Pacific Graphics (2011)
Bing-Yu Chen, Jan Kautz, Tong-Yee Lee, and Ming C. Lin (Editors)

Short Papers

Facial Retargeting by Adding Supplemental Blendshapes

Paul Hyunjin Kim[1], Yeongho Seol[1], Jaewon Song[1] and Junyong Noh[1]

[1]Korea Advanced Institute of Science and Technology, Republic of Korea

Abstract

This paper introduces a novel method to add a minimal set of missing blendshapes to automatically improve the quality of the retargeting result. Our approach compares the expression spaces defined by the source and the target to determine which frames need to be corrected. The expression spaces are represented on the principal axes extracted from the target blendshapes. Those frames that cannot be sufficiently reconstructed by the current set of blendshapes will be augmented by a set of automatically generated blendshapes. The new blendshapes are determined to minimize error between the source animation and the retargeting result. We performed experiments to compare the results created by a basic set of blendshapes and by our method. Our method effectively reduces error between the source and target animation, and produces much visually improved target animation without relying on manual intervention.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation

1. Introduction

Facial retargeting is a widely employed technique to produce facial animation, which applies the performance of an actor/actress or source animation to a target virtual model. It is well known that famous movies such as *Kingkong*, *The lord of the rings* and *Avatar* utilized the facial retargeting technique to create realistic facial expressions on the face of the virtual 3D characters. For the creation of target animation, utilizing a set of blendshapes is a popular choice as it is easy to control blendshapes' parameters to pose desired facial expressions.

However, despite its wide acceptance, the blendshape based facial retargeting process has some drawbacks. The range of expressions that human can form on the face is very wide. Unless hundreds of blenshapes are available from the beginning, which is rarely the case, the initially given set of blendshapes typically has a limited range of facial expressions, as the facial movement is constructed by a linear combination of blendshapes. If there is a facial expression that the target model cannot pose corresponding to the expression in the source animation, additional blendshape needs to be sculpted and added to the set.

This paper introduces a novel method to automatically create appropriate blendshapes during the retargeting process, which increases the quality of retargeting while eliminating manual intervention. We compare the expression space of facial animation created by a source sequence with that of the target model created by its current blendshapes in order to find specific frames in which the source expression is not correctly represented by the target model's blendshapes. For those identified frames, our method measure how much the retargeting results are deviated from the source expressions and creates new blendshapes to fix the retargeting results. Incorporation of the linear shell model [BBA*07] allows easy creation of supplemental blendshapes.

2. Related Work

There have been active researches related to use of blendshapes in facial animation. Bergeron and Lachapelle [BL85] introduced the concept of linear blendshape models. Based on Ekman's FACS [EF78] that decomposes facial expressions into 46 shapes, those linear blendshape models effectively captured and reproduced realistic facial animation [ARL*09], using hand crafted blendshape models. As a way to expedite the process of blendshape construction, a blendshape based facial rig can be generated automatically from photographs [PHL*06] or 3D scan data [ZSCS07]. Li et al. [LWP10] showed the automatic creation of the facial rig

DOI: 10.2312/PE/PG/PG2011short/089-092

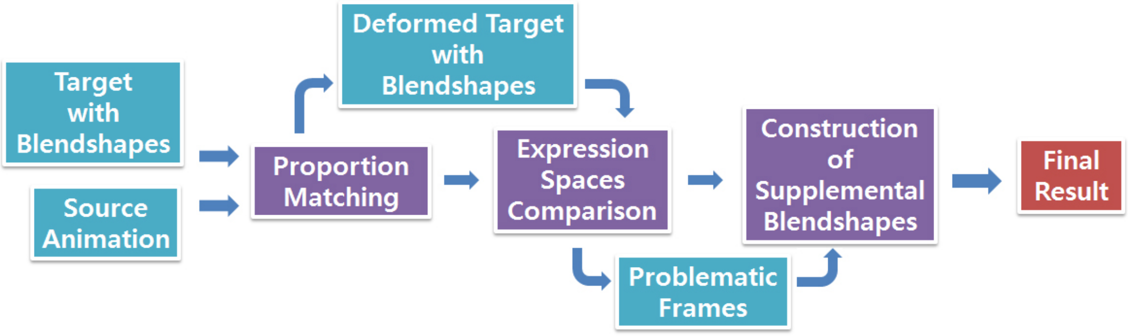

Figure 1: *The conceptual workflow of our method. The process consists of three main steps (purple boxes). In each step, there are input and output values (aqua boxes) which are transferred along the flow of process(blue arrows).*

blendshapes preserving the semantics of a target model. The method for automatic consturction of a blendshape rig is useful in facial retargeting. However, for a different input sequence, the facial rig has to be reconstructed repeatedly.

There has also been a research effort that focuses on improving the quality of facial retargeting, based on blendshapes. Choe et al. [CLK01] adjusted the muscle actuation base to animate blendshape face models. Chuang and Bregler [CB02] analyzed the space of source data to find key shapes from the animation sequence. The weights for the target blendshapes were simply transferred from the source animation. In order to analyze blendshape weights accurately, speech data were synthesized with motion capture data in [DCFN06]. Lewis and Anjyo [LA10] imposed penalty on large weights of blendshapes to prevent distortion of an extreme pose. Previous methods are not related to widening expression space defined by a set of blendshapes. This paper focuses on raising the expressiveness of the target blendshapes by adding missing ones automatically, to faithfully reconstruct the characteristics of the source animation.

3. Method

Given a source animation and a target face model with a reasonable set of initial blendshapes, our method generates supplemental blendshapes and enhances retargeting quality. The method consists of three main steps, as illustrated in Figure (1). In addition, we adopt the local blendshape approach [JTDP05] and segment the face model into two regions such as the upper region and the lower region (Figure (2)). The second and the third step of our method are applied to each region.

3.1. Proportion Matching between the Source and the Target Models

To deform the target model and the associated blendshapes to match the proportion of the source, a small number of feature points are identified between both models. The motion capture markers can be used as feature points. Then, Thin Plate Splines (TPS) deformation [OZS08] which is a special case of Radial Basis Function Warping [CFB97] deforms the target model satisfying the positional constraints defined by the feature points. The target blendshapes are also deformed based on the relationship between the original and the deformed target models (Figure (3)).

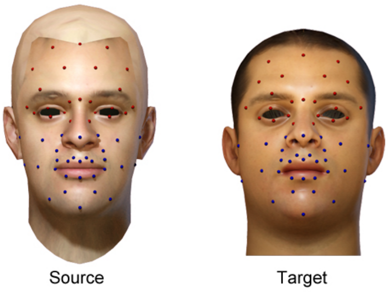

Figure 2: *The source and the target face models have feature points both in the upper region (red) and the lower region (blue).*

3.2. Comparison of Expression Spaces

To identify problematic frames, we employ the method in [CB02] which finds the key poses in the source animation and extend it to serve our purpose. A data matrix **S** for the source animation and **D** for the deformed target blendshapes are constructed. To find the frames that the blendshapes of the deformed target model cannot correctly represent, the space of matrix **D** is compared with the space of matrix **S** using Principal Component Analysis (PCA). Our experiments show that we obtain plausible results when PCA coefficients cover 95% of the deformed target data, determined by the

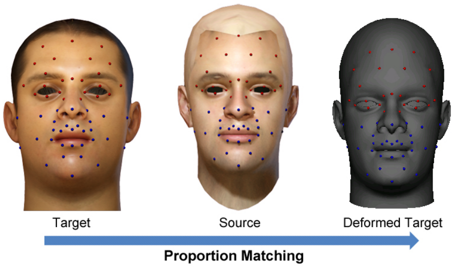

Figure 3: *Using TPS deformation with correspondence of feature points, the target model is deformed to the shape of the source model.*

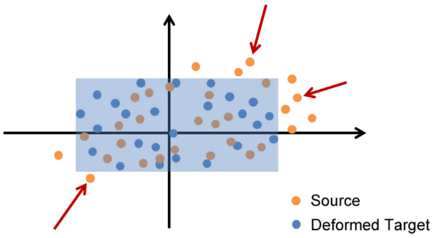

Figure 4: *A conceptual diagram of expression space comparison on the principal axes (black arrows) between the data of source's frames (orange dots) and those of deformed target blendshapes (blue dots). The expression space defined by both max and min values of the deformed target (the transparent blue rectangle) only covers a subset of source data. The points that lie outside (like those indicated by red arrows) are selected as problematic frames.*

size of eigenvalue λ. The matrix of the deformed target data is then projected onto the reduced principal axes. The matrix of the source data is also projected onto the same axes.

For each principal axis, we check if each maximum and minimum value of data representing a source expression falls inside the range defined by the maximum and the minimum value of the deformed target. The maximum and the minimum values of the source falling outside the range indicates that the source expression cannot be correctly represented by the current blendshapes set and additional shapes are needed (Figure (4)). Note that we only focus on the maximum and minimum values of the source, as correcting them will also correct all the expressions represented by in-between values if there are any. Repetition of the process for each principal axis identifies all the source expressions that fall outside the expression space of the deformed target blendshapes. Now we know that which source expressions need to be augmented.

3.3. Construction of Supplemental Blendshapes

This step determines new blendshape expressions that will augment the current set of blendshapes and correct the problematic frames. The desired expression is the one that minimizes error represented by the distances of feature points between the source expression and the retargeting result. The error is measured for each problematic expression found in Section 3.2. As the source model and the target model have different proportions, the error is measured between the source and the retargeting result of the deformed target model. For each problematic expression, when the source's expression is recreated using the blendshapes of the deformed target model, this paper adopts the constrained non-negative least squares method used in [LA10].

Our goal is to sculpt an additional blenshape in the target domain. As the target model has a different proportion than the deformed target model, the magnitude of error should be re-scaled using the inverse operation of Section 3.1. The rescaled magnitude of error is applied to the rest pose of the target model. Since local deformation is needed for sculpting the local blendshape model, the rest pose of the target model with the rescaled error values is deformed via linear shell model in [BBA*07] (Figure (5))

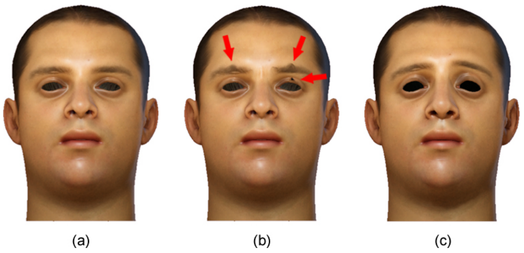
(a) (b) (c)

Figure 5: *(a) The rest pose of the target model. (b) The rescaled error values are applied to the feature points (indicated by red arrows) in the upper region. (c) The linear shell model deforms the target model.*

The deformed result of the rest pose is added to the original blendshapes' group of the target model as a supplemental blenshape. With the new blendshape group, the source animation is retargeted to the target model using NNLS to produce improved retargeting results.

4. Results

The facial animation of a hand crafted model, which consists of 100 frames was retargeted to two target models. In order to demonstrate performance of our system, the retargeting result using NNLS with a basic set of blendshapes (base method) was compared with the result produced by our method. The upper region contains 24 feature points and the lower region contains 30 points in all our experiments.

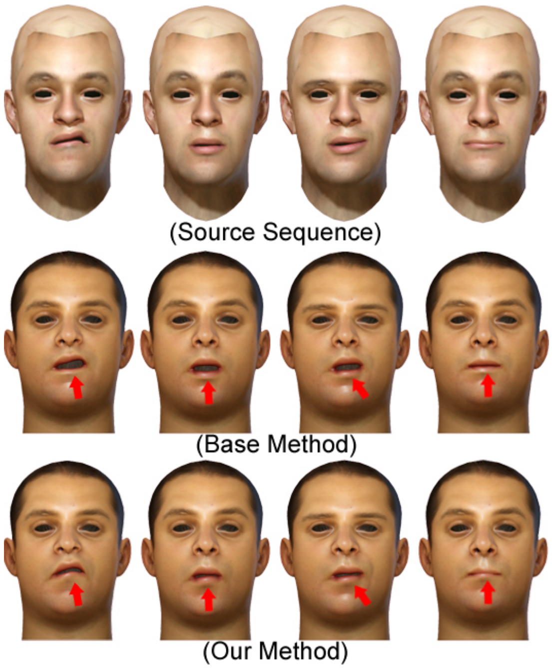

Figure 6: *Retargeting examples of various expressions. Our method faithfully recreates the original expressions and shows improved results compared to those produced by previous method.*

Figure (6) shows facial retargeting to the target model with 33 blendshapes. In the process of finding problematic frames for the lower region, the matrices were projected up to the 9th dimension of principal axes. Three problematic frames were identified from the source animation. For the upper region, the matrices were projected up to the 7th dimension of principal axes to identify 1 problematic frame. The red arrows indicate that our method faithfully recreates the original expressions and shows improved results compared to those produced by the base method.

5. Conclusion

A typically small set of blendshapes has a limited expression space which leads to artifacts in facial retargeting. To remedy the situation, our method identifies problematic frames that cannot be reproduced faithfully by given target model blendshapes. Supplemental blendshapes are then automatically created to widen the expression space of the target. The generation of supplemental blendshapes is minimal as our fix is not performed per frame. However, the generated blendshapes are enough to handle all the problematic frames. Using our approach, tedious trial and error process by the user can be avoided.

Acknowledgement

This work was supported by culture contents industry research and development program of KOCCA/MCST(210-7602-003-10743-01-007, Software Development for 2D to 3D Stereoscopic Image Conversion).

References

[ARL*09] ALEXANDER O., ROGERS M., LAMBETH W., CHIANG M., DEBEVEC P.: The digital emily project: photoreal facial modeling and animation. In *ACM SIGGRAPH 2009 Courses* (2009), ACM, pp. 1–15. 1

[BBA*07] BICKEL B., BOTSCH M., ANGST R., MATUSIK W., OTADUY M., PFISTER H., GROSS M.: Multi-scale capture of facial geometry and motion. *ACM Transactions on Graphics (TOG) 26*, 3 (2007), 33–es. 1, 3

[BL85] BERGERON P., LACHAPELLE P.: {Controlling facial expressions and body movements in the computer generated animated short\'Tony de Peltrie\'}. 1

[CB02] CHUANG E., BREGLER C.: Performance driven facial animation using blendshape interpolation. *Computer Science Technical Report, Stanford University* (2002), 2002–02. 2

[CFB97] CARR J., FRIGHT W., BEATSON R.: Surface interpolation with radial basis functions for medical imaging. *Medical Imaging, IEEE Transactions on 16*, 1 (1997), 96–107. 2

[CLK01] CHOE B., LEE H., KO H.: Performance-driven muscle-based facial animation. *The Journal of Visualization and Computer Animation 12*, 2 (2001), 67–79. 2

[DCFN06] DENG Z., CHIANG P., FOX P., NEUMANN U.: Animating blendshape faces by cross-mapping motion capture data. In *Proceedings of the 2006 symposium on Interactive 3D graphics and games* (2006), ACM, pp. 43–48. 2

[EF78] EKMAN P., FRIESEN W.: Facial action coding system: A technique for the measurement of facial movement, 1978. 1

[JTDP05] JOSHI P., TIEN W., DESBRUN M., PIGHIN F.: Learning controls for blend shape based realistic facial animation. In *ACM SIGGRAPH 2005 Courses* (2005), ACM, pp. 8–es. 2

[LA10] LEWIS J., ANJYO K.: Direct Manipulation Blendshapes. *Computer Graphics and Applications, IEEE 30*, 4 (2010), 42–50. 2, 3

[LWP10] LI H., WEISE T., PAULY M.: Example-based facial rigging. *ACM Transactions on Graphics (TOG) 29*, 4 (2010), 1–6. 1

[OZS08] ORVALHO V., ZACUR E., SUSIN A.: Transferring the rig and animations from a character to different face models. In *Computer Graphics Forum* (2008), vol. 27, Wiley Online Library, pp. 1997–2012. 2

[PHL*06] PIGHIN F., HECKER J., LISCHINSKI D., SZELISKI R., SALESIN D.: Synthesizing realistic facial expressions from photographs. In *ACM SIGGRAPH 2006 Courses* (2006), ACM, pp. 19–es. 1

[WLVGP09] WEISE T., LI H., VAN GOOL L., PAULY M.: Face/off: Live facial puppetry. In *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2009), ACM, pp. 7–16.

[ZSCS07] ZHANG L., SNAVELY N., CURLESS B., SEITZ S.: Spacetime Faces: High-Resolution Capture for˜ Modeling and Animation. *Data-Driven 3D Facial Animation* (2007), 248–276. 1