CAFM: A 3D Morphable Model for Animals

Yifan Sun
Waseda University, Shinjuku, Tokyo, Japan
sunyifan2016@gmail.com

Noboru Murata
Waseda University, Shinjuku, Tokyo, Japan
noboru.murata@eb.waseda.ac.jp

Abstract

We present Cat-like Animals Facial Model (CAFM) – a 3D Morphable Model (3DMM) constructed from 50 samples, including lion, tiger, puma, American Shorthair, Abyssinian cat, etc. To the best of our knowledge, CAFM is the first animal morphable model ever constructed. New animal face images can be registered automatically by fitting pose and shape parameters of CAFM. Moreover, the parametric model regulates the naturalness of the generated animal faces avoiding unreasonable appearance.

Computer vision has recently experienced great advances in automatic facial landmark detection. In this paper, to demonstrate CAFM's application to 3D reconstruction of cat face images, and to put effort towards uniform annotation scheme of immense databases and fair experimental comparison of cat-like animals' facial landmark systems, we improve the labeled cat face data set of 10,000 images with 15 landmarks. Besides, we propose an algorithm matching our model to the input cat face images. With the projection parameters and shape parameter of CAFM, we can generate corresponding 3D meshes.

1. Introduction

3D reconstruction of all generic objects has always been a long term goal in computer vision. For human, the task has achieved good results on human face [10, 8, 11] and body [3, 5]. On the other hand, computer-aided modeling of animal faces is still not being touched, or we can say that we have seen few works on this attempt. The fact that animals are much less cooperative than human beings leads to a lack of 3D animal scans, and the huge diversity of animal types causes 3D reconstruction more challenging.

In this paper, we mainly concentrate on one type of animal, cat-like animal, as shown in Figure 2. 3D Morphable Model (3DMM) is a classic 3D statistic model of human face shape and texture introduced in 1999 [7]. It has become a well-established technology adopted to perform various tasks in many areas, such as computer vision, human behavioral analysis, computer graphics and so on [6, 2, 4].

Figure 1. 3D reconstruction result of matching the Morphable Model CAFM to the input image.

The basic assumption of 3DMM is that a new human face can be expressed as a linear combination of the shape of m exemplar faces. As we know, according to Taxonomy, the same family animals share similar face geometry and head shape. Therefore, 3DMM can be applied in cat-like animal applications.

As a parametric model, 3DMM is constructed by performing a data compression technique, typically Principal Component Analysis (PCA), on training facial meshes. Based on the average shape of the face, a set of shape coefficients can describe a face. With projection parameters, the corresponding rendering can be generated. This is working if and only if each mesh is constructed in a consistent form where the order and number of vertices, the triangulation, and anatomical meaning of every vertex are in accord among all meshes. If the model meshes satisfy the above conditions, we can say they are in dense correspondence. As the correspondence is solved, it is reasonable to build the 3DMM with the meshes.

Especially for larger appearance variations of animals, the powerful priors on 3D face shape of 3DMM provide more discriminative features captured shape information, and it can be leveraged in fitting algorithms of in-the-wild 2D images.

Cats, as a popular choice for pets, play an important role in our life. It means a large number of cat images have been uploaded and shared on the web, which gives us a way to get and label the training data. In this paper, we re-labeled 10,000 cat images improved on previous data set [13] with

Figure 2. The visualization of cat-like animals. Upper: The first row is the big cats, including lion, tiger, puma, leopard. Lower: the second row is the cats, including American Shorthair, British Shorthair, Scottish Fold, Munchkin cat.

15 landmarks, including ears, eyes, nose, and mouth. And we generate the corresponding 3D meshes by computing projection parameters and shape parameter of CAFM.

In summary, in order to generate a 3D Morphable Model of cat-like animals and match this model to 2D images, we address three main challenges:

1. To fill up the blank of 3D reconstruction of the animal face, we propose the Morphable Model, CAFM. With a set of parameters, the parametric model is able to generate a 3D model avoiding unnatural appearance.

2. To match the constructed linear face model to input images, we describe a method matching the model to 2D data and obtaining the shape parameter and projection parameters.

3. To enable the fitting process of the CAFM, we enhance the cat image database containing 2D cat face images and 15 landmarks in pairs.

The constructed cat-like animal face dataset is released at [1] containing pairs of 2D face images with 15 landmarks and 3D face meshes with projection parameters. The Morphable Model CAFM and the fitting algorithm code are also released at the same time.

2. Model

To build a 3D Morphable Model, we need a sample set with a big variety of face shapes. Due to the larger appearance variations of animals compared with human beings, the samples should be representative. The sample set for CAFM consists of 50 animals, including lion, tiger, cougar, leopard, Abyssinian cat, American Shorthair, British Shorthair, Munchkin cat, Persian cat, Scottish Fold and Siamese cat.

2.1. Sample Construction

The samples such as lion, tiger, cougar, and leopard come from The SMAL Model [14]. This team created the animals' models by scanning toy figurines using an Artec

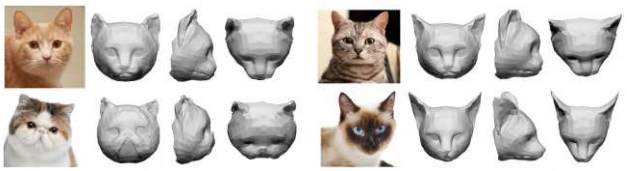

Figure 3. The front view, the side view, and the overlook view of our created 3D cat meshes. The first row shows samples of American Shorthair and British Shorthair. The second row shows samples of Persian cat and Siamese cat.

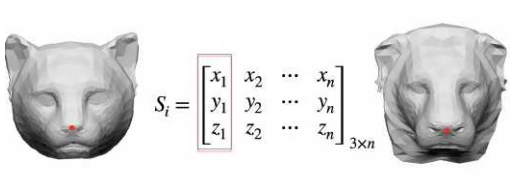

Figure 4. Every index in shape matrix corresponds to the same semantically vertex. In this example, the first index leads to the nose tip.

hand-held 3D scanner. Here we select the head of these models with $n = 1256$ vertices and they shared topology.

We manually create 3D meshes of cat face in ZBrush [12] which is a digital sculpting tool that combines 3D/2.5D modeling, texturing and painting. The created cats include Abyssinian cat, American Shorthair, British Shorthair, Munchkin cat, Persian cat, Scottish Fold and Siamese cat, as shown in Figure 3. Properties of them like size, coat, energy, and shedding vary widely, but they still share similar features. They are consistent with the above structure with $n = 1256$ vertices and the same topology. In other words, the samples are in dense correspondence, as shown in Figure 4. The semantically corresponding vertices such as the nose tip own the same index in every mesh.

We estimate a scaling factor so animals from different backgrounds are in frontal view and comparable in size. The structured 3D models provide semantically corresponding points with the same index in the parametrization domain. To correspondent to the landmarks in cat images, we manually select N landmarks of ears, eyes, nose, and mouth, as shown in Figure 5.

2.2. 3D Morphable Model

The geometry of a face is defined by the shape matrix as follow,

$$S = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \\ z_1 & z_2 & \cdots & z_n \end{bmatrix}_{3\times n}, \quad (1)$$

which contains the x, y, z coordinates of $n = 1256$ vertices. Blanz et al. [6] propose the 3D Morphable Model to de-

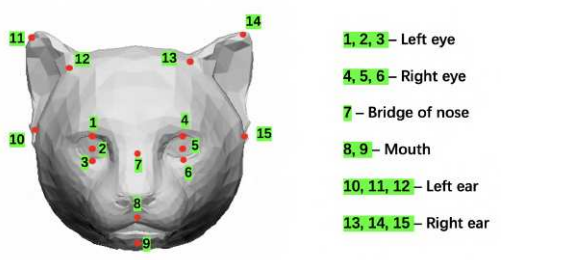

Figure 5. The illustration of 3D landmarks in the sample. We have manually selected landmarks such as ears, eyes, nose, and mouth corresponding to the 2D landmarks in images.

Figure 6. The illustration of 2D landmarks in the input image. We have manually labeled landmarks such as ears, eyes, nose, and mouth corresponding to the 3D landmarks in the mesh.

scribe 3D face space with PCA. With m samples, the new shape of a face is formulated as:

$$S(\boldsymbol{\alpha}) = \bar{S} + \sum_{i=1}^{m-1} \alpha_i s_i = \bar{S} + \boldsymbol{\alpha}\boldsymbol{s}, \tag{2}$$

where $\bar{S} \in \boldsymbol{R}^{3\times n}$ is the mean shape, $\boldsymbol{s} = [s_1, s_2, ..., s_{m-1}]$ is the shape base and $s_i \in \boldsymbol{R}^{3\times n}$, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, ..., \alpha_{m-1}]$ is the corresponding shape parameter. In addition, $\sigma_i \in R$ obtained from the PCA process is the standard deviation of α_i and $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, ..., \sigma_{m-1}] \in \boldsymbol{R}^{m-1}$.

Any 3D face model can be projected onto 2D image space with weak perspective projection:

$$I = fP_rR[\bar{S} + \boldsymbol{\alpha}\boldsymbol{s}] + t_{2d}, \tag{3}$$

where $I \in \boldsymbol{R}^{2\times n}$ is the projection leading to the 2D positions of 3D transformed vertices, f is the scale factor, $P_r = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ is the orthographic projection matrix, $R = R(pitch, yaw, roll)$ is the 3×3 rotation matrix constructed from pitch, yaw, and roll rotation angles, and t_{2d} is the $2 \times n$ translation matrix. The collection of all the model parameters is $p = [f, pitch, yaw, roll, t_{2d}, \boldsymbol{\alpha}]$ included projection parameters and shape parameter.

2.3. Matching the Morphable Model to Images

One of the core tasks of the parametric model is matching the dense 3D Morphable Model to 2D face images. Coefficients of the 3D model and projection parameters can be optimized by minimizing the difference between the projection of the generated 3D model and the input image. Utilizing the exiting cat alignment database labeled with 2D landmarks, we can estimate the parameter p by analysis-by-synthesis. The N 2D landmarks in cat image are in accordance with 3D landmarks in the mesh model, as shown in Figure 6. We denote x and y coordinates of N semantically meaningful landmarks in the image plane as a matrix U:

$$U = \begin{bmatrix} u_1 & u_2 & \cdots & u_N \\ v_1 & v_2 & \cdots & v_N \end{bmatrix}. \tag{4}$$

Since we aim to compare the difference between the 3D model and cat image in the image plane, 3D landmarks are needed to be projected from 3D space to image space. We denote 3D landmarks as matrix $\tilde{S}(\boldsymbol{\alpha})$,

$$\tilde{S}(\boldsymbol{\alpha}) = S(\boldsymbol{\alpha})[:, d], \tag{5}$$

where d is the N-dim index vector indicating the indexes of 3D landmarks in mesh, thus $\tilde{S}(\boldsymbol{\alpha})$ is a $3 \times N$ landmark matrix selected from $S(\boldsymbol{\alpha})$ which contains x, y, z coordinates. The relationship between the 3D landmarks $\tilde{S}(\boldsymbol{\alpha})$ and its projected 2D landmarks matrix $U(p)$ can be described as follow by making use of Equation 3,

$$U(p) = fP_rR\tilde{S}(\boldsymbol{\alpha}) + t_{2d}. \tag{6}$$

We also denote the labeled 2D landmarks in the input image as $U_i(p)$,

$$U_i(p) = \begin{bmatrix} u_{i_1} & u_{i_2} & \cdots & u_{i_N} \\ v_{i_1} & v_{i_2} & \cdots & v_{i_N} \end{bmatrix}. \tag{7}$$

Then we can use the following objective function to estimate parameter p,

$$L(p) = ||U - U_i||_F^2 + \lambda \sum_{j=1}^{m-1} (\frac{\alpha_j}{\sigma_j})^2. \tag{8}$$

Matching 3D shape to an input image is an ill-posed problem. The regularization appended here is to add information in order to solve the ill-posed problem and to prevent overfitting. λ is the parameter that controls the importance of the regularization term.

By minimizing the difference between 2D labeled landmarks in the input image and the projection of 3D landmarks in the model, we can estimate the shape parameter and projection parameters. We initialize the shape parameter and projection parameters and optimize them alternately in every iteration. In a loop, the algorithm generates a projection from the current parameters in p and update them. The landmark matching process is summarized in Algorithm 1.

Algorithm 1 Matching the Morphable Model to Images

Require:
2D labeled landmarks in the input image;

Ensure:
parameters in p included shape parameters $\boldsymbol{\alpha}$ and projection parameters f, R, t_{2d};

1: Initialize shape parameter $\boldsymbol{\alpha} = 0$;
2: Compute projection parameters $[f, R, t_{2d}]$ according to the Gold Standard Algorithm [9];
3: Substitute projection parameters $[f, R, t_{2d}]$ from step 2 into Equation 8 and update shape parameter $\boldsymbol{\alpha}$;
4: Repeat step 2&3.
5: **return** parameters in p containing shape parameter $\boldsymbol{\alpha}$ and projection parameters $[f, R, t_{2d}]$;

3. Experiments

In this section, we present and analyze the morphable model constructed. We also provide the results and some details of matching the model to images.

3.1. CAFM

In this section, we show the representation power of CAFM. According to Section 2.1, there are 50 cat-like animal samples. As shown in Equation 2, m is the number of samples and here equals to 50, the new shape of a face can be expressed as,

$$S(\boldsymbol{\alpha}) = \bar{S} + \sum_{i=1}^{49} \alpha_i s_i = \bar{S} + \boldsymbol{\alpha}\boldsymbol{s}. \tag{9}$$

Given a set of shape parameters $\{\boldsymbol{\alpha}\}_{j=1}^{K}$ that activate every single principal component respectively, the new shape of a face can be generated following Equation 9, as shown in Figure 7. The expressiveness of the Morphable Model is augmented by dividing faces into independent sub-spaces.

As we know, even from the same family, animals still vary from size, coat, energy to shedding. We observe that based on the average cat-like animal face $\bar{S}$, the principal components of the model catch the features of this family, and the deformation is reasonable. Furthermore, the expressiveness of the Morphable Model is augmented, and results show that the generated face is given vitality separated from the samples.

3.2. Matching CAFM to Images

Given cat images and the Morphabel Model CAFM, we can generate 3D reconstructions of input cat images with projection parameters $[f, R, t_{2d}]$ and shape parameter $\boldsymbol{\alpha}$.

Our data set includes 10,000 cat images with 15 landmarks. These in-the-wild cat face images without collecting 3D face scans are near the frontal view. Exploiting the methodology mentioned in Section 2.3, we are able to estimate the projection parameters $[f, R, t_{2d}]$ and shape parameter $\boldsymbol{\alpha}$, and then reconstruct its 3D model, see Figure 8. Generated 3D face almost reproduce the input image, from the position to its identity.

we construct a cat-like animal face dataset containing 10,000 pairs of 2D face images with 15 landmarks and 3D face meshes with projection parameters.

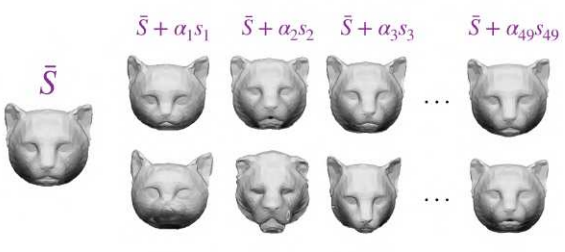

Figure 7. The visualization of CAFM. Right: the average shape of the cat-like animal face which is the $\bar{S}$ in Equation 2. Left: indicates how single principal component effects the results of the model, α_i is randomly selected from 0 to 1.

Figure 8. Cat-like animals from images. From left to right: the input images with landmarks, the visualization of generated 3D meshes, and the value of projection parameters.

4. Conclusion

In this study, we present the first 3D Morphable Model for the animal face which is rarely touched ever. Also, we construct a cat-like animal face dataset. Considering data plays an essential role in deep learning, the work makes an important contribution to the development of the field for animals.

References

[1] https://github.com/sunyifan2017/CAFM-A-3D-Morphable-Model-for-Animals.

[2] O. Aldrian and W. A. Smith. Inverse rendering of faces with a 3d morphable model. *IEEE transactions on pattern analysis and machine intelligence*, 35(5):1080–1093, 2012.

[3] B. Allen, B. Curless, B. Curless, and Z. Popović. The space of human body shapes: reconstruction and parameterization from range scans. In *ACM transactions on graphics (TOG)*, volume 22, pages 587–594. ACM, 2003.

[4] B. Amberg, R. Knothe, and T. Vetter. Expression invariant 3d face recognition with a morphable model. In *2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–6. IEEE, 2008.

[5] M. Andriluka, S. Roth, and B. Schiele. Monocular 3d pose estimation and tracking by detection. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 623–630. IEEE, 2010.

[6] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Transactions on pattern analysis and machine intelligence*, 25(9):1063–1074, 2003.

[7] V. Blanz, T. Vetter, et al. A morphable model for the synthesis of 3d faces. In *Siggraph*, volume 99, pages 187–194, 1999.

[8] J. Booth, A. Roussos, S. Zafeiriou, A. Ponniah, and D. Dunaway. A 3d morphable model learnt from 10,000 faces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5543–5552, 2016.

[9] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[10] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter. A 3d face model for pose and illumination invariant face recognition. In *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 296–301. Ieee, 2009.

[11] L. Tran and X. Liu. Nonlinear 3d face morphable model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7346–7355, 2018.

[12] Wikipedia contributors. Zbrush — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=ZBrush&oldid=926119083, 2019. [Online; accessed 20-December-2019].

[13] W. Zhang, J. Sun, and X. Tang. Cat head detection-how to effectively exploit shape and texture features. In *European Conference on Computer Vision*, pages 802–816. Springer, 2008.

[14] S. Zuffi, A. Kanazawa, D. Jacobs, and M. J. Black. 3D menagerie: Modeling the 3D shape and pose of animals. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, July 2017.