will demonstrate in our results. Finally, our approach allows easy artistic direction over the retargeted solution by providing a simple mechanism to favor or punish certain shape deformations in different regions of the face, or locally exaggerate the retargeting strength, all while staying in a plausible manifold of the target 3D character.

To summarize, our work presents a new practical, robust and flexible method for realistic facial performance retargeting, suitable for today's high demand for realistic digital characters.

[1] e.g, The Irishman (2019) - www.fxguide.com/fxfeatured/de-aging-the-irishman/

2 RELATED WORK

We now review related work in the various areas of performance retargeting.

2.1 General Motion Retargeting

In areas other than the human face, previous works in performance retargeting have focused on skeleton animation [Aberman et al. 2020], full 3D bodies for computer graphics [Baran et al. 2009; Borno et al. 2018] and robotics applications [Morishima et al. 2016; Penco et al. 2018], and hand animation for robotics [Antotsiou et al. 2018; Orbik et al. 2021] and sign language [Ge et al. 2005]. Rigged human body models such as the widely used *SMPL* model [Loper et al. 2015] and its recent variants [Osman et al. 2020; Santesteban et al. 2020] can be trivially re-purposed for retargeting. Since the face is typically parameterized differently than other parts of the body, such methods do not readily apply to facial animation.

2.2 Video Face Retargeting

There exists a large body of work on 2D video-based facial reenactment or face swapping [Chen et al. 2020a,b; Garrido et al. 2014; Kim et al. 2019; Naruniec et al. 0 07; Nirkin et al. 2019; Perov et al. 2021; Ren et al. 2021; Thies et al. 2016; Wang et al. 2021; Zhang et al. 2020]. Colloquially referred to as *DeepFakes*, these techniques have progressed to a point where given a performance, a desired actor's face can be photo-realistically retargeted as a video. In contrast, our method is concerned with 3D geometric retargeting rather than 2D video face swapping. Although certain techniques do use 3D facial geometry within their pipeline [Hong et al. 2021; Thies et al. 2016; Wang et al. 2021], this 3D information tends to be of low resolution as it primarily only serves as prior for the video generation. While these 2D retargeting techniques are indeed impressive, they lack some key benefits of a 3D approach, such as offering artists more control . For a detailed analysis of recent literature in the field of 2D face swapping, we refer to a survey by Mirsky and Lee [2020].

2.3 Real-time Puppeteering

Real-time puppeteering is a special case of retargeting, where the goal is to drive a virtual character's face in real time, and the focus is primarily on speed rather than on the quality of retargeting. Prior to the advent of deep learning, techniques for capturing and animating a digital character's face in real time included Adaptive PCA [Li et al. 2013], wherein the basis vectors of a PCA model were adapted by progressively observing captured frames of an actor. A monocular system for real time face capture was proposed by Cao et al. [2014] which used morphable model regressors to drive a digital character. Real time techniques that are primarily focused towards non-human character animations have also been proposed [Bouaziz et al. 2013; Weise et al. 2011]. These methods rely on a morphable model and are therefore only capable of producing approximate/coarse shapes. As such, these techniques are clearly not suitable for use in high end visual effects production.

With the advent of Telepresence, there has also been a focus on reproducing photo realistic digital doubles [Seymour et al. 2017] in real time. Some recent techniques in this space can indeed reproduce high fidelity avatars [Chen et al. 2021; Lombardi et al. 2018; Ma et al.

2021], however most of these techniques require large amounts of actor specific training data and generalize poorly to multiple actors and test conditions. In contrast, we focus on high quality offline 3D retargeting shapes, without requiring large amounts of training data.

2.4 Data driven 3D Retargeting

Neural face models are becoming increasingly popular owing to their performance and ability to model nonlinear skin deformations. Naturally, some of these models have been used for the purpose of 3D retargeting as well. Chandran et al. [2020] proposed the use of a disentangled variational auto encoder (VAE), which can fully isolate facial identity and expression in its latent space, thereby allowing for the swapping of expression codes across identities in the latent space to achieve 3D human-to-human retargeting. Zhang et al. [2022] recently proposed a framework where human and character specific VAEs share a common latent space; allowing a human face to drive the desired character's face. The use of neural architectures allows for other forms of retargeting; for instance driving a 3D face from an audio input [Karras et al. 2017]. Neural networks that predict the parameters of a rig or a blendshape model have also been developed for retargeting [Aneja et al. 2018; Chaudhuri et al. 2019; Costigan et al. 2014]. Another stream of recent research in 3D retargeting treats the problem similar to 2D face swapping: by performing the retarget first in 2D image space and then regressing rig/model parameters from the retargeted image [Kim et al. 2021; Moser et al. 2021]. The primary drawback of data driven techniques in 3D retargeting is their large requirements of training data and that they only satisfy the stringent quality requirements of production in a single to few character setting. The second drawback is that even if they do generalize across characters, SOTA techniques [Kim et al. 2021; Moser et al. 2021] resort to predicting linear morphable model parameters, resulting in a lack of realism in the retargeted result. Finally, they also offer limited room for artistic intervention.

2.5 Offline Performance Retargeting

Most closely related to our work are methods in offline performance retargeting. Blendshape rigs [Lewis et al. 2014] are an industry wide standard for facial animation. By adjusting the coefficients or blendshape weights of the rig, an artist can intuitively produce a desired expression in a character. In the context of performance retargeting, the coefficients of a source rig are estimated and transferred to the target rig. One such pipeline for estimating blendshape coefficients from a video and applying them to the target is described in Chuang and Bregler [2002]. Although blendshape rigs are intuitive and fast, production rigs with hundreds of shapes are time-consuming to create. Hence researchers have also explored techniques to create character rigs starting from a small subset of shapes [Li et al. 2010], and to maximize rig expressiveness using as few shapes as possible [Carrigan et al. 2020]. Despite their popularity, blendshape rigs have limited expressivity due to their linear nature. To produce subtle nonlinear face deformations, artists are often forced to sculpt hundreds of shapes and keyframe animate their coefficients, making facial animation a massive time-sink in production [Seol et al. 2011]. To address some short comings of such rigs, researchers have proposed several incremental improvements. These include rig augmentation [Kim et al. 2011], coefficient remapping [Song et al. 2011], rigs with skinned bones and corrective shapes [Li et al. 2017], and range of motion calibration between the source and target rigs [Ribera et al. 2017].

Another popular technique for 3D retargeting is deformation Transfer [Sumner and Popović 2004]. Given a source shape in a rest and deformed pose, deformation transfer computes the relative local deformations of triangles in the source and transfers them to a target shape in the rest pose. While extremely efficient, and capable of producing plausible retargets, deformation transfer suffers from the drawback of transferring the smallest wrinkles from the source to the target, resulting in the transfer of high frequency details that may not match the target character. Furthermore, naïve deformation transfer is often artifact prone, leading to self intersections in the geometry and requires additional regularization [Saito 2013]. A simple work-around to alleviate some of the geometric artifacts of deformation transfer while retaining similar properties is simply to perform a per-vertex delta transfer, where the 3D displacements from deformed to rest pose of the source mesh are blindly copied to the target rest mesh, creating the effect of retargeting. While simplistic in theory, this approach has also been used in practice, but also suffers from the drawback of transferring high frequency details from source to target.

Expression cloning [Noh and Neumann 2001] is another technique involving the transfer of expressions using known correspondences of a sparse set of points between the source and the target. Such techniques have been extended to automatically compute mappings between the source and the target [Bouaziz and Pauly 2014; Dutreve et al. 2008] and to consider additional constraints such as contours [Bhat et al. 2013]. Space time expression cloning [Seol et al. 2012] approaches retargeting by assuming that the source and target trajectories must be similar and formulates retargeting by interpreting facial movement as the derivative of position and by constraining the derivative with poisson boundary conditions. Decomposing a facial expression into large and fine scale deformations and transfering them in a single optimization was proposed by Xu et al. [2014]. A notion of locality in face retargeting was introduced by Liu et al. [2011] where the face is automatically segmented into multiple regions and has been used to retarget performances .

In summary, while the problem of facial performance retargeting has been studied for over a decade, current methods are either not suited for high quality human-to-human retargeting for visual effects, or those that are suited, such as blendshape animation and deformation/delta transfer have several drawbacks. We present the first method that does not have a large setup burden and is not prone to geometric artifacts, producing high-fidelity facial performance retargeting suitable for production. We will provide detailed comparisons of our method to the common approaches of blendshape retargeting, deformation transfer and delta transfer, highlighting the superior performance of our approach.

3 LOCAL ANATOMICAL RETARGETING

We now describe our method for local anatomical facial retargeting. Given a single frame from the source character performance, our

goal is to transfer the expression of the source character faithfully to the target character, while preserving the identity and nuances of the target character. We approach this retargeting problem in steps. In the first step, we tackle the retargeting task locally by breaking down a source face into a number of patches and estimating their deformations. These per-patch deformations are transferred over to the target character (Section 3.2) to yield an initial approximation of the retargeted shape. Then, since such a local transfer of deformations can yield inaccurate global face shapes, we perform a second step wherein we fit a character specific anatomical face model to the initial retargeted result, yielding a high fidelity target character shape (Section 3.3). The steps of our method are illustrated in Fig. 2. As our method is directly aimed for a workflow in film production, it also offers artists with several semantically meaningful knobs that they can use to achieve the final look for the retargeted character. Our method operates on a frame-level, and can be trivially parallelized over the whole sequence and is naturally suited for both single shot expression transfer and performance retargeting. Before we explain the details, we first describe a one-time model setup procedure that is needed to build the local and anatomical face models for the two steps of our method (Section 3.1).

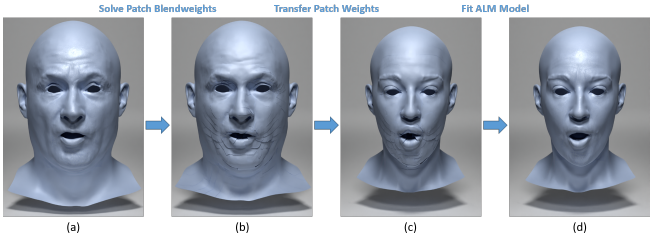

Fig. 2. Our method approximates a given source shape (a), with a collection of patch blendweights (b). These optimized patch blendweights are then transfered to the target model to perform an initial patch-wise retarget of the source shape (c). The result is further processed by an anatomical model to produce the final retargeted shape in high-fidelity (d).

3.1 Model Setup

In order to retarget performances from a source character to a target character, we require a small number of N 3D facial shapes for each character in semantic correspondence, similar to blendshape-based retargeting methods. In contrast to such methods, we require many fewer shapes (eg. all results in this paper are generated with $N = 20$ or fewer input shapes per character) to produce a high fidelity result. These face shapes can be sculpted by artists or scanned using multiview capture setups [Beeler et al. 2011; Fyffe et al. 2015]. Alternatively these shapes can also be created efficiently using automated techniques for rig creation [Carrigan et al. 2020; Li et al. 2010]. Let $\mathcal{S}$ be the set of source shapes, and $\mathcal{T}$ be the set of target shapes, such that $\mathcal{S}_i$ portrays the same expression as $\mathcal{T}_i$. Without loss of generality, let $\mathcal{S}_0$ and $\mathcal{T}_0$ be the neutral expressions. The sets $\mathcal{S}$ and $\mathcal{T}$ should be defined as triangle meshes at the origin of a common canonical coordinate frame. Fig. 3 illustrates a subset of the input shapes used in this paper, however other shape combinations would also be possible (see Section 4.4). In practice, good example shapes to use include extreme expressions like stretching the face wide open, compressing it tightly, smile, puffing air into the cheeks, mouth funneler (e.g. making a *shhh* sound), kiss, eyebrows up and down, and asymmetrical mouth and jaw movements both left and right.

Fig. 3. The local patch layout we use (left), and a subset of the input shapes (right) required for an exemplar source (top) and target (bottom) character.

Patch Blendshape Models. As our method operates at a local level, we divide $\mathcal{S}$ and $\mathcal{T}$ into a number of small spatial patches with overlapping boundaries. An example patch layout is shown in Fig. 3 (left), showing the patches without overlapping boundaries for better visualization. The exact size and distribution of the patches does not greatly affect the retargeting results, and we refer to Section 4.4 for an evaluation of different patch layouts and varying amounts of overlap. Note that the source and target meshes do not need to share the same topology, however we require a consistent mapping between the patches of the source and target models. This is easily achieved if the meshes do all share the same topology, or a UV layout. In other scenarios, this mapping can also be manually specified by an artist. Formally, let $\mathcal{P}$ be the set of all patches and $p \in \mathcal{P}$ represent an individual patch. All patches are modeled analogously in our work and we drop the index of a patch in our equations for brevity. Then $p^{\mathcal{S}}$ is the set of all local shapes in $\mathcal{S}$ for patch p. Given this parameterization, one can define a local patch blendshape model for the shape of the patch $\mathbf{X}_p^{\mathcal{S}}$ as

$$\mathbf{X}_p^{\mathcal{S}} = p^{\mathcal{S}_0} + \sum_{i=1}^{N-1} \alpha_{p,i}(p^{\mathcal{S}_i} - p^{\mathcal{S}_0}), \tag{1}$$

where $\alpha_p \in \mathcal{R}^{N-1}$ are patch specific coefficients used to linearly blend the patch blendshapes (defined as shape displacements between patch $p_i^{\mathcal{S}}$ and the neutral patch $p_0^{\mathcal{S}}$). In essence, a patch blendshape model is analogous to a global blendshape rig, except that each patch has it own set of blending coefficients, thereby leading to a model with greater expressiveness and more degrees of freedom. As the source and target shapes are in semantic correspondence and share the same patch layout, we can similarly define

$$\mathbf{X}_p^{\mathcal{T}} = p^{\mathcal{T}_0} + \sum_{i=1}^{N-1} \alpha_{p,i}(p^{\mathcal{T}_i} - p^{\mathcal{T}_0}), \tag{2}$$

for the target shapes. The patch blendshape models will be used in the first step of retargeting, described in Section 3.2.

Anatomical Local Face Model. In order to obtain high fidelity shapes for the target character face, we extend the Anatomical Local Model (ALM) proposed by Wu et al. [2016], which was designed for monocular face capture. A brief overview of the model is given below, however we refer to the original work for a more detailed description of the construction of the ALM model. In Section 3.3, we describe our novel use of the model for facial retargeting.

An anatomical local model is a character specific model that is built using a set of shapes of a given character. In contrast to blendshape rigs, the ALM model is a local model and is capable of modeling both skin deformations and the interaction of the skin (sliding, folding etc) with the underlying bone structure. Thus the ALM model benefits from locally deforming skin patches while preserving global consistency across these patches through anatomical constraints. Specifically, skin deformation is defined at a patch level, and complete face shapes are parameterized by two main components. The first is a set of patch deformation coefficients that define local patch stretching, bending and general non-rigid deformation in-place (devoid of rigid motion). The second is a set of rigid transformations for each patch. Together, these parameters combine to represent character specific facial expressions. Wu et al. [2016] propose a method to solve for these parameters to match a data term (e.g. monocular video) while respecting character specific anatomical constraints, that come in the form of skin thickness and sliding constraints over the character's rigid skull bone; the position of which is also solved for at the same time. To describe the model formally, we will re-use the same patch layout $\mathcal{P}$ as our patch blendshape models, although this is not a requirement. Let $\mathbf{X}_p$ represent the shape of patch p as defined by the ALM model as

$$\mathbf{X}_p = M_p \left(U_p + \sum_{k=1}^{K} w_{p,k} D_{p,k} \right), \tag{3}$$

where M_p is the rigid motion of the patch, U_p is the average patch shape over all K input shapes, and $D_{p,k}$ is the deformation subspace with corresponding weights $w_{p,k}$. In practice, we use the same N input shapes from the target character patch blendshape models to create the target ALM model (thus $K = N$), although this is not a requirement. An important distinction between the ALM model and the patch blendshape models is that the ALM model removes rigid motion from the patch shapes, such that the $D_{p,k}$ subspace models pure nonrigid deformation, while the rigid motion of the patch is included in the patch blend shape model. The reason for this difference will be made apparent in Section 3.2. As mentioned, the ALM model also consists of an anatomical subspace which is used to constrain the parameters of Eq. 3. These anatomical constraints relate a skin vertex v to a point on the underlying bone b_v, with a skin thickness constraint d_v, along normal direction n_v. More formally, these constraints are defined as follows

$$\tilde{\mathbf{b}}_v = \mathbf{b}_v^0 + \sum_{k=1}^{K} w_{p,k} (\mathbf{b}_v^k - \mathbf{b}_v^0), \tag{4}$$

$$\tilde{\mathbf{n}}_v \cong \mathbf{n}_v^0 + \sum_{k=1}^{K} w_{p,k} (\mathbf{n}_v^k - \mathbf{n}_v^0), \tag{5}$$

$$d_v = d_v^0 + \sum_{k=1}^{K} w_{p,k} (d_v^k - d_v^0). \tag{6}$$

In the equations above, indices p, k continue to refer to the patch and shape indices respectively, and each estimated component (bone point $\tilde{\mathbf{b}}_v$, normal $\tilde{\mathbf{n}}_v$ and skin thickness d_v) are defined in corresponding subspaces to the patch deformation subspace (Eq. 3) such that the subspace weights $w_{p,k}$ semantically correspond. As such, Eq. 4 defines a bone subspace, which supports the sliding of skin over the bone when certain shape weights $w_{p,k}$ are activated. Similarly, Eq. 5 and Eq. 6 are skin normal and thickness subspaces, respectively. Solving for an ALM face pose generally means to solve for M_p and the set of $\{w_{p,k}\}$ for each patch p and then cleverly stitch the patches together. Note that in our work, the ALM model is created only for the target character, as it will be used to anatomically constrain the retargeted performance (Section 3.3). We also only use anatomical constraints derived from the skull bone, which can be automatically fit to the target character using the method of Beeler and Bradley [2014] with minimal overhead. Again, please refer to Wu et al. [2016] for more details.

3.2 Patch-wise Retargeting

We now describe how we use the patch blendshape models defined in Section 3.1 to obtain an initial estimate of the retargeted performance from a source to a target character. We assume we are given patch blendshape models corresponding to both characters, and that we will process the frames of the source performance individually. Let us denote the current source performance shape that is to be retargeted as $\mathbf{X}^{S'}$. At a high level, we approach the problem by estimating the coefficients α of all the patches of the source model (Eq. 1) that can accurately describe the local skin deformations required to match the shape $\mathbf{X}^{S'}$. We then transfer these coefficients to the target model (Eq. 2) to obtain an estimate of the retargeted expression. During this process, we will add several methods to artistically control the result. The resulting per-patch deformations of the target model will be passed on to the final step in Section 3.3.

To solve for the coefficients α that best fit the source patch blendshape model to $\mathbf{X}^{S'}$ we employ a least squares optimization, defined by the following fitting energy

$$E_{Fit} = \sum_{p \in \mathcal{P}} (\mathbf{X}_p^{S'} - R\mathbf{X}_p^{\mathcal{S}}). \tag{7}$$

Here, $\mathbf{X}_p^{\mathcal{S}}$ is the source model defined in Eq. 1, and R is a global rigid transformation for the entire model. We include this transformation to accommodate a practcal scenario where the shape $\mathbf{X}^{S'}$ may come from a facial performance capture system and may not lie at the canonical origin. While methods for removing rigid head motion from performances do exist [Beeler and Bradley 2014], these techniques are not perfect and there is often some residual rigid motion remaining, which we account for by optimizing R along with the shape coefficients α.

Since blindly optimizing for patch coefficients α_p is an underconstrained problem, we further regularize the patch coefficients to remain close to zero as

$$E_{Reg} = \sum_{p \in P} \sum_{i=0}^{N-1} (\alpha_{p,i})^2, \tag{8}$$

and to stay consistent across adjacent patches with an overlap energy, defined as

$$E_O = \sum_{p \in P} \sum_{q \in \mathcal{N}(p)} \sum_{i=0}^{N-1} (\alpha_{p,i} - \alpha_{q,i}), \tag{9}$$

where $\mathcal{N}(p)$ defines the patches neighboring p. The final energy for fitting our patch blendshape model to a source shape $\mathbf{X}^{S'}$ is the weighted sum of these energies,

$$E_{PBS} = \lambda_{Fit} E_{Fit} + \lambda_{Reg} E_{Reg} + \lambda_O E_O. \tag{10}$$

The result of fitting the patch-wise blendshapes to a source performance shape is illustrated in Fig. 2 (a) and (b). An important distinction of our patch-wise retargeting model in comparison to the ALM model is the presence of rigid motion in our patch blendshapes. As an ALM model optimization solves for both per-patch rigid transformations and deformation coefficients, a least squares solver prefers to explain as much of skin deformation as possible using the rigid transform and only dials in the patch coefficients when necessary. While this property may be beneficial in the context of face tracking, it extends poorly to retargeting, as transferring rigid transformations of source patches to the target is undesirable due to differences in scale and range of motions between the two characters. Therefore, by not separating the rigid motion from the blendshapes in the patch-wise retargeting step, we expect the patch coefficients to explain all of the skin deformation, which translates into better, more character-specific expression transfer during retargeting. In Section 4.4 we will show a visual comparison of transferring ALM coefficients (rigid transform + patch coefficients) from the source to the target character vs. our local transfer, where our method clearly outperforms the ALM transfer (refer to Fig. 16).

Artistic Control. In practice, we found that it is beneficial to allow some level of artistic control over the patch fitting process. For example, even simply increasing or decreasing the "strength" of the retergeting can be powerful, which is easily accomplished by post-multiplying the resulting α coefficients by a user-defined scalar value. Importantly, as our method is local, we can support a spatially-varying strength control parameter, where the retarget strength of each local patch can be individually specified (typically accomplished through a texture map lookup). An additional way to add user control is to allow artists to provide blendshape preferences, with a per-shape preference weight γ_i, where $0 \leq i \leq N-1$ and $0 \leq \gamma_i \leq 1$. By default, $\gamma_i = 1$ for all i, but this preference parameter allows to penalize the use of certain shapes by setting the corresponding γ_i to a value below 1, or favor a shape by setting all other shape values to less than 1. Again, in practice we can even allow spatially-varying shape preferences, different for each patch p, and thus the preference weight is formally defined as $\gamma_{p,i}$. To incorporate the shape preferences in our optimization, we modify the source model from Eq. 1 to be

$$\mathbf{X}_p^S = p^{S_0} + \sum_{i=1}^{N-1} \alpha_{p,i} \cdot \gamma_{p,i} (p^{S_i} - p^{S_0}), \tag{11}$$

effectively scaling the blendshapes by the user preference values. This has the effect that when a user preference is less than 1, the corresponding blendshape is scaled closer to the neutral shape, and the system must use a higher corresponding $\alpha_{p,i}$, contradicting Eq. 8, and so if possible a different combination of shapes to achieve the same goal will be chosen by the optimization instead. In the end, we transfer the weighted shape coefficients $\alpha_{p,i} \cdot \gamma_{p,i}$ to the target blendshape model to account for the scaling during optimization.

The result of the patch-wise retargeting is a set of deformed target patch shapes $\mathbf{X}_p^{T'}$ for all p, which approximate the desired target shape corresponding to the source input shape (Fig. 2 (c)). For all experiments reported in this paper, unless explicitly mentioned, we set λ_{Fit} to 1, λ_O to 100 and λ_{Reg} to 35.

3.3 Anatomical Reconstruction

In the previous section we described the main retargeting procedure, which transfers per-patch deformations from the source to the target character. Now, the goal is to convert the patch-wise retargets into a globally consistent target face shape (Fig. 2 (d)). Even though we aim to obtain spatial consistency in the per-patch deformations (via the overlap regularizer in Eq. 9), there will inevitably be discontinuities at the patch boundaries. For this reason, we employ the character specific ALM model of Wu et al. [2016] described in Section 3.1 to provide the final target shape.

Following Wu et al. [2016], this is achieved by solving for the model parameters M_p and $\{w_{p,k}\}$ in Eq. 3 in another optimization. Contrary to Wu et al. [2016], who formulate the optimization to match a data term coming from monocular video, we instead formulate a new data term from the patch-wise retargets, formulated in 3D space. Specifically, we create an alternate data term as

$$E_D = \lambda_D \sum_{v \in \mathcal{V}} \sum_{p \in P(v)} (\mathbf{X}_p(v) - \mathbf{X}_p^{T'}(v)), \tag{12}$$

where $\mathcal{V}$ is the set of all vertices in the target character mesh, $P(v)$ denotes all patches that contain vertex v, $\mathbf{X}_p(v)$ is the ALM model (Eq. 3) evaluated at v, and $\mathbf{X}_p^{T'}(v)$ are the retargeted patch shapes evaluated at v. Using this new data term in the model fitting procedure of Wu et al. [2016] (combined with their standard anatomical and overlap constraints), we obtain the final retargeted shape, as illustrated in Fig. 2 (d). Note that we do not solve for the skull position in the ALM model, but instead fix it in space since we wish to perform the retargeting in a canonical space. While estimating the parameters of the ALM model, we set λ_D to 10, the weight for the anatomical constraint from Wu et al. [2016] λ_{A1} to 10 and their overlapping constraint weight λ_O to 0.85.

4 RESULTS AND EVALUATION

We now present the results of our facial retargeting method and compare it to alternatives. The dataset we use for evaluation consists of several performance sequences of different actors captured using

a production capture system[2] based on Wu et al. [2016], plus a single hand-crafted fantasy creature.

4.1 Qualitative Results

We start by showcasing several qualitative examples of our local, anatomically constrained retargeting technique in Fig. 1 and Fig. 4. Our method is successfully able to retarget a wide variety of facial performances, ranging from dialogues, emotions, and facial workouts from a variety of source characters to a range of target characters. Each result captures the subtle facial deformations of the source character, without altering the target's identity. As such, our method can be an invaluable tool for facial animation and retargeting in visual effects and high-end applications. We further highlight the flexibility of our method by retargeting performances from human characters to a target fantasy creature in Fig. 5, achieved with the same algorithm and no additional parameter tuning. We kindly refer you to our supplemental video for more results.

4.2 Comparisons with Existing Techniques

We now compare the performance of our approach with that of common methods used in the industry today. Specifically, we will compare to global blendshape-based retargeting [Lewis et al. 2014], deformation transfer [Sumner and Popović 2004] and simple delta transfer, as described in Section 2.5. For the global blendshape model, we use the same 20 shapes as our approach for a fair comparison (later we will also compare to a large 236-shape blendshape rig similar to what is used in production settings).

A qualitative comparison is provided in Fig. 7, where four different source expressions are transferred to four different target characters, using each of the methods. Both deformation transfer and delta transfer tend to generate unrealistic shapes in the eyes and mouth regions, especially when the target character is more dissimilar in shape from the source character. Both methods also incorrectly transfer the wrinkle details from the source character to the target (e.g. the forehead in row 2). As well, sometimes deformation transfer suffers from geometric artifacts (e.g. the eye region in rows 1 and 4). The 20-shape global blendshape model does not have enough expressiveness to reach the necessary facial deformations, resulting in the loss of the intended expression as seen by the closing eyes in row 1, and the changed mouth expression in row 4. The supplemental video shows that the global blendshape model also has problems with temporal stability. In contrast, our method produces expressive, stable and artifact-free retargets.

We also provide a quantitative evaluation of the methods. This is achieved by leaving out 4 of the 20 shapes and building retargeting models from the 16 remaining shapes, and then evaluating the retargeted result on the 4 validation shapes. Results are shown in Fig. 8. Starting with an open mouth expression (rows 1 and 2), we retarget from two different source characters to the same target character. In the ideal case, the resulting target shape would be the same, independent of the source character. Notice the large differences in the result for the deformation and delta transfer methods. Also, evaluating the per-vertex error with respect to the held out ground truth shape via the heat map shows that our method achieve the most accurate results. Row 3 of Fig. 8 adds a second held-out expression, retargeted from the same source character as row 2, and again our method produces the most accurate result.

As a final comparison, we demonstrate that our method can achieve quality on par with large-scale blendshape rigs often used in production, but with far fewer input shapes. To this end, we employed a publicly-available model containing 236 shapes[3], which we mapped to our own characters for comparison. Fig. 6 shows retargeting results for two different source-target pairs. Visually, the results using the larger blendshape rig are naturally more appealing than the 20-shape rig, and our results are comparable to the production rig results while using only 20 input shapes.

4.3 User study

In order to further compare our approach with common methods used in practice, we performed a user study to gain insight into the best retargeting method and the approach that generates the most realistic facial animation overall. The study consisted of several examples of 3D human source characters being retargeted to 3D human target characters. We simultaneously showed participants the results of our method, 20-shape global blendshape retargeting, deformation transfer and delta transfer. As we will illustrate in this section, our proposed technique clearly outperforms the others both in terms of retargeting accuracy and animation realism.

We performed the study on both static expression retargets as well as dynamic retargeted performances. The static expressions allowed users to take their time and analyze nuances in the resulting shapes, while the performance animations provided a more holistic view of the retargeting quality. 10 different source/target retargeting examples were shown for each of the individual expressions and performance animations, spanning 5 different source characters and 6 additional target characters (source characters were never used as target characters and vice-versa). The performances ranged from dialog speech to fast facial expression transitions.

An example frame from the user study is shown in Fig. 9. For the static expressions, the users were asked: *Which of the Target expressions (A,B,C, or D in blue) is the best retargeting of the Source expression (in blue) to the Target identity? Please refer to example expressions of the Target in gray, to help understand their identity.* To help the users understand the identity and expressions of the target character, six ground truth expressions were shown at the bottom of the screen (in gray). The order of the four results was randomized for each example. The dynamic performances were presented in the same manner, and the users were asked two questions, *Which of the Target performances (A,B,C, or D in blue) is the best retargeting of the Source performance (in blue) to the Target identity? Please refer to example expressions of the Target in gray, to help understand their identity*, and as well: *Which Target performance looks overall the most realistic?* Participants were allowed to choose more than one answer if they could not decide. 45 participants from various backgrounds took part in the survey (45% were not familiar with computer graphics, 44% were experienced in graphics but not in retargeting methods, and 11% were familiar with retargeting). The results of the user study are illustrated in Fig. 10 and Table 1. Fig. 10

[2] https://studios.disneyresearch.com/anyma/

[3] www.eisko.com/louise/virtual-model

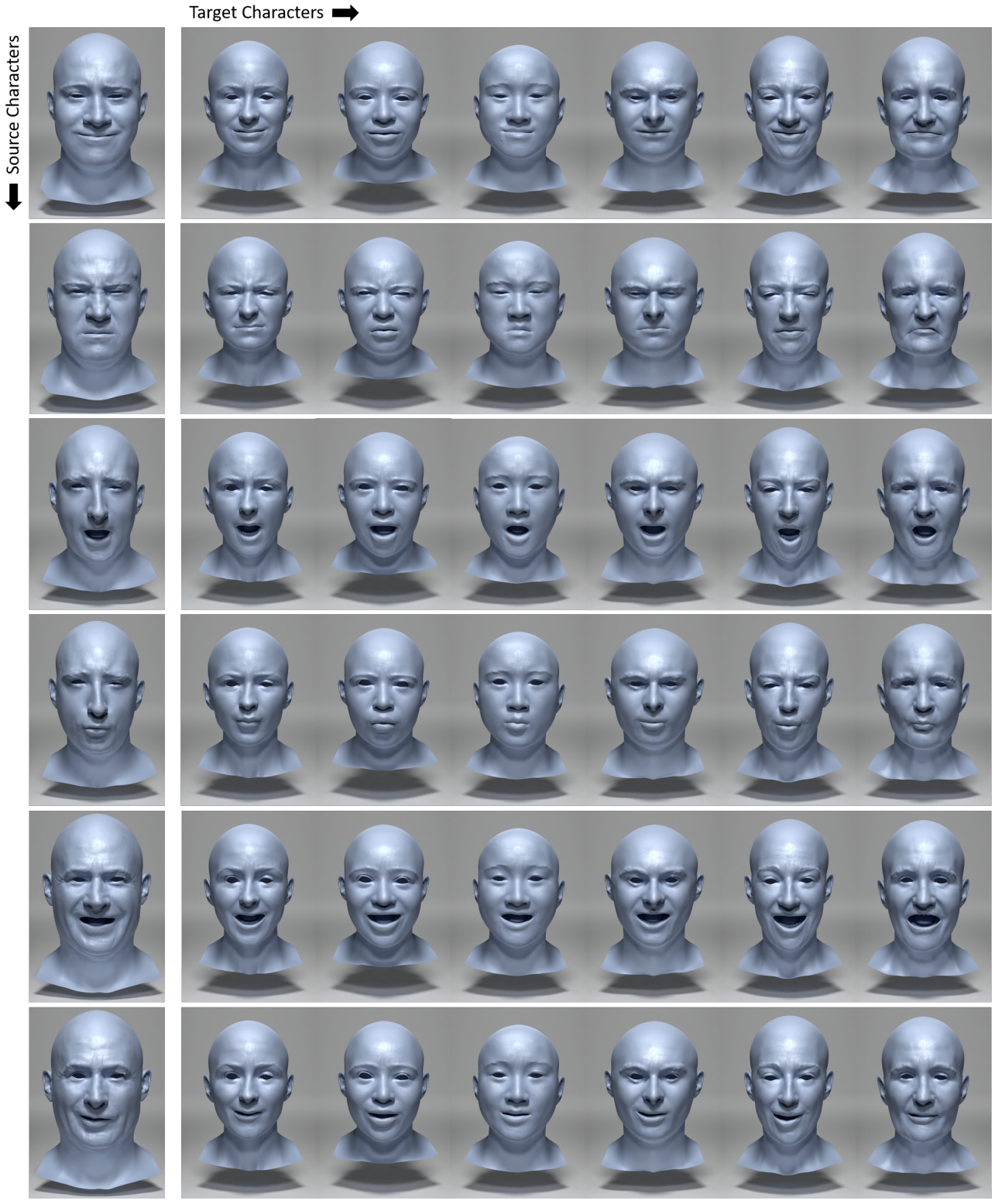

Fig. 4. Several examples of high quality facial performance retargeting obtained by our method. Each row corresponds to a unique source performance and each column is a unique target character. Our method is consistently able to output convincing performances with a high degree of realism while staying faithful to the target character's facial anatomy.

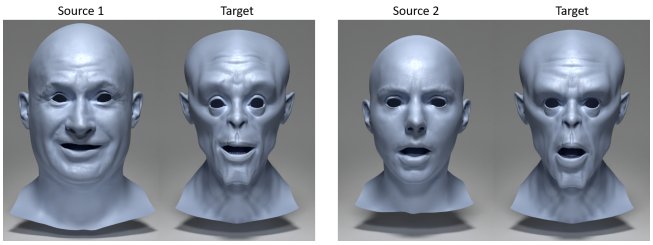

Fig. 5. We highlight the flexibility of our method by retargeting performances from several human characters to a fantasy creature.

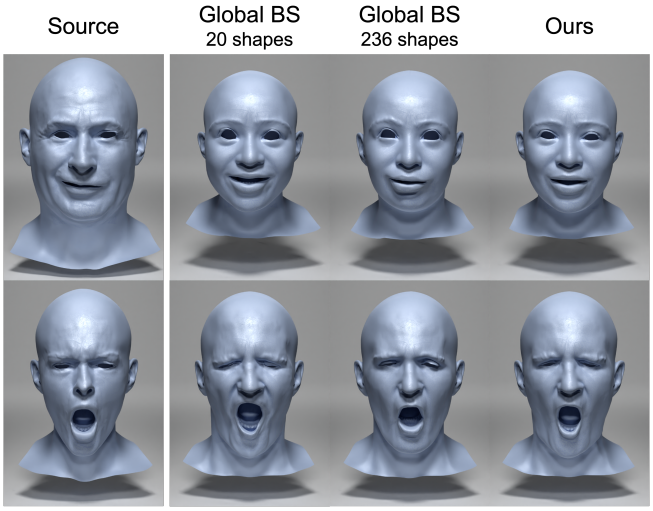

Fig. 6. Our results (using only 20 input shapes) are comparable to large-scale production blendshape rigs containing hundreds of shapes. The 20-shape global blendshape result is included for comparison.

tallies the total number of votes for each method over all 10 static examples and all 10 animations, separated by question. As can be clearly seen, our method (blue) was the most popular choice for all categories. Table 1 additionally shows the sum over the set of retargeting examples where each method was the chosen winner, per question. Again, the proposed technique was clearly a favorite, independent of the source/target character pair, independent of static expressions versus dynamic performances, and across all questions. Interestingly, the users were able to identify that global blendshape retargeting (with so few shapes) is unsuitable for high quality animations, as this method was least preferred. Deformation transfer and delta transfer showed a similar performance, likely owing to their similar algorithmic nature.

4.4 Ablation Studies and Evaluations

We now show the effects when varying certain parameters of our method, starting with its dependence on the patch layout. Fig. 11 shows the effects of varying both the number and the layout of the patches . For this experiment, we held out a subset of validation shapes from both the source and target models, and compared

Table 1. Number of retargeting examples where the method was chosen as top performing (out of 10). Note that ties are counted twice.

Method	Expressions (best retarget)	Animations (best retarget)	Animations (most real)
Global Blendshapes	1	0	0
Deformation Transfer	1	1	1
Delta Transfer	1	1	2
Ours	**7**	**9**	**8**

the reconstruction accuracy for one of the held-out shapes under different configurations. As indicated by the error maps, accuracy decreases when there are too few patches (last two layouts), but for a sufficient density of patches the exact layout has little effect (first two layouts). All our results are created with the first layout.

A second parameter that is user-controllable is the amount of overlap between patches, defined by the number of closed vertex rings in the mesh connectivity. Fig. 12 illustrates the effect of different overlap values during retargeting. The quality of the patch retargeting step is severely degraded with too little overlap, while too much overlap it results in over-smoothed shapes. In all our results, we use 6 overlap rings. Furthermore, the weight for the overlap consistency term λ_O in Eq. 10 also has an effect on the results, as we illustrate in Fig. 13. The first row shows the patch fit results to a source shape with corresponding error maps for λ_O values of 0, 25, 100 and 500. Rows two and three illustrate retargeting results to two different characters using local blendshape transfer only (without the anatomical reconstruction step). When the overlap weight is very small, individual patches fit the source shape better but the patches are extremely disconnected. When the overlap weight is very high, the local patches align almost perfectly, but at the cost of losing expression fidelity to the source shape. A good tradeoff is found around $\lambda_O = 100$; the value we use for all of our results.

Our local model for retargeting is naturally more expressive than a global blendshape rig, given the same number of input shapes. In Fig. 14, we show the accuracy of reconstructing a source shape using a global blendshape model in comparison to our local model while varying the number of shapes in the models. The ground truth that is being evaluated was left out of both models. As can be seen in the heatmaps and in the accompanying plot, our method with only as few as 7 shapes still significantly outperforms a traditional blendshape rig with as many as 19 shapes. Continuing the evaluation of input shape cardinality, Fig. 15 illustrates how our retargeting pipeline degrades with fewer and fewer input shapes . As shown, reducing from 20 to 15 shapes introduces only a small error, which becomes increasingly larger with fewer shapes. For this experiment, we progressively removed shapes so as to keep the most overall facial deformation within the given shape budget.

We also evaluate our decision to embed the rigid motion of the patches into the blendshapes, in contrast to separating the rigid and non-rigid components as Wu et al. [2016] . As described in Section 3.2, separating the rigid motion is undesirable due to differences in scale and range of motions between characters. Fig. 16 illustrates this issue on a retargeted character obtained in two different ways, one where the rigid motion is separated from the blendshapes and ours, which leads to fewer artifacts.

Fig. 7. We present qualitative comparisons of the proposed method against commonly used facial retargeting techniques in production. The source subject is shown in the first column and the retargeted expression for a unique target character is shown in each row. Our method clearly produces the most expressive, yet artifact free retargeting in all cases. Kindly refer to our supplemental video for additional comparisons.

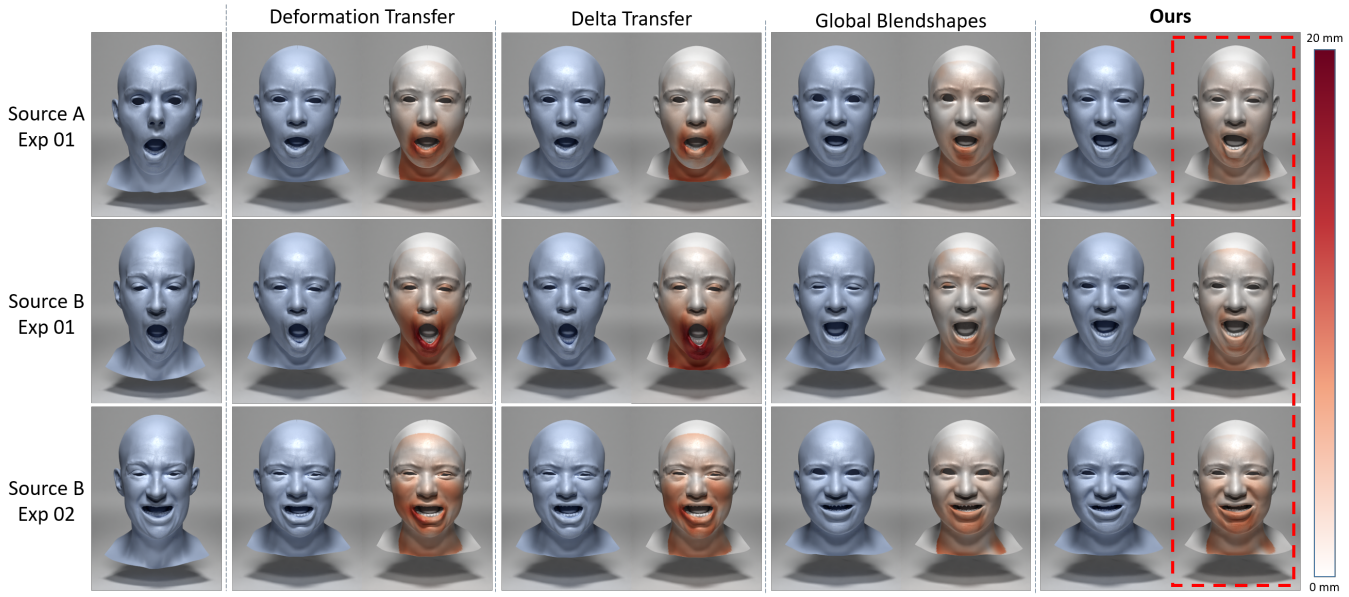

Fig. 8. Quantitative Comparison of the proposed method against common approaches, achieved by leaving a subset of shapes out of the models for validation. Rows 1 and 2 show one of the held-out expressions retargeted from two different source characters to the same target character. Row 3 adds a new held-out expression from the same source character as row 2.

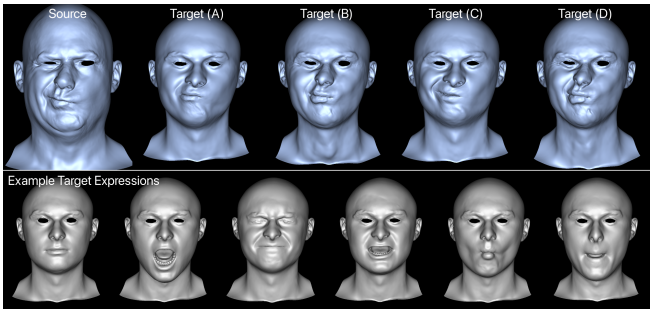

Fig. 9. Example frame from the user study, showing the source expression and resulting target expressions from the different methods (top in blue) as well as example real target expressions for guidance (bottom in gray).

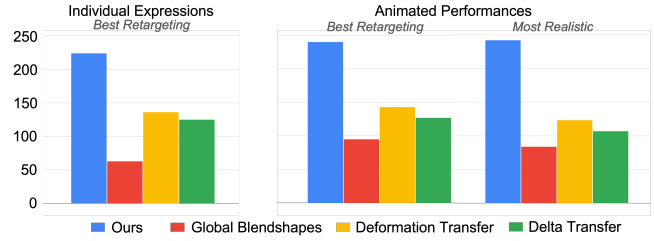

Fig. 10. We performed a user study to compare our method (blue) with global blendshape retargeting (red), deformation transfer (yellow) and delta transfer (green), on both individual expression transfers (left) and animated performance retargets (right) of multiple source and target pairs.

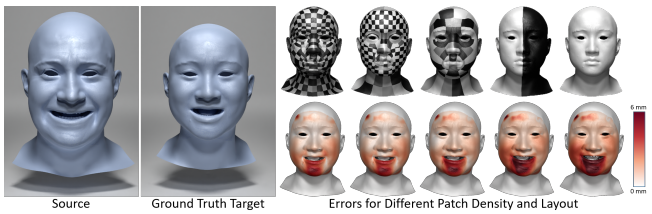

Fig. 11. Here we compare different number and layout of local patches for our method on a left out shape. As indicated by the error maps, accuracy suffers when there are too few patches (last two layouts), but for a sufficient density of patches the exact layout has little effect (first two layouts).

4.5 Run time analysis

Our two step retargeting technique takes 1 min per frame in total on a standard desktop CPU with an Intel(R) Core(TM) i7-7700K processor and 32GB of RAM. This run time was measured while using 20 blendshapes, and 400 patches, with each shape having 95,000 vertices. A bulk of this time (almost 90%) is spent in optimizing the ALM model [Wu et al. 2016] for anatomical reconstruciton, where modern GPU solvers [Fratarcangeli et al. 2020] could offer substantial speed ups. Our method is trivially parallelizable across frames and works seamlessly on performance data without temporal regularization. All results in our paper were produced using a CPU based non-linear least squares solver [Agarwal et al. 2010].

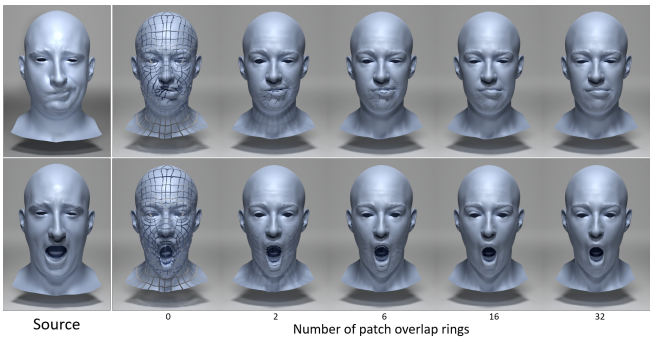

Fig. 12. The amount of overlap between the local patches (in vertex rings) affects the quality of the retarget. Too little overlap results in larger discontinuities between patches, and too much overlap results in oversmooth retargets. We use 6 overlap rings.

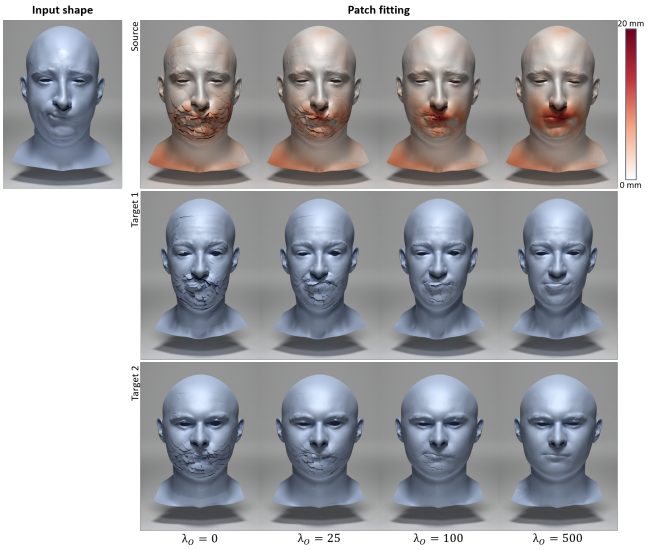

Fig. 13. The patch overlap weight λ_O affects the quality of the retarget. Too small and the patches are very disconnected. Too large and the desired source expression is compromised. We use $\lambda_O = 100$.

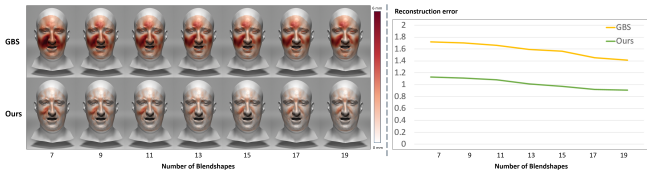

Fig. 14. We show the effect of varying the number of input shapes with our approach vs. a standard global blendshape rig on a source shape reconstruction task. Our model achieves a lower reconstruction error with only 7 shapes than what global blendshapes achieves with 19 shapes.

4.6 Artistic Manipulation

In addition to providing high fidelty results, our method allows a certain amount of artistic control. We first demonstrate the user-definable *retargeting strength* map, which is a spatially-varying

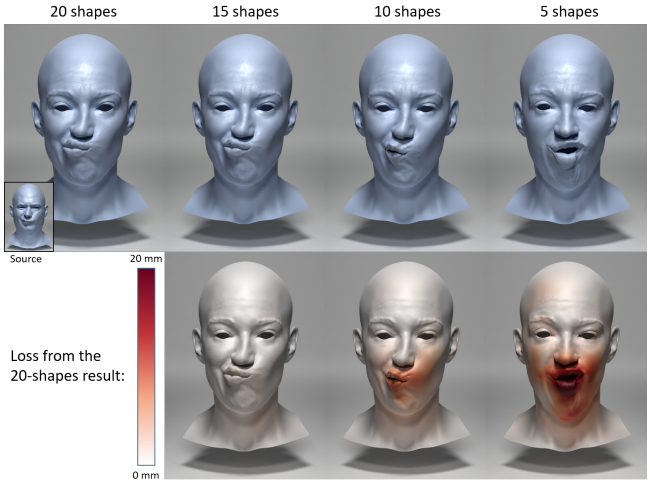

Fig. 15. We show a retargeting result using models with varying input shapes, starting from our usual 20 shapes down to 15, 10 and 5 (top row). The error as compared to the 20-shape result (bottom row) indicates that the quality of the retarget degrades naturally with fewer input shapes.

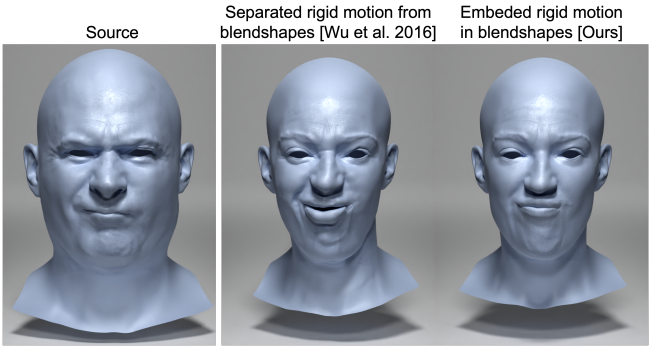

Fig. 16. Separating the rigid motion from the local blendshapes leads to artifacts during retargeting (center), compared to our approach of embedding the rigid motion into the patch blendshapes (right).

scalar value that increases or decreases the expressiveness of the retarget. Fig. 17 illustrates using this strength map in extreme situations, like retargeting to only parts of the face as well as exaggerating the result with a strength of 1.5.

Additionally, although our model makes use of only a handful of shapes for retargeting, it can certainly leverage additional blendshapes when being incorporated into an existing workflow. When extra corresponding shapes between the source and target characters are available (we refer to these as calibration shapes), instead of naively including them into the model, which might further underconstrain the retarget and increase solve times, we propose a simple calibration step that optimizes for a spatially varying weight map (akin to the retargeting strength map defined above), given a source and target model. For each source calibration shape, when we fit the source model to obtain a collection of patch coefficients α which when transfered over to the target model, should ideally

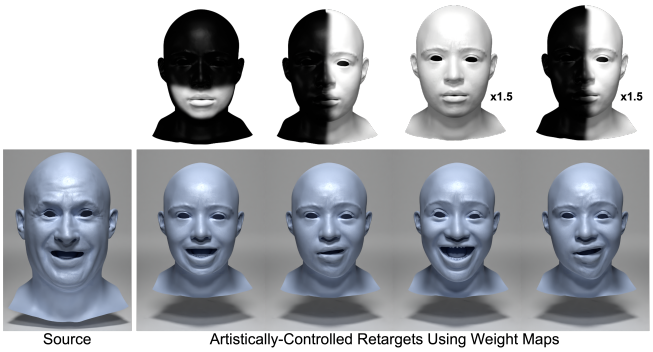

Fig. 17. A retargeting strength map (top) can be painted to spatially control the retargeting result (bottom). For columns 4 and 5, the patch blendweights are amplified by a factor of 1.5 in the masked regions.

produce the corresponding target calibration shape. Based on this insight, we optimize for a per-patch retargeting strength scalar that re-weights the patch coefficients in a spatially varying manner, such that the difference between the transferred target shape and the true target calibration shape is minimized. Once such a map is optimized for, it essentially remaps the patch coefficients during subsequent retargeting, to respect the target manifold better. In Fig. 18, we show the spatially varying retarget strength map resulting from such a calibration, between a source character (left) and two target characters (right half, first and second row). The spatially varying weight map is applied before subsequent retargeting analogous to the weightmap applied in Fig. 17. The calibration step can introduce subtle variations in the retargeted performance as visualized in the heatmap in Fig. 18. The optimized weight map can also serve as a starting point for artists to achieve interesting retargeting effects.

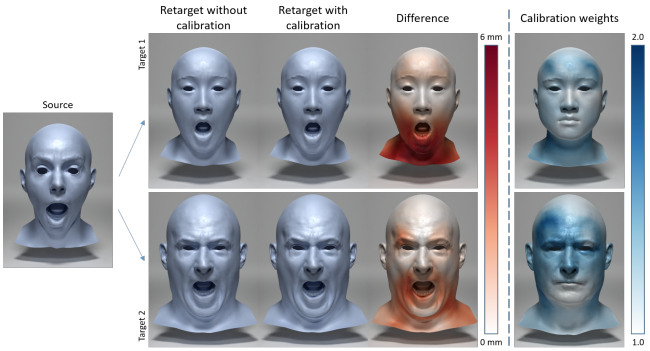

Fig. 18. We show the effect of calibrating a weight map between a source character (left) and two target characters (two rows on the right)

5 LIMITATIONS

While our method only requires a small number of shapes for each character, we do require that these shapes are in semantic correspondence and creating such shapes through capture or sculpting requires time and effort from artists. This is a problem that we do not address in this work and believe that techniques like Li et al. [2010] can mitigate to a certain extent. In the absence of a shared topology, a predefined mapping between the source and target patches needs to also be provided. Finally, though our method provides several ways for artists to intuitively control the retargeted result, these edits (Fig. 12, Fig. 17) require re-solving the entire sequence.

6 CONCLUSION

In conclusion, we present a local anatomically constrained method for high fidelity facial performance retargeting that is ready for use in demanding production pipelines. Our offline algorithm leverages the expressive power of local blendshape rigs to obtain an initial estimate of the retargeted performance. Then in a second step, an anatomical model built using the target character's facial geometry is used to constrain the retargeted performance to an anatomically plausible subspace. The result is a powerful method that can perform highly realistic retargeting given only a handful of shapes in correspondence (20 shapes) when compared to full blown production rigs with hundreds of shapes. Our method additionally allows artists to control several aspects of the retargeted performance in order to achieve the perfect look for their animation. We demonstrate several benefits of our method with a detailed user study. We hope that our new tool benefits animators that spend innumerable hours in producing a realistic facial animation.

ACKNOWLEDGMENTS

We wish to thank Maurizio Nitti and Doriano van Essen for creating the creature character in Fig. 1 and Fig. 5.

REFERENCES

Kfir Aberman, Peizhuo Li, Dani Lischinski, Olga Sorkine-Hornung, Daniel Cohen-Or, and Baoquan Chen. 2020. Skeleton-Aware Networks for Deep Motion Retargeting. *ACM Trans. Graphics (Proc. SIGGRAPH)* 39, 4 (2020).

Sameer Agarwal, Keir Mierle, and Others. 2010. Ceres Solver. http://ceres-solver.org.

Deepali Aneja, Bindita Chaudhuri, Alex Colburn, Gary Faigin, Linda Shapiro, and Barbara Mones. 2018. Learning to Generate 3D Stylized Character Expressions from Humans. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 160–169.

Dafni Antotsiou, Guillermo Garcia-Hernando, and Tae-Kyun Kim. 2018. Task-Oriented Hand Motion Retargeting for Dexterous Manipulation Imitation. In *ECCV Hands Workshop*.

Ilya Baran, Daniel Vlasic, Eitan Grinspun, and Jovan Popović. 2009. Semantic Deformation Transfer. *ACM Trans. Graphics (Proc. SIGGRAPH)* 28, 3, Article 36 (2009).

Thabo Beeler and Derek Bradley. 2014. Rigid Stabilization of Facial Expressions. *ACM Trans. Graphics (Proc. SIGGRAPH)* 33, 4, Article 44 (jul 2014).

Thabo Beeler, Fabian Hahn, Derek Bradley, Bernd Bickel, Paul Beardsley, Craig Gotsman, Robert W. Sumner, and Markus Gross. 2011. High-quality passive facial performance capture using anchor frames. *ACM Trans. Graphics (Proc. SIGGRAPH)* 30, Article 75 (2011). Issue 4.

Kiran S. Bhat, Rony Goldenthal, Yuting Ye, Ronald Mallet, and Michael Koperwas. 2013. High Fidelity Facial Animation Capture and Retargeting with Contours. In *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. Association for Computing Machinery, 7–14.

Mazen Al Borno, Ludovic Righetti, Michael J. Black, Scott L. Delp, Eugene Fiume, and Javier Romero. 2018. Robust Physics-based Motion Retargeting with Realistic Body Shapes. In *ACM / Eurographics Symposium on Computer Animation*.

Sofien Bouaziz and Mark Pauly. 2014. Semi-Supervised Facial Animation Retargeting.

Sofien Bouaziz, Yangang Wang, and Mark Pauly. 2013. Online Modeling for Realtime Facial Animation. *ACM Trans. Graphics (Proc. SIGGRAPH)* 32, 4, Article 40 (2013).

Chen Cao, Qiming Hou, and Kun Zhou. 2014. Displaced Dynamic Expression Regression for Real-Time Facial Tracking and Animation. *ACM Trans. Graphics (Proc. SIGGRAPH)* 33, 4, Article 43 (2014).

Emma Carrigan, Eduard Zell, Cédric Guiard, and Rachel McDonnell. 2020. Expression Packing: As-Few-As-Possible Training Expressions for Blendshape Transfer. *Computer Graphics Forum* 39 (2020).

P. Chandran, D. Bradley, M. Gross, and T. Beeler. 2020. Semantic Deep Face Models. In *2020 International Conference on 3D Vision (3DV)*. IEEE Computer Society, 345–354.

Bindita Chaudhuri, Noranart Vesdapunt, and Baoyuan Wang. 2019. Joint face detection and facial motion retargeting for multiple faces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9719–9728.

Lele Chen, Chen Cao, Fernando De la Torre, Jason M. Saragih, Chenliang Xu, and Yaser Sheikh. 2021. High-Fidelity Face Tracking for AR/VR via Deep Lighting Adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*. Computer Vision Foundation / IEEE, 13059–13069.

Renwang Chen, Xuanhong Chen, Bingbing Ni, and Yanhao Ge. 2020a. *SimSwap: An Efficient Framework For High Fidelity Face Swapping*. Association for Computing Machinery, 2003–2011.

Renwang Chen, Xuanhong Chen, Bingbing Ni, and Yanhao Ge. 2020b. SimSwap: An Efficient Framework For High Fidelity Face Swapping. In *MM '20: The 28th ACM International Conference on Multimedia*. ACM, 2003–2011.

Erika Chuang and Christoph Bregler. 2002. Performance Driven Facial Animation using Blendshape Interpolation. *Computer Science Technical Report, Stanford University* 2 (01 2002).

Timothy Costigan, Mukta Prasad, and Rachel McDonnell. 2014. Facial Retargeting Using Neural Networks. Association for Computing Machinery, 31–38.

Ludovic Dutreve, Alexandre Meyer, and Saïda Bouakaz. 2008. Feature Points Based Facial Animation Retargeting. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*. Association for Computing Machinery, 197–200.

Marco Fratarcangeli, Derek Bradley, Aurel Gruber, Gaspard Zoss, and Thabo Beeler. 2020. Fast Nonlinear Least Squares Optimization of Large-Scale Semi-Sparse Problems. *Computer Graphics Forum* (2020).

Graham Fyffe, Andrew Jones, Oleg Alexander, Ryosuke Ichikari, and Paul Debevec. 2015. Driving High-Resolution Facial Scans with Video Performance Capture. *ACM Trans. Graphics (Proc. SIGGRAPH)* 34, 1, Article 8 (2015).

Pablo Garrido, Levi Valgaerts, Ole Rehmsen, Thorsten Thormählen, Patrick Pérez, and Christian Theobalt. 2014. Automatic Face Reenactment. *2014 IEEE Conference on Computer Vision and Pattern Recognition* (2014), 4217–4224.

Chunbao Ge, Yiqiang Chen, Changshui Yang, Baocai Yin, and W. Gao. 2005. Motion Retargeting for the Hand Gesture. In *WSCG*.

Yang Hong, Bo Peng, Haiyao Xiao, Ligang Liu, and Juyong Zhang. 2021. HeadNeRF: A Real-time NeRF-based Parametric Head Model. *CoRR* abs/2112.05637 (2021). arXiv:2112.05637

Tero Karras, Timo Aila, Samuli Laine, Antti Herva, and Jaakko Lehtinen. 2017. Audio-Driven Facial Animation by Joint End-to-End Learning of Pose and Emotion. *ACM Trans. Graph.* 36, 4, Article 94 (2017).

Hyeongwoo Kim, Mohamed Elgharib, Hans-Peter Zollöfer, Michael Seidel, Thabo Beeler, Christian Richardt, and Christian Theobalt. 2019. Neural Style-Preserving Visual Dubbing. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 178:1–13.

Paul Hyunjin Kim, Yeongho Seol, Jaewon Song, and Junyong Noh. 2011. Facial Retargeting by Adding Supplemental Blendshapes. In *Pacific Graphics Short Papers*, Bing-Yu Chen, Jan Kautz, Tong-Yee Lee, and Ming C. Lin (Eds.). The Eurographics Association.

Seonghyeon Kim, Sunjin Jung, Kwanggyoon Seo, Roger Blanco i Ribera, and Junyong Noh. 2021. Deep Learning-Based Unsupervised Human Facial Retargeting. *Computer Graphics Forum* 40, 7 (2021), 45–55.

J. P. Lewis, K. Anjyo, Taehyun Rhee, M. Zhang, Frédéric H. Pighin, and Z. Deng. 2014. Practice and Theory of Blendshape Facial Models. In *Computer Graphics Forum (Proc. Eurographics)*.

Hao Li, Thibaut Weise, and Mark Pauly. 2010. Example-Based Facial Rigging. *ACM Trans. Graphics (Proc. SIGGRAPH)* 29, 3 (July 2010).

Hao Li, Jihun Yu, Yuting Ye, and Chris Bregler. 2013. Realtime Facial Animation with On-the-Fly Correctives. *ACM Trans. Graphics (Proc. SIGGRAPH)* 32, 4, Article 42 (2013).

Tianye Li, Timo Bolkart, Michael J. Black, Hao Li, and Javier Romero. 2017. Learning a Model of Facial Shape and Expression from 4D Scans. *ACM Trans. Graph.* 36, 6, Article 194 (2017).

Ko-Yun Liu, Wan-Chun Ma, Chun-Fa Chang, Chuan-Chang Wang, and Paul Debevec. 2011. A framework for locally retargeting and rendering facial performance. *Computer Animation and Virtual Worlds* 22, 2-3 (2011), 159–167.

Stephen Lombardi, Jason Saragih, Tomas Simon, and Yaser Sheikh. 2018. Deep appearance models for face rendering. *ACM Transactions on Graphics* 37, 4 (Aug 2018), 1–13.

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A Skinned Multi-Person Linear Model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 34, 6 (2015), 248:1–248:16.

Shugao Ma, Tomas Simon, Jason M. Saragih, Dawei Wang, Yuecheng Li, Fernando De la Torre, and Yaser Sheikh. 2021. Pixel Codec Avatars. *CoRR* abs/2104.04638 (2021). arXiv:2104.04638

Yisroel Mirsky and Wenke Lee. 2020. The Creation and Detection of Deepfakes: A Survey. *CoRR* abs/2004.11138 (2020). arXiv:2004.11138

Saori Morishima, Ko Ayusawa, Eiichi Yoshida, and Gentiane Venture. 2016. Whole-body motion retargeting using constrained smoothing and functional principle component analysis. In *Int. Conf. on Humanoid Robotics*.

Lucio Moser, Chinyu Chien, Mark Williams, Jose Serra, Darren Hendler, and Doug Roble. 2021. Semi-Supervised Video-Driven Facial Animation Transfer for Production. *ACM Trans. Graphics (Proc. SIGGRAPH)* 40, 6, Article 222 (2021).

Jacek Naruniec, Leonhard Helminger, Christopher Schroers, and Romann Weber. 2020-07. High-Resolution Neural Face Swapping for Visual Effects. *Computer Graphics Forum* 39, 4 (2020-07), 173 – 184.

Yuval Nirkin, Yosi Keller, and Tal Hassner. 2019. FSGAN: Subject agnostic face swapping and reenactment. In *Proceedings of the IEEE International Conference on Computer Vision*. 7184–7193.

Jun-yong Noh and Ulrich Neumann. 2001. Expression Cloning. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '01)*. Association for Computing Machinery, 277–288.

Jedrzej Orbik, Shile Li, and Dongheui Lee. 2021. Human hand motion retargeting for dexterous robotic hand. In *Int. Conf. on Ubiquitous Robots*.

Ahmed A. A. Osman, Timo Bolkart, and Michael J. Black. 2020. STAR: Sparse Trained Articulated Human Body Regressor. In *European Conference on Computer Vision (ECCV)*, Vol. LNCS 12355. 598–613.

L. Penco, B. Clement, V. Modugno, E. Mingo Hoffman, G. Nava, D. Pucci, Nikos G. Tsagarakis, J.-B. Mouret, and S. Ivaldi. 2018. Robust Real-Time Whole-Body Motion Retargeting from Human to Humanoid. In *Int. Conf. on Humanoid Robotics*.

Ivan Perov, Daiheng Gao, Nikolay Chervoniy, Kunlin Liu, Sugasa Marangonda, Chris Umé, Mr. Dpfks, Carl Shift Facenheim, Luis RP, Jian Jiang, Sheng Zhang, Pingyu Wu, Bo Zhou, and Weiming Zhang. 2021. DeepFaceLab: Integrated, flexible and extensible face-swapping framework. arXiv:2005.05535 [cs.CV]

Yurui Ren, Ge Li, Yuanqi Chen, Thomas H. Li, and Shan Liu. 2021. PIRenderer: Controllable Portrait Image Generation via Semantic Neural Rendering. arXiv:2109.08379 [cs.CV]

Roger Blanco i Ribera, Eduard Zell, J. P. Lewis, Junyong Noh, and Mario Botsch. 2017. Facial Retargeting with Automatic Range of Motion Alignment. *ACM Trans. Graph.* 36, 4, Article 154 (2017).

Jun Saito. 2013. Smooth Contact-Aware Facial Blendshapes Transfer. In *Proceedings of the Symposium on Digital Production*. Association for Computing Machinery, 7–12.

Igor Santesteban, Elena Garces, Miguel A. Otaduy, and Dan Casas. 2020. SoftSMPL: Data-driven Modeling of Nonlinear Soft-tissue Dynamics for Parametric Humans. *Computer Graphics Forum (Proc. Eurographics)* (2020).

Yeongho Seol, J.P. Lewis, Jaewoo Seo, Byungkuk Choi, Ken Anjyo, and Junyong Noh. 2012. Spacetime Expression Cloning for Blendshapes. *ACM Trans. Graphics (Proc. SIGGRAPH)* 31, 2, Article 14 (apr 2012).

Yeongho Seol, Jaewoo Seo, Hyunjin Kim, J.P. Lewis, and Junyong Noh. 2011. Artist Friendly Facial Animation Retargeting. *ACM Trans. Graphics (Proc. SIGGRAPH)* 30 (12 2011), 162.

Mike Seymour, Chris Evans, and Kim Libreri. 2017. Meet Mike: Epic Avatars. In *ACM SIGGRAPH 2017 VR Village*. Association for Computing Machinery, Article 12.

Jaewon Song, Byungkuk Choi, Yeongho Seol, and Junyong Noh. 2011. Characteristic Facial Retargeting. *Comput. Animat. Virtual Worlds* 22, 2–3 (2011), 187–194.

Robert W. Sumner and Jovan Popović. 2004. Deformation Transfer for Triangle Meshes. *ACM Trans. Graphics (Proc. SIGGRAPH)* 23, 3 (2004), 399–405.

Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Niessner. 2016. Face2Face: Real-Time Face Capture and Reenactment of RGB Videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Yuhan Wang, Xu Chen, Junwei Zhu, Wenqing Chu, Ying Tai, Chengjie Wang, Jilin Li, Yongjian Wu, Feiyue Huang, and Rongrong Ji. 2021. HifiFace: 3D Shape and Semantic Prior Guided High Fidelity Face Swapping. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*. International Joint Conferences on Artificial Intelligence Organization, 1136–1142.

Thibaut Weise, Sofien Bouaziz, Hao Li, and Mark Pauly. 2011. Realtime Performance-Based Facial Animation. *ACM Trans. Graphics (Proc. SIGGRAPH)* 30, 4, Article 77 (2011).

Chenglei Wu, Derek Bradley, Markus Gross, and Thabo Beeler. 2016. An Anatomically-constrained Local Deformation Model for Monocular Face Capture. *ACM Trans. Graphics (Proc. SIGGRAPH)* 35, 4, Article 115 (2016).

Feng Xu, Jinxiang Chai, Yilong Liu, and Xin Tong. 2014. Controllable High-Fidelity Facial Performance Transfer. *ACM Trans. Graph.* 33, 4, Article 42 (2014).

Juyong Zhang, Keyu Chen, and Jianmin Zheng. 2022. Facial Expression Retargeting From Human to Avatar Made Easy. *IEEE Transactions on Visualization and Computer Graphics* 28, 2 (2022), 1274–1287.

Jiangning Zhang, Xianfang Zeng, Mengmeng Wang, Yusu Pan, Liang Liu, Yong Liu, Yu Ding, and Changjie Fan. 2020. FReeNet: Multi-Identity Face Reenactment. In *CVPR*. 5326–5335.