# Pinscreen: 3D Avatar from a Single Image

Shunsuke Saito Lingyu Wei Jens Fursund Liwen Hu Chao Yang Ronald Yu
Kyle Olszewski Stephen Chen Isabella Benavente Yen-Chun Chen Hao Li

Pinscreen, Inc.

**Figure 1:** *Our system can automatically build a complete avatar model including face and hair meshes from a single input image.*

## Abstract

The age of social media and immersive technologies has created a growing need for processing detailed visual representations of ourselves as virtual and augmented reality is growing into the next generation platform for online communication, connecting hundreds of millions of users. A realistic simulation of our presence in a mixed reality environment is unthinkable without a compelling and directable 3D digitization of ourselves. With the wide availability of mobile cameras and internet images, we introduce a technology that can build a realistic 3D avatar from a single photograph. This textured 3D face model includes hair and can be instantly animated by anyone in real-time through natural facial performances captured from a regular RGB camera. Immediate applications include personalized gaming and VR-enabled social networks using automatically digitized 3D avatars, as well as mobile apps such as video messengers (e.g., Snapchat) with face-swapping capabilities. As opposed to existing solutions, our technology enables the automatic generation of a complete head model from a fully unconstrained image.

**Keywords:** face capture, avatar generation, hair modeling

**Concepts:** •**Computing methodologies** → **Image manipulation;**

## 1 Introduction

Our objective is to build a fully textured blendshape model of a subject's face (arbitrary lighting, facial expressions, etc.) from a single unconstrained image including a hair mesh. While the fitting of 3D face models to images has been extensively explored in the vision and graphics community, the digitization of hair models has only been possible recently with some manual input. One of the key steps for hair modeling requires a pixel-level segmentation of hair regions. As opposed to faces, whose geometry and appearances can be approximated effectively using linear models, the dimensionality and complexity of hairstyle variations are substantially larger. With the recent success of deep learning techniques for semantic segmentation tasks, we introduce a deep convolutional neural network for robust hair segmentation and develop a data-driven framework for 3D hair modeling from segmented hair images. A 3D hair mesh is generated using boundary constraints from the hair segmentation, the head scalp of a 3D fitted face model, as well as shape priors obtained from a large hairstyle database [Hu et al. 2015]. The face and hair textures are synthesized using a combination of data-driven inpainting and a neural synthesis approach. Once the face model is fitted to the input image, we directly obtain identity and expression blendshape coefficients of the subject's face.

This head model can then be animated using a standard RGB-based facial performance capture framework which produces the corresponding animation coefficients of a performer's face [Saito et al. 2016]. These animation curves are finally used for retargeting the generated avatar which can be the user's face or someone else's. Our live demonstration runs on a simple MacBook Pro. We will demonstrate using both the integrated iSight camera as well as a higher quality Logitech webcam and display the results on a larger TV screen. Our system consists of a standard Mac OS X application which communicates with a Linux program (installed on a virtual machine that runs in parallel) for access to a deep convolutional neural network framework (Caffe [Jia et al. 2014]). In summary, we introduce the first system that can automatically build a complete avatar model that includes face and hair meshes from a single input image. Our hair modeling algorithm combines recent advances in deep learning-based segmentation and data-driven shape modeling.

**Acknowledgements.** We would like to thank Bagelmo for the meatful discussions.

## References

HU, L., MA, C., LUO, L., AND LI, H. 2015. Single-view hair modeling using a hairstyle database. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2015) 34*, 4 (July).

JIA, Y., SHELHAMER, E., DONAHUE, J., KARAYEV, S., LONG, J., GIRSHICK, R., GUADARRAMA, S., AND DARRELL, T. 2014. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, ACM, 675–678.

SAITO, S., LI, T., AND LI, H. 2016. Real-time facial segmentation and performance capture from rgb input. In *Proceedings of the European Conference on Computer Vision (ECCV)*.


SA '16 Emerging Technologies, December 05-08, 2016, Macao
ISBN: 978-1-4503-4514-9/16/12
DOI: http://dx.doi.org/10.1145/2988240.3014572