ACM SIGGRAPH

SIS Conference Reporting

506 - Submission Reviews - By Person

Generated: March 09, 2017 13:50:44 EST

Question	Response
Id:	papers_0506
Title:	Avatar Digitization From a Single Image
Reviewer #16:	
1) Description	This submission presents a system for automatically creating a 3D avatar including the face and hair from a single unconstrained image. The authors incorporated some latest techniques for facial, hair modeling and high-quality texture synthesis to build a complete framework for single image based avatar digitization. It is a very challenging task but the system chooses proper techniques to overcome the difficulties and the results are convincing. The single image based face digitization is achieved using the method proposed in [Saito et al 2016], which reconstructs a face using multilinear PCA model [Thies et al 2016a]. The high-fidelity face texture is inferred using a DL-based transfer algorithm. The hair strands are initially found and warpped to the input image using the methods proposed in [Hu et al 2015, Chai et al 2016], then converted to strips. The hair texture is synthesized using a data-driven method. The authors integrated many existing methods, and the main contribution of the submission is a deep learning-based hair texture synthesis methods. I like the very interesting system, but won't strongly recommend to accept it as a technical paper.
2) Resubmission Evaluation	
3) Clarity of Exposition	This submission is well written and easy to follow. the texts, illustrations and images are clearly presented. However, since the section 4 is almost equivalent to [Saito et al 2016], it can be simplified a little bit. There are several small typos: 1. From line 277, both of the two consecutive sentences contain 'first'. 2. In line 279, the word "based" seems superfluous. 3. Line 519, we -> We
4) Quality of References	The references are good and cited properly thoughout the paper. However, I think the following paper should be cited as the multilinear PCA face model is similiar to the the submission: Blanz et al, Reanimating Faces in Images and Video, 2003
5) Reproducibility	Yes it's reproducible. The algorithms are well described and some techniques used in the system have been properly cited as well.
6) Rating	3.0
8) Explanation of Rating	This submission presents a complete solution to generate an avatar head from a single image. The system integrated many existing advanced techniques. The only novelty lies in the hair texture synthesis using a deep learning method. The face digitization and head digitization processes are also totally separated. There are not interaction between these two parts. It is unclear how the hairs are added onto the reconstructed heads. Is there any collisiion avoidance or adaption applied? The submission looks like a simple combinition of two problems. The comparison shown in Fig. 11. seems not that fair. Generating geometric hair models are not the purpose of [Thies et al 2016] and [Ichim et al 2015] at all. In my opinion, the authors can simplify the section 4 and focus more on single image based hair modeling. In summary, although the submission proposed a powerful system composed of solid techniques, my conclusion is that it is still a borderline paper.
Reviewer #94:	
1) Description	This paper proposes a fully automatic method for generating a 3D face model with skin texture, rig, and hair for a given single 2D image of a face. The deep learning techniques are then applied to generate the texture and hair, based on the large database of skin and hair, respectively. In addition, many other state-of-the-art techniques are employed to make the high quality face model. I hope the model would be good for video game use. I think this submission is fun to know how we can make a realistic 3D face model from a single image input, using the various state-of-the-art techniques. On the other hand, I could not understand the limitations and drawbacks of the method. It's also not clear for me about the ultimate goal, application(s) and future direction of this research.
2) Resubmission Evaluation	
3) Clarity of Exposition	Yes, it is well described, but I think the submission is a bit long. For example, since one of the deep learning approaches for skin texture generation has been recently published, Sect. 4 can be shortened.
4) Quality of References	Yes, I think the references are good enough for me.
5) Reproducibility	I think so, while we need to know lots of state-of-the-art techniques employed in the proposed method.
6) Rating	3.1
	This submission shows a great challenge of making a realistic 3D face model that can be animated, with a single face image as input. The authors make careful and nice treatment of the state-of-the-art techniques to make this possible. We can therefore know the "state-of-the-art" integration approach developed by the authors. In this sense I think this submission is a good paper. On one hand, it'd be nice to describe more about the relation between the

8) Explanation of Rating	techniques and the workflow steps (in Figure 2), for example, as a diagram or chart. On one hand, as the authors told in the paper, the resultant model is not compelling, while I agree that it could be used for a video game. It would be great if the authors can show a few concrete application examples (even in the game would be fine). Or, as mentioned earlier, the authors should discuss the ultimate goal, application(s) and future direction of this research. The author could then evaluate the proposed method in the scenario of this challenging research. As for the skin texture technique, it's been recently published as a CVPR paper. So I think the authors should make a more short and concise description of section 4. The main technical novelty of the method may then lie in the hair modeling step. I would recommend the authors to detail this part and to show various hairstyle examples, because this will demonstrate the power of this method, and differentiate this paper clearly from the CVPR paper. This submission is well written, and the framework in Figure 2 is interesting. However I'd like to see more convincing results, such as, showing hairstyle variations or more realistic face models.

Reviewer #41:

1) Description	This paper presents a fully automatic framework for creating a complete 3D avatar from a single unconstrained image. This framework digitizes an entire model using a textured mesh representation for the head and volumetric strips for the hair. Digitized models can be integrated into existing game engines. The proposed system integrates recent advances in facial shape modeling, appearance inference, and a pipeline for single-view hair generation based on hairstyle retrieval from a database, followed by a strand-to-hair-strip conversion method. The paper introduces an algorithm for realistic hair texture synthesis for the strips based on feature correlation analysis using a deep neural network. The contributions of this paper include a fully automatic framework for 3D avatar creation from a single image, a pipeline for single-view strip-based hair generation. The magnitude of contribution is limited by the fact that the proposed framework and pipeline basically assemble techniques from multiple existing papers.
2) Resubmission Evaluation	
3) Clarity of Exposition	Clear
4) Quality of References	Adequate
5) Reproducibility	This work could be reproduced. System details have been adequately discussed. Limitations have been discussed.
6) Rating	2.8
8) Explanation of Rating	This paper demonstrates that the fully automatic digitization of 3D avatars, including hair, is possible from a single image. The proposed system integrates recent advances in facial shape modeling, appearance inference, and a pipeline for single-view hair generation based on hairstyle retrieval from a database. It also performs realistic hair texture synthesis by adapting an existing deep learning based algorithm for image stylization. However, the contribution of this paper is limited by the fact that the proposed framework basically assembles techniques from multiple existing papers. The proposed system demonstrates those integrated techniques do work, but does not provide fresh ideas and insights itself. Even though claimed so, the pipeline for single-view hair generation does not look novel. A novel pipeline should have at least one novel component. However, in the proposed pipeline, none of the components is novel. The steps for strand-based hair generation are based on [Hu et al. 2015] and [Chai et al. 2016]. Hair mesh generation is based on marching cubes and the method from [Zhu and Bridson 2005]. Hair strip generation is based on [Luo et al. 2013]. By the way, the generated hair strips do not seem to have high quality for short hairstyles. Likewise, the algorithm for hair texture synthesis does not look novel either since it is the same as one of the components from [Gatys et al. 2016] originally developed for image style transfer.

Reviewer #42:

1) Description	This paper introduces a fully automatic system for creating a virtual avatar from a single input image. The created avatar consists of both expression blendshapes and a strip-based hair model with synthesized textures. Compared with state-of-the-art avatar generation techniques that all requires multiple images as input, the described system only needs a single image, which makes it more appealing to consumer applications. However, the novelty of the work and the quality of the generated avatars are below the SIGGRAPH standard (see comments below).
2) Resubmission Evaluation	
3) Clarity of Exposition	Yes, the exposition is clear.
4) Quality of References	References are good.
5) Reproducibility	It's doable, but not easy as it requires to integrate a few previous work. Limitations are discussed.
6) Rating	2.5
	The problem studied in this paper is quite appealing to VR/mobile applications. The proposed system consists of two independent components: one for the face fitting and the other for the hair modeling. The face fitting part is a straightforward combination of [Kazemi and Sullivan 2014; Thies et al. 2016a] (for geometry) and [Saito et al. 2016b] (for texture). The hair modeling part is very similar to the fully automatic pipeline of AutoHair [Chai et al. 2016] that generates strand-based hair models. The technical contributions of the work are limited to (all in page 6):

8) Explanation of Rating	- a strand-to-hair-strip conversion method to generate the strip-based hair models, and - a hair texture synthesis for the strips. Actually, in my opinion, as automatic methods for fitting a morphable face model (with texture) to a single image have long been available (dating back to 1999), fully automatic generation of dyanmic avatars (with hair) has been enabled by the AutoHair pipeline from last year. The system presented in this paper only added an extra strand-to-hair-strip conversion step to that pipeline. In this regard, the scientific contribution of the paper is minimal. Results presented in the paper and video are not satisfactory. A lot of disturbing artifacts (e.g. in regions around the silhouette) can be observed in almost all hair models shown in the paper. I seriously doubt if the quality is good enough for games or VR applications. For the comparisons shown in Fig. 11, I'd like to see the full models in the video. I also want to see the comparisons between AutoHair and the present system. It's also necessary to rotate the models to let people see the back side of the models. Overall, I am not convinced that the paper is acceptable to SIGGRAPH due to the lack of scientific contributions and the unconvincing results.

Reviewer #79:

1) Description	Given a single photo (in the wild) the method produces an animatable avatar of the person that includes the face + hair. The paper proposed a hair representation via hair strips (rather than hair strands). The results presented in the video and paper are impressive! State of the art can either reconstruct a face+hair model from multiple views (Cao et al), from a video (Ichim et al.), or rough models from Internet photos (Liang et al.). This paper creates very cool avatars from a single photo. Steps of the algorithm: 1) segment the hair and face using Cao et al. 2) fit the a PCA linear face model to the face part, and a PCA model to the texture part. use Saito et al for filling in missing texture parts 3) add generic teeth, tongue, and eyeballs models 4) hair modeling: the hair inside the segmented mask is matched to a dataset of artist created hairs following Hu et al. The match is adjusted to match the silhouette of the input hair. Hair strips are created to reduce complexity (from hair strands) following Luo et al. Texture of the hair is first estimated globally and then details are added per strip. 5) animation is done via blend shapes, while keeping the hair attached and move with the face
2) Resubmission Evaluation	
3) Clarity of Exposition	the exposition is clear
4) Quality of References	yes references are great, some additional related works: https://pdfs.semanticscholar.org/2e91/8398b29c073580549f3b4183ffb38b34ce46.pdf and commercial similar algorithms, see loom.ai: http://demo.loomai.com/signup would be interesting to compare to loom.ai results too.
5) Reproducibility	The work is possible but challenging to reproduce since lots of components are included in the system. While all the individual components are published works, they require lots of knowhow and the code is not published. The limitations of the work are clear.
6) Rating	4.0
8) Explanation of Rating	I loved the results, very cool. This is the first time someone shows animatable face + hair modeling from a single photo. I would like to see in the rebuttal a few additional discussions: 1) for how many photos does it work, and more importantly when does the system fail. Under which conditions, is it lighting, bad fiducial detection, problems in segmentation, etc. I'd like to see a discussion of all the components and how they may fail. 2) which hairstyles are problematic to this system. please explain and show results in the paper if it is accepted. 3) please show and discuss how the back of the head looks. Since the reconstruction happens from a single photo, the back of the head cannot be reconstructed reliably.

Submission Information: