Technical Report

Department of Computer Science
and Engineering
University of Minnesota
4-192 Keller Hall
200 Union Street SE
Minneapolis, MN 55455-0159 USA

TR 17-002

Data-Driven Variation for Virtual Facial Expressions

Nick Sohre, Moses Adeagbo, Stephen Guy

March 16, 2017

Varied Happy Smiles: Data-Driven Variation in Facial Expressions

Nick Sohre, Moses Adeagbo, Stephen J Guy
University of Minnesota, Department of Computer Science & Engineering
sohre, sjguy @cs.umn.edu, adeag001@umn.edu

Figure 1: *A variety of happy smiles generated by our method.*

Abstract

Animating digital characters has an important role in computer assisted experiences, from video games to movies to interactive robotics. A critical component of digital character interaction is the animation of the human face. Here we explore a data-driven method to produce variation in animated smiles. We define a low-dimensional parameter space for learning based on key feature points of the face, which generalizes to arbitrary digital models. We perform a large-scale user study to annotate a systematic sweep of faces, and train a non-parametric classifier to predict the level of perceived happiness. This model is tuned to balance between precision and the variation in its predictions. New happy faces are then sampled from this model, resulting in a variety of generated faces that display a targeted level of happiness. This diversity can allow rich interactions with digital characters to be built automatically, without the need for hand-crafted expressions.

Keywords: data-driven facial animation, digital character emotion, computer graphics

1 Introduction

Virtual humans are increasingly a part of our games and other digital media. They appear in movies as animated actors, video games as interactive non-player characters, personal avatars in games, virtual reality and social media, and are even used to control human-like robots. A critical component of creating compelling interactions with digital characters is the animation of the human face. Humans use and expect faces to produce a variety of cues for communicating things like intonation and emotion. Understanding the relationships between facial movements and these cues has implications extending outside the field of computer graphics, including both medicine and psychology.

As technology has become more sophisticated and graphics capabilities increase, the desire for more realistic characters in both appearance and motion also grows. Research has shown that variety in appearance of digital characters is an important factor in creating experiences that exhibit realism and immersion [McDonnell et al. 2008b; OSullivan 2009]. Specifically, the human face is of particular interest, as it is crucial for interacting with characters in a virtual world [Sinha et al. 2006]. Diversity of facial expressions is not only important for creating a variety of characters, but also making individual characters more rich and life-like.

Happiness is among one of the most basic and important emotions conveyed by the human face. As smiles vary widely both within and across individuals, creating compelling virtual characters must also exhibit these kinds of diversity. For this work, we wish to automate the process of creating multiple faces that are all varied, but still are happy. We introduce the problem of automatically producing realistic facial expressions that convey a target emotion with controlled extent for a digital character, and study it within the context of happiness. This task presents several challenges that need to be addressed. In order to extend to a variety of applications, our solution must be versatile enough to be applicable to many digital characters. We also need to capture the complex relationship between emotions and facial movements in a model that supports facial synthesis.

Current methods for producing facial expressions tend towards utilizing facial performance capture, physically based models of facial movement, or using animations handcrafted by artists. Facial performance capture, while effective, can involve tedious calibration and/or require high end equipment, as well as actors that can produce every desired facial expression. Physically based models have shown great promise, but depend on very complex simulations and require extensive knowledge of facial tissue dynamics. Many digital characters are hand designed by artists, which can be tedious for the artists and costly for practitioners. To alleviate these issues and overcome the aforementioned challenges, we present the following as contributions for this work:

- *A data-driven method for automatically discovering new faces* We train a non-parametric classifier on the annotated faces in our standardized space to predict the quality of new faces. This model produces a variety of new faces with the desired predicted happiness.
- *A low dimensional feature space for facial expressions* We propose a novel basis for representing facial expressions. This space is based off of key facial feature points, resulting in a highly flexible representation that is agnostic of facial animation technique. The space is semantically meaningful, readily extending to other fields of study.

- *A variance maximizing classification method* We introduce the problem of generating faces with a targeted level of happiness, and formulate it as a classification task that prefers high precision over high accuracy in order to maximize variation in the targeted class.

The result is a system capable of automatically and quickly producing a variety of faces with a targeted happiness level. The system is generic, and may be re-purposed to work on arbitrary facial models.

2 Background

The animation of digital human-like faces has a rich history in the literature, from performance capture to modeling, to human perception of facial actions, and creating facial expressions for digital characters. Below, we briefly highlight some closely related works. For a more complete overview of the field, we refer the reader to the survey by [Vinayagamoorthy et al. 2006].

2.1 Modeling & Manipulating the Human Face

Several methods have been proposed for the digital representation and manipulation of the human face. Common to all methods in computer graphics is the use of a 3D spatial mesh to represent the face, that is then deformed according to some model of facial movement. Such models include those based on physical interactions of skin, subcutaneous tissue, muscle, and skeletal structure [Waters 1987; Cong et al. 2015; Lee et al. 1995]. These approaches can achieve very realistic behavior, at the cost of a high level of complexity in the model. Parametric descriptors based on these models have also been developed to categorize muscle movements as facial actions [Ekman and Friesen 1977; Essa and Pentland 1997].

Another model involves mixing amounts of predefined deformations of a base mesh. Known as animation blendshapes, these are widely used in both industry and research. A large body of work involves producing facial animations based on the facial performance of an actor. Tracking the movements of an actor's face, called performance capture, is then translated into deformations of a 3D facial mesh in a process called retargeting. Many approaches to this task have been proposed [Zhang et al. 2016; Bouaziz et al. 2013; Li et al. 2013; Xu et al. 2014]. Researchers have propose various methods, from adaptive dimensionality reduction [Li et al. 2013], to neural networks [Costigan et al. 2014] to local patch alignment [Zhang et al. 2016]. Others focus on constructing detailed, realistic facial models from depth & reflectance sensor data [Pighin et al. 2006] and generating blendshape segmentation schemes [Joshi et al. 2005].

2.2 Human Perception of Digital Characters

Using computer animated faces for both communicating and studying human perception of emotion has been used in previous efforts [Griesser et al. 2007]. User studies involving digital characters have been used to study the realism and effectiveness of digital characters. These include studying the realism of facial animations and its impact on digital character appeal [Kokkinara and McDonnell 2015; McDonnell et al. 2008a; McDonnell 2012]. Other studies have shown that a variety of digital character appearances can produce more compelling experiences [McDonnell et al. 2008b], including the human face specifically [OSullivan 2009]. In addition, the literature shows that user studies of human perception of virtual characters is useful for studying their emotional expressiveness [Liu et al. 2016]

2.3 Generating Facial Expressions

Generative methods for digital character facial expressions have also recently been explored. The authors of [Si and McDaniel 2015] investigate how best to create smiles for robotic and digital characters. Other works, as in that of in [Cassel and Stone 1994] focuses on generating facial expressions synchronous with dialogue. Other systems generate facial expressions from dialogue audio and text transcripts [Marsella et al. 2013].

3 Data Collection

Our approach utilizes annotated faces as a basis to learn a model of happiness. To build such a dataset, we performed a large-scale user study at the 2015 Minnesota State Fair. The stimuli were composed of digital facial animations representing a systematic sweep of anatomically plausible faces. Participants were shown these expressions on a tablet device, and asked to evaluate each in terms of emotional intent, as well as assign a quality score for how well the face portrayed a smile (Figure 2). Over 900 subjects participated in the survey, providing over 10,000 responses in total. The stimuli contained mostly smile-like faces, but also had some negatively angled facial expressions, which served as controls.

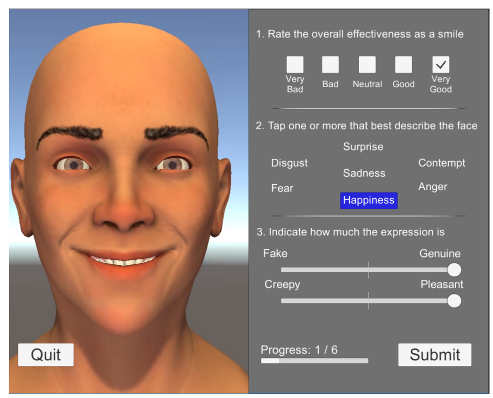


Figure 2: *A screenshot of the application used to conduct the user study. Subjects responded to stimuli by rating each in terms of smile effectiveness and emotional intent.*

A single response contains, for a given facial expression, the smile quality score the participant assigned. These responses were then aggregated for each facial expression, producing the average quality score for each face. The result is a dataset composed of 63 facial expressions annotated with perceived smile quality. Although the user study stimuli were rendered on a single facial model, the design and use of facial space is indented (and subsequently verified in Section 7.1) to generalize the findings to other faces.

The results of our user study are summarized in Figure 4. The range of quality scores is well covered by the systematic sweep of facial space points, with the majority of the standard errors being reasonably small. There are three faces in particular that were outliers in terms of standard error, and are suppressed from use in learning. These faces exhibited similar animation artifacts that contributed to a lack of consensus in quality scores.

We bin the data to produce groups that exhibit a high level of statistical significance for learning. This helps alleviate negative effects of uncertainty in the quality estimates. An ANOVA test shows high statistical significance with 4 bins, with $[F(3, 60) = 863.5, p < 0.001]$. A post-hoc analysis also confirms statistical significance between all pairs of bins. We label these bins by level of happiness

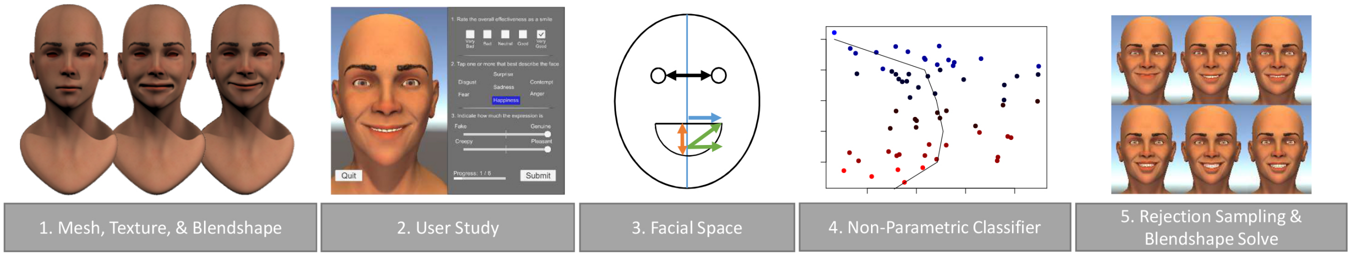
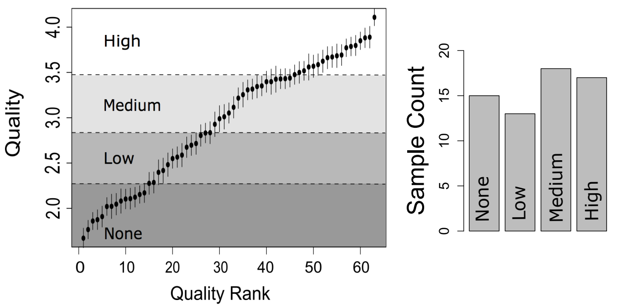


Figure 3: *A graphical overview of our pipeline.*

as *None, Low, Medium*, and *High*. The distribution of faces in these bins is depicted in Figure 4b.



(a) *Smile Quality Scores* **(b)** *Happiness Bins*

Figure 4: ***User Study Data.*** ***(a)*** *Each face's smile quality with its standard error. The data is binned into effective happiness bins.* ***(b)*** *The distribution of faces by happiness bins.*

4 Face Generation

The design of our facial generation pipeline allows new facial expressions to be automatically generated from user data. Figure 3 shows a high level overview of this process. The first step of collecting facial expressions annotated with perceived happiness by real humans captures the relationships between movements of the face and the emotional content. Dimensionality reduction focused around mouth features is used to assist with learning a model of happiness and preserving the data's semantic content. Predicting happiness levels of new faces is enabled by training a classifier on the annotated faces. Additional faces in our low dimensional space can then be generated via rejection sampling, keeping random samples having the desired predicted happiness level and discarding others. To create representations of the resulting smiles that can be rendered, the samples are transferred onto an existing 3D facial mesh. Some examples drawn with high quality real-time rendering software are shown in Figure 1.

The choice of facial feature space can have a strong impact on both the amount of data needed to generate a model, and the quality of the resulting faces. Section 5 proposes a feature space designed to generalize well to other facial meshes. Additionally, the learning method is a key element of our smile generation process. In Section 6, we introduce a new learning algorithm designed to maximize the precision of our classifier while maintaining a variety of generated faces.

5 Facial Space

We propose a low dimensional human facial configuration space to parameterize expressions based on key facial feature points. These feature points are chosen to focus on the region of the mouth. The mouth has been shown to be a principal source of variation in facial expressions [Köhn 2006], as well as the primary source of information for detecting happiness [Nusseck et al. 2008]. Further motivated by the mouth's relevance in facial medicine and psychology, this new space is constructed as the combination of three facial features: *lip corner angle, lip corner extension,* and *dental show*. To compute these features, we track three control points as they move with the face: the intersections of the upper and lower lips and the saggital plane, and the left mouth corner (we simplify facial movement by assuming symmetry across the saggital plane). The values for each parameter are computed as relationships between these control points (see Figure 5). Each measure is normalized relative to the inter-pupilary distance of the face to standardize across different face sizes, which is known to be proportional to other features of the face [Stephan 2003]. The low dimensional nature of this feature space is conducive for learning what makes a facial expression appear happy. Additionally, facial space is a standardized space, and generalizes to any facial model that exhibits these facial features. Being derived from key facial points, each dimension is semantically meaningful, which extends well to other fields of study.

To render points in facial space to an image, they must be transferred onto a digital facial model. Here, we use a 3D mesh with interpolative blendshapes as defined by an artist. Then the task of adjusting the facial model to match the target facial space point becomes one of transforming the point into the space of blendshape weights. As this transformation is non-linear, we compute a local linear gradient in facial space with respect to the blendshape weights and implement an iterative optimization algorithm to minimize the squared error. We note that this method can be applied to any facial animation system that is locally controllable with respect to facial space.

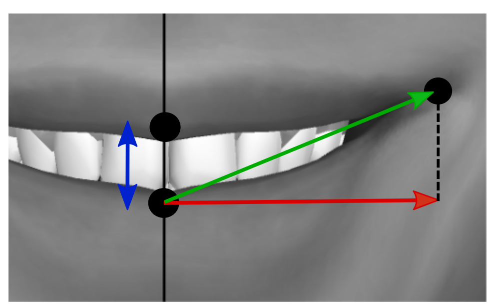


Figure 5: *Computation of facial space features. Control points are shown in black, and the solid black line shows the position of the saggital plane. Angle is computed as the angle between the green and red arrows. Extent is the length of the red arrow, and dental show is the extent of the blue arrows.*

6 Precision-Variety Learning

The approach outlined in Section 4 makes use of a binary classifier to identify faces that match the targeted level of happiness. Traditional binary classifiers seek to maximize predictive accuracy. However, our work relies more on the precision, that is, the rate of correct classification only among the faces classified as happy. Because our model only generates positively classified faces, faces incorrectly classified as unhappy will be discarded, unseen by any user. In this way our model does not suffer from loss of accuracy due to false negatives. This property suggests evaluating our classifier based on precision. However, with any binary classification task there is a fundamental trade-off between precision and variety; maximizing one comes at the cost of the other. Below, we present a new classification method that exposes this trade-off to allow us to maximize variety in positively classified faces while retaining as much precision as possible.

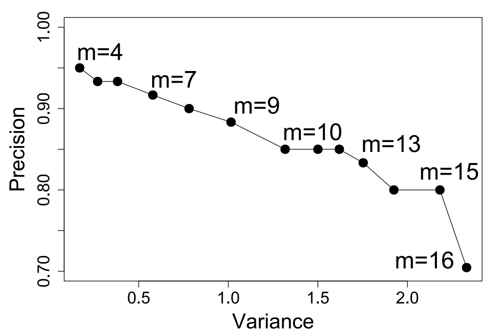


Figure 6: *Precision-Variance Tradeoff curve over m for the High happiness bin.*

6.1 Precision Variance Trade-off

High precision can be ensured by carefully selecting which positive samples are allowed in to the training set. For example, choosing to only include positive training samples that are far away from negative samples increases the precision of the model at the cost of false negatives, which is a favorable trade given our goals. However, including too few positive training samples results in very little variety, which is an equally important objective. Varying the positive samples allowed into the training set with respect to their nearest negative neighbor exposes this trade-off for tuning between precision and variety. We define a new parameter m to be the number of samples in decreasing distance from their closest negative neighbor included in the training set. This effectively manipulates the margin of separation between the positive and negative training samples. Allowing more positive samples increases the variety in the predictions, at the cost of some precision due to a narrower margin.

6.2 Precision Variety Learning (PVL)

Our annotated facial data serves as a basis for learning a smile effectiveness model. A variant of the k-nearest neighbors (KNN) classifier is used to build this model. A one-vs-rest binary classification scheme allows multiple happiness levels to be learned simultaneously. Our algorithm determines the predicted class of a sample by looking for training samples within a defined radius r. If the proportion of positive training samples found exceeds a threshold $t, 0 \leq t \leq 1$, the test point is predicted as belonging to the positive class. An n parameter specifies the minimum number of training samples within the radius to test for classification. For our classifier, we choose $n = 6$ based off the density of our training data, $r = 0.4$ based on the distribution of inter-point distances, and $t = 0.3$ via manual tuning.

Algorithm 1: PVL Prediction

Input : $sample, trainData, pClass, m, r, n, t$
Output: Prediction
$pos \leftarrow$ getPositiveSamples($trainData, pClass$);
$neg \leftarrow$ getNegativeSamples($trainData, pClass$);
$pos \leftarrow$ sortByDistanceToNearest(pos, neg);
$pos \leftarrow$ firstN($positive, m$);
$trainData \leftarrow$ union($positive, negative$);
$witnesses \leftarrow 0$;
$votes \leftarrow 0$;
for i = *1:size($trainData$)* **do**
  if *distance($sample, trainData[i]$)$< r$* **then**
    $witnesses \leftarrow witnesses + 1$;
    $votes \leftarrow votes +$ isClass($trainData[i]$,pClass);
  end
end
if $witnesses < n \vee votes/witnesses < t$ **then**
  return($False$);
end
return($True$);

To obvserve the effect of m on classification, we compute precision using hold-one-out cross validation with varying m, and estimating the variety of each resulting classifier as the average variance of 100 positively classified points sampled each fold. The results are shown in Figure 6: for small m, the precision of the model remains high, but results in a classifier that produces little variety when sampled. Conversely, for large m, a larger variety of points can be generated, at the cost of precision. Thus, m allows us to tune the precision/variety trade-off in the learning process. We call this approach *Precision-Variance Learning* (PVL). The resulting algorithm is presented in Algorithm 1.

The profile of the precision variance trade-off of the classifier varies depending on the positive class. Figure 7 shows the curves when the None, Low, and Medium bins are chosen as the desired happiness level. Regardless of the actual shape, the trade-off persists between the precision and the variance of the resulting classifier, and m can be tuned to maximize variance while maintaining a desired level of precision.

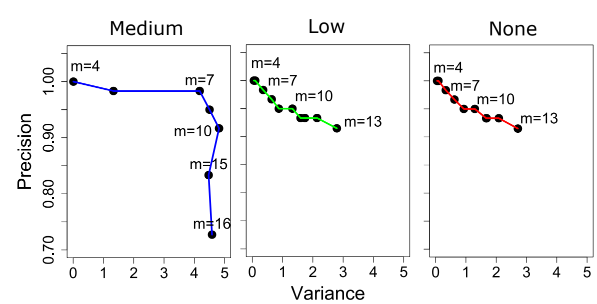


Figure 7: ***Response Curves for Lower Bins.*** *Trade-off curves over m for Happiness bins Medium, Low, and None*

7 Results

Our method is capable of producing a variety of facial expressions with a targeted level of happiness. Figure 1 demonstrates some examples where we asked our classifier for faces with high happiness, and setting $m = 8$. As this value for m allows for variety without a large loss of precision (Figure 6), the result is faces that differ in appearance, but are all happy. The PVL model can also be trained to produce faces from the different happiness categories. The middle and bottom rows of Figure 10 show faces generated from categories Medium and None.

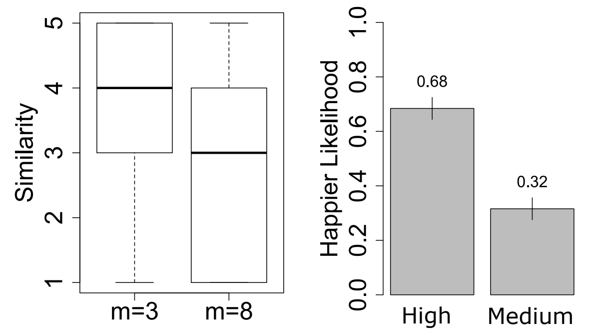
Figure 8: *Faces generated with High happiness, $m = 8$ applied to a new facial model with different appearance and blendshape set.*

The design of facial space also allows our method to extend to new facial models. We obtained a new facial model differing in appearance from our original, with blendshapes defined to perform natural movements of the face. Although these blendshapes are unknown a priori to our model, we may sample new faces with desired happiness level from the PVL classifier and perform the blendshape solve on the new model. Figure 8 shows some results for the High happiness level with $m = 8$.

7.1 Validation

To validate our method's effectiveness in creating a variety of faces within a targeted happiness level in addition to its extendability to new models, we performed a second user study. In this two-part study, 19 participants (11 women and 8 men with average age 28.3) were shown side-by-side pairs of generated animations and asked to assess them in terms of their similarity and relative happiness. The first section of the study analyzed the PVL learning approach proposed in Section 6. Here, pairs of faces were shown generated from sets with the same value m (i.e., either two faces from the m=3 set or two faces from the m=8 set). Some examples of the faces shown can be seen in the top row of Figure 10. Participants were asked to indicate on a Likert scale how similar the two facial expressions appeared. The expected results was that pairs from sets with lower m would be more similar (have less variety) that sets from higher m. Figure 9 summarizes the similarity ratings for the two types of comparisons. A t-test confirms ($p < 0.01$) that comparisons between two expressions from the m = 3 group were significantly more likely to be labeled as similar than those generated with m = 8, confirming our hypothesis and validating our PVL learning approach.

The second section of the study aimed at validating the predictive accuracy of our PVL classification approach. Here we used a Two-alternative forced choice (2AFC) design where participants were asked to indicate which of two smiles appeared happier. In every pair of smiles, one came from a set of smiles with a predicted



(a) *Variety Test* **(b)** *Happiness Test*

Figure 9: ***(a)*** *Faces generated with $m = 3$ appeared more similar to participants than those generated with $m = 8$.*
(b) *Faces generated with High happiness were significantly more likely to appear happier than those generated with Medium happiness.*

"Medium" category smile and one with a predicted "High" category smile. Both sets of smiles were generated with m=8, so as to have a variety of different smiles. Smile videos were displayed using the character shown in Figure 8, so as to test our prediction ability on an entirely new character not used during training. Figure 9 shows the results. Smiles form the predicted High happiness category were rated happier than ones from from the Medium category to a statistically significant degree ($p < .001$, student t-test). This serves not only to validate our ability to generate faces with varying levels of happiness, but also demonstrates our method can extend to mulitple facial meshes.

8 Discussion & Analysis

There are several trends and trade-offs worth noting in our method. The m parameter is not the only parameter that can be used to tune the behavior of the PVL algorithm. For example, increasing t has a dampening effect on the model's sensitivity to m, and increasing r an amplifying effect. Appropriate values for n will depend on these other two, as well as the density of the training data. We also note that the model response for different targeted happiness levels will also impact the choice of m (Figure 7). These choices have a significant impact on the results, and the optimal values will depend on the nature of the training data. Likewise, different expressions may inherently allow different amounts of variation. As shown in Figure 7, it is easy find a very large variety of smiles which are "not happy very high precision. This is because there is naturally a wider variety of ways to express displeasure or discontent than there are ways to express pure happiness.

The performance of our method depends on several quantities. Among these are the dimensionality of facial space and the underlying control space of digital faces used for rendering (in our case, this is the number of blendshapes). Overall, the runtime complexity of our system is $O(mb + n^2)$, where m is the dimensionality of facial space, n is the number of training samples, and b is the number of blendshapes. This means our algorithm can scale well with increasing blendshapes or increasing facial space dimensions (though not necessarily both). The runtime growth of our system increases polynomialy with number of training samples, making training over a larger sample sets a potential bottleneck in building the classifier. This could be alleviated by utilizing spatial datastructures to accelerate the computation of finding neighbors.

While this work was limited to the generation of faces with differ-

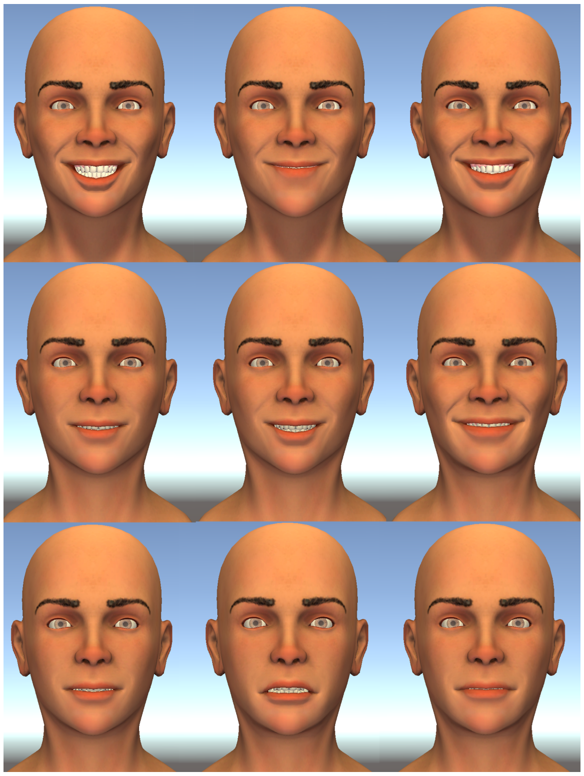


Figure 10: *Example set of generated faces with High (top row) Medium (middle row) and None (bottom row) targeted happiness levels.*

ent amounts of happiness, our system is not limited to a single emotion. The stimuli composing the user study was by majority smile or smile-like expressions, and consequently supports studying happiness much more effectively. Performing data collection on a larger range of facial expressions would allow for studying different emotions, provided they are sufficiently captured by the annotated data. This also requires blendshapes that allow for a broad range of facial movement. Similarly, we note that increasing the coverage and density of the annotated faces could allow for more granular categories or regression classifiers to be trained. This could support more fine-tuned control over the target emotions or be used to generate mixtures of emotions.

In our proposed feature space there are limitations in considering only feature points near the mouth. For our study of happiness, this was sufficient, but as our algorithms are independent of facial space, it could be extended to include other features corresponding to other control points on the face. However, this (as with increasing the number of control parameters for digital faces) will contribute to increasing dimensionality and its impact on computational and classification performance.

9 Conclusions

In this work, we have proposed and implemented a complete system for the generation of a variety of happy faces for use in digital characters. We performed a large scale user study to gather data on their perceived emotional intent. We then transformed them into *facial space*, a new space we proposed based on key facial features. These data were used as training samples for a non-parametric classifier to predict whether or not a new facial coordinate would be perceived as having a targeted level of happiness. To maximize the effectiveness of the classifier, we introduced *Precision-Variance* learning, which allows a balance between precision and variety of the model. The classifier was then used to generate a variety of faces with different happiness levels novel to the existing data.

9.1 Limitations

Some limitations of our method motivate further study. One such area is in the limited coverage our user study data has of plausible facial positions. While a wider range of facial space coverage was not necessary to develop our method, data covering a larger range of faces can enable the study of a more diverse set of emotions. Another limitation is the fact that blendshapes have extent, thereby necessitating constrained optimization. This constraint could be relaxed by allowing blendshapes to be extrapolated past their original bounds.

9.2 Future Work

In the future we intend to explore the generation of other emotions and mixtures of emotions by conducting additional user studies to add to our data pool. Similarly, we plan to extend facial space to include control points in other regions of the face that are anticipated to be important in expressing other emotions (for example, eye and eyebrow motion may be very important to express surprise). Building data-driven models that capture how perceived emotional intent relates to movements of key facial features has implications beyond making compelling digital characters. As digital representations and computer graphics allow us to permute facial positions in a way that human actors cannot, another exciting area of future work is to investigate faces with large asymmetry or other issues which may arise from facial trauma or nervous system damage. This can allow our work to inform areas of medicine such as facial reconstructive surgery, emotional recognition therapy, and psychologists looking to quantitatively study how treatment and intervention can help patients better express their emotional intent.

References

BOUAZIZ, S., WANG, Y., AND PAULY, M. 2013. Online modeling for realtime facial animation. *ACM Transactions on Graphics (TOG) 32*, 4, 40.

CASSEL, PELACHAUD, B. S. A. B. D. P., AND STONE. 1994. Animated conversation: Rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. Tech. rep.

CONG, M., BAO, M., BHAT, K. S., FEDKIW, R., ET AL. 2015. Fully automatic generation of anatomical face simulation models. In *Proceedings of the 14th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, ACM, 175–183.

COSTIGAN, T., PRASAD, M., AND MCDONNELL, R. 2014. Facial retargeting using neural networks. In *Proceedings of the Seventh International Conference on Motion in Games*, ACM, New York, NY, USA, MIG '14, 31–38.

EKMAN, P., AND FRIESEN, W. V. 1977. Facial action coding system.

ENNIS, C., AND EGGES, A. 2012. *Perception of Complex Emotional Body Language of a Virtual Character*. Springer Berlin Heidelberg, Berlin, Heidelberg, 112–121.

ESSA, I. A., AND PENTLAND, A. P. 1997. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Trans-*

actions on Pattern Analysis and Machine Intelligence 19, 7 (Jul), 757–763.

GRIESSER, R. T., CUNNINGHAM, D. W., WALLRAVEN, C., AND BÜLTHOFF, H. H. 2007. Psychophysical investigation of facial expressions using computer animated faces. In *Proceedings of the 4th symposium on Applied perception in graphics and visualization*, ACM, 11–18.

HYDE, J., CARTER, E. J., KIESLER, S., AND HODGINS, J. K. 2014. Assessing naturalness and emotional intensity: a perceptual study of animated facial motion. In *Proceedings of the ACM Symposium on Applied Perception*, ACM, 15–22.

JOSHI, P., TIEN, W. C., DESBRUN, M., AND PIGHIN, F. 2005. Learning controls for blend shape based realistic facial animation. In *ACM SIGGRAPH 2005 Courses*, ACM, 8.

KÖHN, H.-F. 2006. Combinatorial individual differences scaling within the city-block metric. *Computational statistics & data analysis 51*, 2, 931–946.

KOKKINARA, E., AND MCDONNELL, R. 2015. Animation realism affects perceived character appeal of a self-virtual face. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*, ACM, New York, NY, USA, MIG '15, 221–226.

LEE, Y., TERZOPOULOS, D., AND WATERS, K. 1995. Realistic modeling for facial animation. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, ACM, 55–62.

LI, H., YU, J., YE, Y., AND BREGLER, C. 2013. Realtime facial animation with on-the-fly correctives. *ACM Trans. Graph. 32*, 4, 42–1.

LIU, K., TOLINS, J., TREE, J. E. F., NEFF, M., AND WALKER, M. A. 2016. Two techniques for assessing virtual agent personality. *IEEE Transactions on Affective Computing 7*, 1, 94–105.

MARSELLA, S., XU, Y., LHOMMET, M., FENG, A., SCHERER, S., AND SHAPIRO, A. 2013. Virtual character performance from speech. In *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, ACM, 25–35.

MCDONNELL, R., JÖRG, S., MCHUGH, J., NEWELL, F., AND O'SULLIVAN, C. 2008. Evaluating the emotional content of human motions on real and virtual characters. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, ACM, 67–74.

MCDONNELL, R., LARKIN, M., DOBBYN, S., COLLINS, S., AND O'SULLIVAN, C. 2008. Clone attack! perception of crowd variety. In *ACM Transactions on Graphics (TOG)*, vol. 27, ACM, 26.

MCDONNELL, R. 2012. *Appealing Virtual Humans*. Springer Berlin Heidelberg, Berlin, Heidelberg, 102–111.

NUSSECK, M., CUNNINGHAM, D. W., WALLRAVEN, C., AND BÜLTHOFF, H. H. 2008. The contribution of different facial regions to the recognition of conversational expressions. *Journal of vision 8*, 8, 1–1.

OSULLIVAN, C. 2009. Variety is the spice of (virtual) life. In *International Workshop on Motion in Games*, Springer, 84–93.

PIGHIN, F., HECKER, J., LISCHINSKI, D., SZELISKI, R., AND SALESIN, D. H. 2006. Synthesizing realistic facial expressions from photographs. In *ACM SIGGRAPH 2006 Courses*, ACM, 19.

SI, M., AND MCDANIEL, J. D. 2015. Creating genuine smiles for digital and robotic characters: An empirical study. In *Games Entertainment Media Conference (GEM), 2015 IEEE*, IEEE, 1–6.

SINHA, P., BALAS, B., OSTROVSKY, Y., AND RUSSELL, R. 2006. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE 94*, 11, 1948–1962.

STEPHAN, C. N. 2003. Facial approximation: An evaluation of mouth-width determination. *American journal of physical anthropology 121*, 1, 48–57.

VINAYAGAMOORTHY, V., GILLIES, M., STEED, A., TANGUY, E., PAN, X., LOSCOS, C., SLATER, M., ET AL. 2006. Building expression into virtual characters.

WATERS, K. 1987. A muscle model for animation three-dimensional facial expression. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, ACM, New York, NY, USA, SIGGRAPH '87, 17–24.

XU, F., CHAI, J., LIU, Y., AND TONG, X. 2014. Controllable high-fidelity facial performance transfer. *ACM Transactions on Graphics (TOG) 33*, 4, 42.

ZHANG, J., YU, J., YOU, J., TAO, D., LI, N., AND CHENG, J. 2016. Data-driven facial animation via semi-supervised local patch alignment. *Pattern Recognition 57*, 1–20.